Peter Carr and Qiji Jim Zhu

Convex Duality and Financial Mathematics

April 11, 2017

To Carol and Olivia

To Lilly and Charles.

And in memory of Jonathan Borwein (1951-2016) with respect.

Preface

Convex duality plays an essential role in many important financial problems. For example, it arises both in the minimization of convex risk measures and in the maximization of concave utility functions. Together with generalized convex duality, they also appear when an optimization is not immediately apparent, for instance in implementing dynamic hedging of contingent claims. Recognizing the role of convex duality in financial problems is crucial for several reasons. First, considering the primal and dual problem together gives the financial modeler the option to tackle the more accessible problem first. Usually, knowledge of the solution of one helps in solving the other. Moreover, the solution to the dual problem can usually be given a financial interpretation. As a result, the dual problem often illuminates an alternative perspective, which is not easily achieved by examining the primal problem in isolation. When flipping from the primal to the dual, a surprise insight typically awaits, irrespective of past experience. Finally, as an added benefit, the primal and the dual can often be paired together to provide better numerical solutions than when either side is considered in isolation.

The goal of this book is to provide a concise introduction to this growing research field. Our target audience is graduate students and researchers in related areas. We begin in Chapter 1 with a quick introduction of convex duality and related tools. We emphasize the relationship between convex duality and the Lagrange multiplier rule for constrained optimization problems. We then give a quick overview of the intrinsic duality relationship in several diverse financial problems.

In Chapter 2, we consider the simplest possible financial market model. In particular, we consider a one period economy with a finite number of possible states. Using this simple financial market model, we showcase convex duality in a number of important financial problems. We begin with the Markowitz portfolio theory, which involves a particularly simple convex programming problem: optimizing a quadratic function with linear constraints. Duality plays two important roles in Markowitz portfolio theory. First, while the primal problem may involve hundreds or even thousands of variables representing the risky assets potentially included in the portfolio, the dual problem has only two variables related to the two constraints on the initial endowment and the expected return. In fact, the key observation of Markowitz is that one can evaluate the performance of a portfolio in the dual space using the variance - expected return pair. Second, the duality relationship between the primal Markowitz portfolio problem and its dual help us to understand that the set of optimal portfolios is an affine set, which leads to the important two fund theorem. The core methodology of optimizing a quadratic function with linear constraints was also used in the capital asset pricing model, which leads to the widely used Sharpe ratio. Duality also plays a crucial role in this problem.

Next, we consider portfolio optimization from the perspective of maximizing expected utility. There has been a very long history of using utility functions in economics. In financial problems, utility functions are increasing concave functions of wealth. The concavity of the utility function captures the risk aversion of an investor. Arrow and Pratt introduced widely used measures of the level of risk aversion. It turns out that there is a precise way of using generalized convexity to characterize Pratt–Arrow risk aversion. This application illustrates the relevance of generalized convexity in dealing with financial problems. It is even more interesting to consider the dual of the expected utility maximization problem. It turns out that in the absence of arbitrage, solutions to the dual problem are in essence the equivalent martingale measures (also called risk-neutral probabilities), which are widely used in pricing financial derivatives. Considering the expected utility maximization problem along with its dual leads us to rediscover the fundamental theorem of asset pricing. An added benefit of this alternative approach is that martingale measures can be related to the risk aversion of agents in the market.

The last application that we cover in Chapter 2 concerns the dual representation of coherent risk measures. Coherent risk measures are motivated by the common regulatory practice of assigning each position in a risky asset with the appropriate amount of cash reserves. Hence, they are widely used to analyze risks. Mathematically, a coherent risk measure is characterized by a sub-linear function: a convex function with positive homogeneity. It is well known that the dual of a sub-linear function is an indicator function. Thus, using dual representation, a coherent risk measure is just the support function of a closed convex set. Financially, we can view the generating set of a coherent risk measure as the probabilities assigned to risky scenarios in a stress test. Duality also generates numerical methods for calculating some important coherent risk measures such as the conditional value at risk.

We expand our discussion to a more general multi-period financial market model in Chapter 3. This more general setting allows us to model dynamic trading. The added complexity in dealing with a multi-period model mainly involves capturing the increase in information using an information structure. After laying out the multi-period financial market model, we show that the fundamental theorem of asset pricing also arises in a multi-period financial market model. After that we also discuss two new topics: super (sub) hedging and conic finance. In general, the absence of arbitrage leads to multiple (usually infinitely many) pricing martingale measures in an incomplete market. Thus, the no arbitrage principle usually determines a price range for a contingent claim with upper and lower bounds, which are given by the supremum and the infimum of the expectation of the payoff under the martingale measures, respectively. If a market price falls outside of these bounds, then an arbitrage opportunity occurs. It turns out that the dual solution to the optimization problem of finding the upper or lower no arbitrage bounds provides a trading strategy that one can use to take advantage of such an arbitrage opportunity. Conic finance is used to describe financial markets for which the absolute value of the price depends on whether one is buying or selling. In other words, conic finance describes realistic financial markets with a strictly positive bid-ask spread. In such a model, the cash flows that can be achieved from implementing acceptable trading strategies form a convex cone. This observation provides the rationale for the name conic finance. Despite the added complication of dealing with a conic constraint, we show that most of the duality relationships that are observed under zero bid–ask spread still prevail when the spread is positive.

We then move to continuous-time financial models in Chapter 4. The most noteworthy duality relationship developed in this chapter is the observation that the classical Black-Scholes formula for pricing a contingent claim with a convex payoff is, in fact, a Fenchel-Legendre transform. We show that the function describing cash borrowings while delta hedging a short position in a contingent claim is just the Fenchel conjugate of the contingent claim pricing function. The flip side is that the contingent claim pricing function can itself be viewed as a Fenchel conjugate of the function describing these cash borrowings. This provides a new perspective on the convex function linking the price of the contingent claim to the underlying spot price. With the availability of many tradable contingent claims such as those embedded in ETF's, the ability to dynamically hedge a contingent claim with other contingent claims is increasingly becoming a financial reality. Interestingly, when using contingent claims as hedging instruments, one discovers a similar duality relationship between the contingent claim pricing function and the cash borrowings function in terms of generalized convexity. Many useful applications are also discussed in this chapter. We examine the convexity and generalized convexity of the Bachelier and Black-Scholes option pricing formulae with respect to volatility as well. Generalizations of these properties might be useful in dealing with financial products related to volatility and be a potentially fruitful future research direction.

The material in this book grew out of slides used to teach a joint doctoral seminar at New York University's Courant Institute in the fall of 2015. Part of the materials has also been used previously for graduate topic courses on optimization and modelling at Western Michigan University. We thank our colleagues at both NYU and WMU for providing us with supportive research environments. Professor Robert Kohn helped to arrange us becoming neighbors, which facilitated our collaboration in no small part. Conversations with Professors Marco Avellaneda, Jonathan Goodman and Fang-Hua Lin have been most helpful. We are also indebted to the participants of these courses for many stimulating discussions. In particular, we thank Monty Essid, Tom Li, Matthew Foreman, Sanjay Karanth, Jay Treiman, Mehdi Vazifadan, and Guolin Yu whose detailed comments on various parts of our lecture notes have been incorporated into the text.

New York, NY *Peter Carr*
Kalamazoo, MI *Qiji Jim Zhu*
April, 2017

Contents

1

Convex Duality

Summary. We present a concise description of the convex duality theory in this chapter. The goal is to lay a foundation for later application in various financial problems rather than to be comprehensive. We emphasize the role of the subdifferential of the value function of a convex programming problem. It is both the set of Lagrange multiplier and the set of solutions to the dual problem. These relationships provide much convenience in financial applicaitons. We also discuss generalized convexity, conjugacy and duality.

1.1 Convex Sets and Functions

1.1.1 Definitions

Definition 1.1.1 (Convex Sets and Functions) *Let X be a Banach space. We say that a subset C of X is a* convex set *if, for any $x, y \in C$ and any $\lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in C$. We say an extended-valued function $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is a* convex function *if its domain,* dom f*, is convex and for any $x, y \in \operatorname{dom} f$ and any $\lambda \in [0,1]$, one has*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y).$$

We call $f\colon X \to [-\infty, +\infty)$ a concave function *if $-f$ is convex.*

In some sense convex functions are the simplest functions next to linear functions. Convex sets and functions are intrinsically related. For example, it is easy to verify that C is a convex set if and only if $\iota_C(x) := \inf\{\|x - c\| \mid c \in C\}$, its indicator function, is a convex function. On the other hand if f is a convex function then, the epigraph of f, $\operatorname{epi} f := \{(x,r) \mid f(x) \le r\}$ and $f^{-1}((-\infty, a]) := \{x \mid f(x) \in (-\infty, a]\}$, $a \in \mathbb{R}$ are convex sets. In fact, we can check that the convexity of $\operatorname{epi} f$ characterizes that of f. This geometric characteization is very useful in many situations. For instance, it is easy to see that the intersection of a class of convex sets is convex. Now let f_α be a class of convex functions we can see that

$$\operatorname{epi} \sup_\alpha f_\alpha = \cap_\alpha \operatorname{epi} f_\alpha$$

and, thus, $\sup_\alpha f_\alpha$ is convex. In particular, the *support function* of a set $C \subset X$ defined on the dual space X^* by

$$\sigma_C(x^*) = \sigma(C; x^*) := \sup\{\langle x, x^* \rangle \mid x \in C\}$$

is always convex. Note that allowing the extended value $+\infty$ in the definition of convex function is important in establishing those relations.

An important property of convex functions related to applications in economics and finance is the Jensen inequality.

Proposition 1.1.2 (Jensen's Inequality) *Let f be a convex function. Then, for any random variable X on a finite probability space,*

$$f(\mathbf{E}[X]) \leq \mathbf{E}[f(X)].$$

When X has only finite states this result directly follows from the definition. The general result can be proven by approximation.

A special kind of convex set – convex cone is very useful.

Definition 1.1.3 *Let X be a finite dimensional Banach space. We say $K \subset X$ is a convex cone if for any $x, y \in K$ and any $\alpha, \beta \geq 0$, $\alpha x + \beta y \in K$. Moreover, we say K is pointed if $K \cap (-K) = \{0\}$.*

A pointed convex cone K induces a partial order $\leq_K$ by defining $x \leq_K y$ if and only if $y - x \in K$. We can easily check that $\leq_K$ is reflexive ($x \leq_K x$), antisymmetric ($x \leq_K y$ and $y \leq_K x$ implies $x = y$) and transitive ($x \leq_K y$ and $y \leq_K z$ implies that $x \leq_K z$). The definition of convexity can easily be extended to mappings whose image space has such a partial order.

Definition 1.1.4 (Convex Mappings) *Let X and Y be two Banach spaces. Assume that Y has a partial order $\leq_K$ generated by the pointed convex cone $K \subset Y$. We say that a mapping $f\colon X \to Y$ is K-*convex *provided that, for any $x, y \in$* dom f *and any $\lambda \in [0, 1]$, one has*

$$f(\lambda x + (1 - \lambda)y) \leq_K \lambda f(x) + (1 - \lambda) f(y).$$

1.1.2 Convex Programming

We will often encounter various forms of the general convex programming problems below in financial applications in subsequent chapters. Let X, Y and Z be finite dimensional Banach spaces. Assume that Y has a *partial order* $\leq_K$ generated by the pointed convex cone K. We denote the *polar cone* of K by

$$K^+ := \{y^* \in Y^* : \langle y^*, y \rangle \geq 0 \text{ for all } y \in K\}.$$

Consider the following class of constrained optimization problems

$$P(y,z) \qquad \begin{aligned} \text{Minimize} \;\; & f(x) \\ \text{Subject to} \;\; & g(x) \le_K y, \\ & h(x) = z, \\ & x \in C, \end{aligned} \tag{1.1.1}$$

where C is a closed set, $f : X \to \mathbb{R}$ is lower semicontinuous, $g : X \to Y$ is lower semicontinuous with respect to $\le_K$, and $h : X \to Z$ is continuous. We will use $v(y,z)$ to represent the *optimal value* function

$$v(y,z) := \inf\{f(x) \colon g(x) \le_K y, h(x) = z, x \in C\},$$

which may take values $\pm\infty$ (in infeasible or unbounded below cases), and $S(y,z)$ the (possibly empty) solution set of problem $P(y,z)$.

A concrete example is

$$\begin{aligned} \text{Minimize} \;\; & f(x) \\ \text{Subject to} \;\; & g_i(x) \le y_m, m = 1,2,\dots,M, \\ & h_k(x) = z_k, k = 1,2,\dots,K \\ & x \in C \subset \mathbb{R}^N, \end{aligned} \tag{1.1.2}$$

where C is a closed subset, $f, g_m : \mathbb{R}^N \to \mathbb{R}$ are lower semicontinuous and $h_k : \mathbb{R}^N \to \mathbb{R}$ are continuous. Defining vector valued function $g = (g_1, g_2, \dots, g_M)$ and $h = (h_1, h_2, \dots, h_k)$, problem (1.1.2) becomes problem (1.1.1) with $\le_K = \le_{\mathbb{R}^{M}+}$. Beside Euclidean spaces, for applications in this book we will often need to consider the Banach space of random variables.

It turns out that the optimal value function of a convex programming problem is convex.

Proposition 1.1.5 (Convexity of Optimal Value Function) *Suppose that in the constrained optimization problem (1.1.1), function f is convex, mapping g is $\le_K$ convex, and mapping h is affine and set C is convex. Then the optimal value function v is convex.*

Proof. Consider $(y^i, z^i), i = 1,2$ in the domain of v and an arbitrary $\varepsilon > 0$. We can find x^i_ε feasible to the constraint of problem $P(y^i, z^i)$ such that

$$f(x^i_\varepsilon) < v(y^i, z^i) + \varepsilon, \;\; i = 1,2. \tag{1.1.3}$$

Now for any $\lambda \in [0,1]$, we have

$$\begin{aligned} f(\lambda x^1_\varepsilon + (1-\lambda) x^2_\varepsilon) &\le \lambda f(x^1_\varepsilon) + (1-\lambda) f(x^2_\varepsilon) \\ &< \lambda v(y^1, z^1) + (1-\lambda) v(y^2, z^2) + \varepsilon. \end{aligned} \tag{1.1.4}$$

It is easy to check that $\lambda x^1_\varepsilon + (1-\lambda) x^2_\varepsilon$ is feasible for problem $P(\lambda(y^1, z^1) + (1-\lambda)(y^2, z^2))$. Thus, $v(\lambda(y^1, z^1) + (1-\lambda)(y^2, z^2)) \le f(\lambda x^1_\varepsilon + (1-\lambda) x^2_\varepsilon)$. Combining with inequality (1.1.4) and letting $\varepsilon \to 0$ we arrive at

$$v(\lambda(y^1, z^1) + (1-\lambda)(y^2, z^2)) \le \lambda v(y^1, z^1) + (1-\lambda) v(y^2, z^2),$$

that is to say v is convex. •

This is a very potent result that can help us to recognize the convexity of many other functions. For example, let C be a convex set then the distance function to C defined by $d_C(z) = \inf[\|z - c\| : c \in C]$ is a convex function because we can rewrite it as the optimal value of the following special case of problem (1.1.1)

$$d_C(z) = \inf[\|x\| : x + c = z, c \in C].$$

Similarly, the inf-convolution function defined below is convex

$$\begin{aligned} f\Box g(z) &:= \inf_y [f(z-x) + g(x)] \\ &= \inf\{f(u) + r : g(x) - r \leq 0, u + x = z\} \end{aligned}$$

While the value function of a convex programming problem is always convex it is not necessarily smooth even if all the data involved are smooth. The following is an example

$$v(y) = \inf[x : x^2 \leq y] = \begin{cases} -\sqrt{y} & y \geq 0 \\ +\infty & y < 0. \end{cases}$$

1.2 Subdifferential and Lagrange Multiplier

Many naturally arising nonsmooth convex functions lead to the definition of subdifferential as a replacement for the nonexisting derivative.

1.2.1 Definition

Definition 1.2.1 (Subdifferential) *Let X be a finite dimensional Banach space. The subdifferential of a lower semicontinuous function $\phi : X \to \mathbb{R} \cup \{+\infty\}$ at $x \in \mathrm{dom}\, \phi$ is defined by*

$$\partial\phi(x) = \{x^* \in X^* : \phi(y) - \phi(x) \geq \langle x^*, y - x\rangle \ \ \text{for all } y \in X\}.$$

We define the *domain* of the subdifferential of ϕ by

$$\mathrm{dom}\, \partial\phi = \{x \in X \mid \partial\phi(x) \neq \emptyset\}.$$

An element of $\partial\phi(x)$ is called a *subgradient* of ϕ at x.

Definition 1.2.2 (Normal Cone) *For a closed convex set $C \subset X$, we define the* normal cone *of C at $\bar{x} \in C$ by $N(C; \bar{x}) = \partial\iota_C(\bar{x})$.*

Sometimes we will also use the notation $N_C(\bar{x}) = N(C; \bar{x})$. A useful characterization of the normal cone is $x^* \in N(C; x)$ if and only if, for all $y \in C$, $\langle x^*, y - x\rangle \leq 0$.

It is easy to verify that if f has a continuous derivative at x then $\partial f(x) = \{f'(x)\}$. At a nondifferentiable point a convex function's subdifferential is usually a set. Here are a few examples.

Example 1.2.3 We can easily verify that

- $\partial\|\cdot\|(0) = B_1(0)$, the closed ball centered at 0 with radius 1, in particular, $\partial|\cdot|(0) = [-1, 1]$.
- $\partial(\cdot)^+(0) = [0, +\infty)$.
- $\partial(\cdot)^-(0) = (-\infty, 0]$.

1.2.2 Nonemptyness of Subdifferential

A natural and important question is that when can we ensure the subdifferential is nonempty. The following Fenchel - Rockafellar theorem provides a basic form of sufficient conditions.

Theorem 1.2.4 (Fenchel - Rockafellar Theorem on Nonemptiness of Subdifferential) *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Suppose that $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. Then the subdifferential $\partial f(\bar{x})$ is nonempty.*

Proof. We observe that $(\bar{x}, f(\bar{x}))$ is a boundary point of the closed set $\operatorname{epi} f$ which has a nonempty interior. Thus by the Hahn-Banach extension theorem there exists a supporting hyperplane of $\operatorname{epi} f$ at $(\bar{x}, f(\bar{x}))$ whose normal vector is $(0, 0) \neq (x^*, r) \in X^* \times \mathbb{R}$. Now, for any $x \in \operatorname{dom} f$ and $u \geq f(x)$, we have

$$r(u - f(\bar{x})) + \langle x^*, x - \bar{x}\rangle \geq 0. \tag{1.2.5}$$

Since $u \geq f(x)$ is arbitrary, $r \geq 0$. Moreover, if $r = 0$, then $\bar{x} \in \operatorname{int} \operatorname{dom} f$ would also imply $x^* = 0$, which yield a contradiction. Thus, $r > 0$. Letting $u = f(x)$ in (1.2.5) we see that $-x^*/r \in \partial f(\bar{x})$. ●

Remark 1.2.5 (Constraint Qualification: Relative Interior) The Fenchel–Rockafellar Theorem is a fundamental result that we will use often in the sequel. Condition $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ is a sufficient condition that can be improved. Notice that we don't need to worry about points at which $f = \infty$. Thus, we need only check the condition of Theorem 1.2.4 on $\operatorname{span}(\operatorname{dom} f)$. Thus, condition $\bar{x} \in \operatorname{int}(\operatorname{dom} f)$ can be revised to $x \in \operatorname{ri}(\operatorname{dom} f)$ and f is lower semicontinuous, where ri signifies the *relative interior*, i.e. interior points on $\operatorname{span}(\operatorname{dom} f)$.

Remark 1.2.6 (Constraint Qualification: Polyhedral Problem) Recall that a set is polyhedral if it is the intersection of finitely many closed half-spaces. A function is polyhedral if its epigraph is a polyhedral set. For a polyhedral function its subdifferential is nonempty in any point of its domain (see e.g. [7]). This sufficient condition is very useful in dealing with linear programming problems.

The conclusion $\partial f(\bar{x}) \neq \emptyset$ can be stated alternatively as there exists a linear functional x^* such that $f - x^*$ attains its minimum at $\bar{x}$. This is a very useful perspective on the use of variational arguments – deriving results by observing a certain auxiliary function attains a minimum or maximum.

1.2.3 Calculus

For more complicated convex functions we need the help of a convenient calculus for calculating or estimating its subdifferential. It turns out that the key for developing such a calculus is to combine a decoupling mechanism with the existence of subgradient. We summarize this idea in the following lemma.

Lemma 1.2.7 (Decoupling Lemma) *Let the functions $f\colon X \to \mathbb{R}$ and $g\colon Y \to \mathbb{R}$ be convex and let $A\colon X \to Y$ be a linear transform. Suppose that f, g and A satisfy the condition*

$$0 \in \operatorname{ri}[\operatorname{dom} g - A \ \operatorname{dom} f]. \tag{1.2.6}$$

Then there is a $y^* \in Y^*$ such that for any $x \in X$ and $y \in Y$,

$$p \leq [f(x) - \langle y^*, Ax\rangle] + [g(y) + \langle y^*, y\rangle], \tag{1.2.7}$$

where $p = \inf_{x\in X}\{f(x) + g(Ax)\}$.

Proof. Define an optimal value function $v\colon Y \to [-\infty, +\infty]$ by

$$\begin{aligned} v(u) &= \inf_{x\in X}\{f(x) + g(Ax + u)\} \\ &= \inf_{x\in X}\{f(x) + g(y) : y - Ax = u.\} \end{aligned} \tag{1.2.8}$$

Proposition 1.1.5 implies that v is convex. Moreover, it is easy to check that $\operatorname{dom} v = \operatorname{dom} g - A \ \operatorname{dom} f$ so that the constraint qualification condition (1.2.6) ensures that $\partial v(0) \neq \emptyset$. Let $-y^* \in \partial v(0)$. By definition we have

$$v(0) = p \leq v(y - Ax) + \langle y^*, y - Ax\rangle \leq f(x) + g(y) + \langle y^*, y - Ax\rangle. \tag{1.2.9}$$

●

We apply the decoupling lemma of Lemma 1.2.7 to establish a sandwich theorem.

Theorem 1.2.8 (Sandwich Theorem) *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ and $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ be convex functions and let $A\colon X \to Y$ be a linear map. Suppose that $f \geq -g \circ A$ and f, g and A satisfy condition* (1.2.6)*. Then there is an affine function $\alpha\colon X \to R$ of the form $\alpha(x) = \langle A^*y^*, x\rangle + r$ satisfying $f \geq \alpha \geq -g \circ A$. Moreover, for any $\bar{x}$ satisfying $f(\bar{x}) = -g \circ A(\bar{x})$, we have $-y^* \in \partial g(A\bar{x})$.*

Proof. By Lemma 1.2.7 there exists $y^* \in Y^*$ such that for any $x \in X$ and $y \in Y$,

$$0 \leq p \leq [f(x) - \langle y^*, Ax\rangle] + [g(y) + \langle y^*, y\rangle]. \tag{1.2.10}$$

For any $z \in X$ setting $y = Az$ in (1.2.10) we have

$$f(x) - \langle A^*y^*, x\rangle \geq -g(Az) - \langle A^*y^*, z\rangle. \tag{1.2.11}$$

Thus,

$$a := \inf_{x\in X}[f(x) - \langle A^*y^*, x\rangle] \geq b := \sup_{z\in X}[-g(Az) - \langle A^*y^*, z\rangle].$$

Picking any $r \in [a, b]$, $\alpha(x) := \langle A^*y^*, x\rangle + r$ is an affine function that separates f and $-g \circ A$. Finally, when $f(\bar{x}) = -g \circ A(\bar{x})$, it follows from (1.2.10) that $-y^* \in \partial g(A\bar{x})$. ●

We now use the tools established above to deduce calculus rules for the convex functions. We start with a sum rule playing a role similar to the sum rule for derivatives in calculus.

Theorem 1.2.9 (Convex Subdifferential Sum Rule) *Let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ and $g\colon Y \to \mathbb{R} \cup \{+\infty\}$ be convex functions and let $A\colon X \to Y$ be a linear map. Then at any point x in X, we have the sum rule*

$$\partial(f + g \circ A)(x) \supset \partial f(x) + A^*\partial g(Ax), \tag{1.2.12}$$

with equality if condition (1.2.6) holds.

Proof. Inclusion (1.2.12) is easy and left as an exercise. We prove the reverse inclusion under condition (1.2.6). Suppose $x^* \in \partial(f + g \circ A)(\bar{x})$. Since shifting by a constant does not change the subdifferential of a convex function, we may assume without loss of generality that

$$x \to f(x) + g(Ax) - \langle x^*, x\rangle$$

attains its minimum 0 at $x = \bar{x}$. By the sandwich theorem there exists an affine function $\alpha(x) := \langle A^*y^*, x\rangle + r$ with $-y^* \in \partial g(A\bar{x})$ such that

$$f(x) - \langle x^*, x\rangle \geq \alpha(x) \geq -g(Ax).$$

Clearly equality is attained at $x = \bar{x}$. It is now an easy matter to check that $x^* + A^*y^* \in \partial f(\bar{x})$. ●

Note that when A is the identity mapping and both f and g are differentiable Theorem 1.2.9 recovers sum rules in calculus. The geometrical interpretation of this is that one can find a hyperplane in $X \times \mathbb{R}$ that separates the epigraph of f and hypograph of $-g$ i.e. $\{(x, r) : -g(x) \geq r\}$.

By applying the subdifferential sum rule to the indicator functions of two convex sets we have parallel results for the normal cones to the intersection of convex sets.

Theorem 1.2.10 (Normals to an Intersection) *Let C_1 and C_2 be two convex subsets of X and let $x \in C_1 \cap C_2$. Suppose that $C_1 \cap \operatorname{int} C_2 \neq \emptyset$. Then*

$$N(C_1 \cap C_2; x) = N(C_1; x) + N(C_2; x).$$

Proof. Applying the subdifferential sum rule to the indicator functions of C_1 and C_2. ●

The condition (1.2.6) is often referred to as a constraint qualification. Without it the equality in the convex subdifferential sum rule may not hold (Exercise **??**).

1.2.4 Role in Convex Programming

Subdifferential plays important roles in convex programming. First for unconstrained convex minimization problem we have Fermat's rule:

Proposition 1.2.11 (Subdifferential at Optimality) *Let X be a Banach space and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a proper convex function. Then the point $\bar{x} \in X$ is a (global) minimizer of f if and only if the condition $0 \in \partial f(\bar{x})$ holds.*

Proof. We only need to observe that $\bar{x} \in X$ is a minimizer of f if and only if

$$f(x) - f(\bar{x}) \geq 0 = \langle 0, x - \bar{x} \rangle,$$

which by definition is equivalent to $0 \in \partial f(\bar{x})$. ●

Alternatively put, minimizers of f correspond exactly to "zeroes" of ∂f.
Consider the constrained convex optimization problem of

$$\mathcal{CP} \qquad \begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in C \subset X, \end{array} \qquad (1.2.13)$$

where C is a closed convex subset of X and $f\colon X \to \mathbb{R} \cup \{+\infty\}$ is a convex lower semicontinuous function. Combining the Fermat's rule with the subdifferential sum rule we derive a characterization for solutions to $\mathcal{CP}$.

Theorem 1.2.12 (Pshenichnii–Rockafellar Conditions) *Let C be a closed convex subset of $\mathbb{R}^N$ and let $f\colon X \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Suppose that $C \cap \operatorname{cont} f \neq \emptyset$ and f is bounded below on C. Then $\bar{x}$ is a solution of $\mathcal{CP}$ if and only if it satisfies*

$$0 \in \partial f(\bar{x}) + N(C; \bar{x}).$$

Proof. Apply the convex subdifferential sum rule of Theorem 1.2.9 to $f + \iota_C$ at $\bar{x}$. ●

Finally we turn to the relationship between subdifferential of optimal value functions in convex programming and Lagrange multipliers. We shall see from the two versions of Lagrange multiplier rules given below, the subdifferential of the optimal value function completely characterizes the set of Lagrange multipliers (denoted λ in these theorems).

Theorem 1.2.13 (Lagrange Multiplier without Existence of Optimal Solution) *Let $v(y, z)$ be the optimal value function of the constrained optimization problem $P(y, z)$. Then $-\lambda \in \partial v(0, 0)$ if and only if*

(i) (nonnegativity) $\lambda \in K^+ \times Z^*$; *and*
(ii) (unconstrained optimum) *for any $x \in C$,*

$$f(x) + \langle \lambda, (g(x), h(x)) \rangle \geq v(0, 0).$$

Proof. (a) The "only if" part. Suppose that $-\lambda \in \partial v(0,0)$. It is easy to see that $v(y,0)$ is non-increasing with respect to the partial order $\leq_K$. Thus, for any $y \in K$,

$$0 \geq v(y,0) - v(0,0) \geq \langle -\lambda, (y,0) \rangle$$

so that $\lambda \in K^+ \times Z^*$. Conclusion (ii) follows from the fact that for all $x \in C$,

$$f(x) + \langle \lambda, (g(x), h(x)) \rangle \geq v(g(x), h(x)) + \langle \lambda, (g(x), h(x)) \rangle \geq v(0,0). \quad (1.2.14)$$

(b) The "if" part. Suppose λ satisfies conditions (i) and (ii). Then we have, for any $x \in C$, $g(x) \leq_K y$ and $h(x) = z$,

$$f(x) + \langle \lambda, (y,z) \rangle \geq f(x) + \langle \lambda, (g(x), h(x)) \rangle \geq v(0,0). \quad (1.2.15)$$

Taking the infimum of the leftmost term under the constraints $x \in C$, $g(x) \leq_K y$ and $h(x) = z$, we arrive at

$$v(y,z) + \langle \lambda, (y,z) \rangle \geq v(0,0). \quad (1.2.16)$$

Therefore, $-\lambda \in \partial v(0,0)$. ●

If we denote by $\Lambda(y,z)$ the multipliers satisfying (i) and (ii) of Theorem 1.2.13 then we may write the useful set equality

$$\Lambda(0,0) = -\partial v(0,0).$$

The next corollary is now immediate. It is often a useful variant since h may well be affine.

Corollary 1.2.14 (Lagrange Multiplier without Existence of Optimal Solution) *Let $v(y,z)$ be the optimal value function of the constrained optimization problem $P(y,z)$. Then $-\lambda \in \partial v(0,0)$ if and only if*

(i) (nonnegativity) $\lambda \in K^+ \times Z^*$;
(ii) (unconstrained optimum) *for any $x \in C$, satisfying $g(x) \leq_K y$ and $h(x) = z$,*

$$f(x) + \langle \lambda, (y,z) \rangle \geq v(0,0).$$

When an optimal solution for the problem $P(0,0)$ exists, we can also derive a so called *complementary slackness* condition.

Theorem 1.2.15 (Lagrange Multiplier when Optimal Solution Exists) *Let $v(y,z)$ be the optimal value function of the constrained optimization problem $P(y,z)$. Then the pair $(\bar{x}, \lambda)$ satisfies $-\lambda \in \partial v(0,0)$ and $\bar{x} \in S(0,0)$ if and only if*

(i) (nonnegativity) $\lambda \in K^+ \times Z^*$;
(ii) (unconstrained optimum) *the function*

$$x \mapsto f(x) + \langle \lambda, (g(x), h(x)) \rangle$$

attains its minimum over C at $\bar{x}$;
(iii) (complementary slackness) $\langle \lambda, (g(\bar{x}), h(\bar{x})) \rangle = 0$.

Proof. (a) The "only if" part. Suppose that $\bar{x} \in S(0,0)$ and $-\lambda \in \partial v(0,0)$. As in the proof of Theorem 1.2.13 we can show that $\lambda \in K^+ \times Z^*$. By the definition of the subdifferential and the fact that $v(g(\bar{x}), h(\bar{x})) = v(0,0)$, we have

$$0 = v(g(\bar{x}), h(\bar{x})) - v(0,0) \geq \langle -\lambda, (g(\bar{x}), h(\bar{x}))\rangle \geq 0,$$

so that the complementary slackness condition $\langle \lambda, (g(\bar{x}), h(\bar{x}))\rangle = 0$ holds.

Observing that $v(0,0) = f(\bar{x}) + \langle \lambda, (g(\bar{x}), h(\bar{x}))\rangle$, the strengthened unconstrained optimal condition follows directly from that of Theorem 1.2.13.

(b) The "if" part. Let λ and $\bar{x}$ satisfy conditions (i), (ii) and (iii). Then, for any $x \in C$ satisfying $g(x) \leq_K 0$ and $h(x) = 0$,

$$\begin{aligned} f(x) &\geq f(x) + \langle \lambda, (g(x), h(x))\rangle \\ &\geq f(\bar{x}) + \langle \lambda, (g(\bar{x}), h(\bar{x}))\rangle = f(\bar{x}). \end{aligned} \tag{1.2.17}$$

That is to say $\bar{x} \in S(0,0)$.

Moreover, for any $g(x) \leq_K y, h(x) = z$, $f(x) + \langle \lambda, (y,z)\rangle \geq f(x) + \langle \lambda, (g(x), h(x))\rangle$. Since $v(0,0) = f(\bar{x})$, by (1.2.17) we have

$$f(x) + \langle \lambda, (y,z)\rangle \geq f(\bar{x}) = v(0,0). \tag{1.2.18}$$

Taking the infimum on the left hand side of (1.2.18) yields

$$v(y,z) + \langle \lambda, (y,z)\rangle \geq v(0,0),$$

which is to say, $-\lambda \in \partial v(0,0)$. ●

We can deduce from Theorems 1.2.13 and 1.2.15 that $\partial v(0,0)$ completely characterizes the set of Lagrange multipliers.

1.3 Fenchel Conjugate

Obtaining Lagrange multipliers by using the convex subdifferential is closely related to convex duality theory based on the concept of conjugate functions introduced by Fenchel.

1.3.1 The Fenchel Conjugate

The *Fenchel conjugate* of a function (not necessarily convex) $f\colon X \to [-\infty, +\infty]$ is the function $f^*\colon X^* \to [-\infty, +\infty]$ defined by

$$f^*(x^*) = \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\}.$$

The operation $f \to f^*$ is also called a Fenchel-Legendre transform. The function f^* is convex and if the domain of f is nonempty then f^* never takes the value $-\infty$. Clearly the conjugate operation is *order-reversing*: for functions $f, g\colon X \to [-\infty, +\infty]$, the inequality $f \geq g$ implies $f^* \leq g^*$.

1.3.2 The Fenchel–Young Inequality

This is an elementary but important result that relates conjugate operation with the subgradient.

Proposition 1.3.1 (Fenchel–Young Inequality) *Let* $f\colon X \to \mathbb{R}\cup\{+\infty\}$ *be a convex function. Suppose that* $x^* \in X^*$ *and* $x \in \operatorname{dom} f$*. Then*

$$f(x) + f^*(x^*) \geq \langle x^*, x\rangle. \tag{1.3.1}$$

Equality holds if and only if $x^* \in \partial f(x)$*.*

Proof. The inequality (1.3.1) follows directly from the definition. We have the equality

$$f(x) + f^*(x^*) = \langle x^*, x\rangle,$$

if and only if, for any $y \in X$,

$$f(x) + \langle x^*, y\rangle - f(y) \leq \langle x^*, x\rangle.$$

That is

$$f(y) - f(x) \geq \langle x^*, y - x\rangle,$$

or $x^* \in \partial f(x)$. ●

Remark 1.3.2 When f is differentiable, taking derivative with respect to x in the Fenchel equality we have $x^* = f'(x)$. Then the Fenchel-Legendre transform has the following explicit form as a function of x

$$f^*(f'(x)) = \langle x, f'(x)\rangle - f(x).$$

In Chapter 4 we will see that when f is the price of a contingent claim as a function of a forward price x, the Fenchel -Legendre transform is related to the delta hedging . Its derivative is also relevant when we deal with dynamical hedging. We can directly verify the following representation of the derivative of the Fenchel -Legendre transform

$$D_x f^*(f'(x)) = D_x\langle x, f'(x)\rangle - f'(x) = [D_x, f'(x)I]x,$$

where D_x is the differential operator with respect to x, I is the identity operator and $[A, B] = AB - BA$ represents the commutator of operator A and B. Symmetrically we also have

$$D_{x^*} f((f^*)'(x^*)) = [D_{x^*}, (f^*)'(x^*)I]x^*.$$

We can consider the conjugate of f^* called the *biconjugate* of f and denoted f^{**}. This is a function on X^{**}. When X is a reflexive Banach space, i.e. $X = X^{**}$ it follows from the Fenchel–Young inequality (1.3.1) that $f^{**} \leq f$. The function f^{**} is the largest among all the convex function dominated by f and is called the convex hull of f. Many important convex functions f on $X = \mathbb{R}^N$ equal to their biconjugate

f^{**}. Such functions thus occur as natural pairs, f and f^*. Table 1.1 shows some elegant examples on $\mathbb{R}$ and Table 1.2 describes some simple transformations of these examples. Checking the calculation in Table 1.1 and verifying the formulas in Table 1.2 are good exercises to get familiar with concept of conjugate functions.

Note that the first four functions in Table 1.1 are special cases of indicator functions on $\mathbb{R}$. A more general result is:

Example 1.3.3 Let C be a closed convex set in the reflexive Banach spacce X. Then $\iota_C^* = \sigma_C$ and $\sigma_C^* = \iota_C$.

$f(x) = g^*(x)$	$\operatorname{dom} f$	$g(y) = f^*(y)$	$\operatorname{dom} g$
0	$\mathbb{R}$	0	$\{0\}$
0	$\mathbb{R}_+$	0	$-\mathbb{R}_+$
0	$[-1,1]$	$\|y\|$	$\mathbb{R}$
0	$[0,1]$	y^+	$\mathbb{R}$
$\|x\|^p/p, \quad p>1$	$\mathbb{R}$	$\|y\|^q/q \quad (\frac{1}{p}+\frac{1}{q}=1)$	$\mathbb{R}$
$\|x\|^p/p, \quad p>1$	$\mathbb{R}_+$	$\|y^+\|^q/q \quad (\frac{1}{p}+\frac{1}{q}=1)$	$\mathbb{R}$
$-x^p/p, \quad 0<p<1$	$\mathbb{R}_+$	$-(-y)^q/q \quad (\frac{1}{p}+\frac{1}{q}=1)$	$-\operatorname{int}\mathbb{R}_+$
$-\log x$	$\operatorname{int}\mathbb{R}_+$	$-1-\log(-y)$	$-\operatorname{int}\mathbb{R}_+$
e^x	$\mathbb{R}$	$\begin{cases} y\log y - y & (y>0) \\ 0 & (y=0) \end{cases}$	$\mathbb{R}_+$

Table 1.1. Conjugate pairs of convex functions on $\mathbb{R}$.

Combining Fenchel–Young inequality and the sandwich theorem we can show that $f^{**} = f$ for convex lsc function f.

Theorem 1.3.4 (Biconjugate) *Let X be a finite dimensional Banach space. Then $f^{**} \le f$ in $\operatorname{dom} f$ and equality holds at point $x \in \operatorname{int}\operatorname{dom} f$.*

Proof. It is easy to check $f^{**} \le f$ and we leave it as an exercise. For any $\bar{x} \in \operatorname{int}\operatorname{dom} f$, $\partial f(\bar{x}) \neq \emptyset$. Let $x^* \in \partial f(\bar{x})$. By the Fenchel–Young inequality we have

$f = g^*$	$g = f^*$
$f(x)$	$g(y)$
$h(ax) \quad (a \neq 0)$	$h^*(y/a)$
$h(x+b)$	$h^*(y) - by$
$ah(x) \quad (a > 0)$	$ah^*(y/a)$

Table 1.2. Transformed conjugates.

$$f(\bar{x}) = \langle x^*, \bar{x} \rangle - f^*(x^*) \leq \sup_{y^*}[\langle y^*, \bar{x} \rangle - f^*(y^*)] = f^{**}(\bar{x}) \leq f(\bar{x}).$$

●

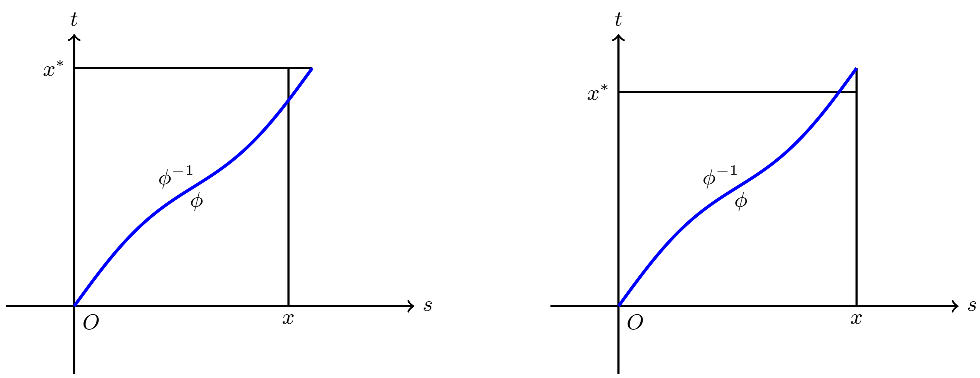


Fig. 1.1. Fenchel-Young inequality

1.3.3 Graphic Illustration and Generalizations

For increasing function ϕ, $\phi(0) = 0$, $f(x) = \int_0^x \phi(s)ds$ is convex and $f^*(x^*) = \int_0^{x^*} \phi^{-1}(t)dt$. Graphs Fig. 1.1 illustrate the Fenchel-Young inequality graphically. The additional areas enclosed by the graph of ϕ^{-1}, $s = x$ and $t = x^*$ or that of ϕ, $s = x$ and $t = x^*$ beyond the area of the rectangle $[0, x] \times [0, x^*]$ generates the additional area that leads to a strict inequality. We also see that equality holds when $x^* = \phi(x) = f'(x)$ and $x = \phi^{-1}(x^*) = (f^*)'(x^*)$ in Fig. 1.2.

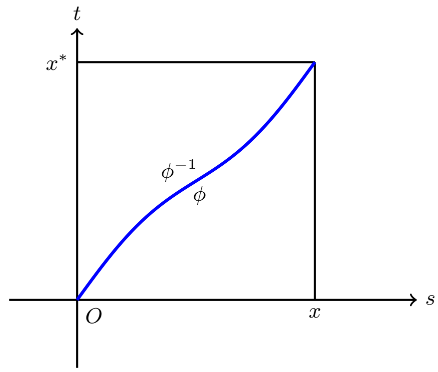


Fig. 1.2. Fenchel-Young equality

1.4 Convex Duality Theory

Using the Fenchel–Young inequality for each constrained optimization problem we can write its companion *dual problem*. There are several different but equivalent perspectives.

1.4.1 Rockafellar Duality

We start with the Rockafellar formulation of bi-conjugate. It is very general and — as we shall see — other perspectives can easily be written as its special cases.

Consider a two-variable function $F(x, y)$ on $X \times Y$ where X, Y are Banach spaces. Treating y as a parameter, consider the parameterized optimization problem

$$v(y) = \inf_x F(x, y). \tag{1.4.1}$$

Our associated *primal* optimization problem[1] is

$$p = v(0) = \inf_{x \in X} F(x, 0) \tag{1.4.2}$$

and the *dual* problem is

$$d = v^{**}(0) = \sup_{y^* \in Y^*} -F^*(0, -y^*). \tag{1.4.3}$$

Since v dominates v^{**} as the Fenchel-Young inequality establishes, we have

$$v(0) = p \geq d = v^{**}(0).$$

This is called *weak duality* and the non-negative number $p - d = v(0) - v^{**}(0)$ is called the *duality gap* — which we aspire to be small or zero.

[1] The use of the term 'primal' is much more recent than the term 'dual' and was suggested by George Dantzig's father Tobias when linear programming was being developed in the 1940's.

Let $F(x,(y,z)) := f(x) + \iota_{\mathrm{epi}(g)}(x,y) + \iota_{\mathrm{graph}(h)}(x,z)$. Then problem $P(y,z)$ in (1.1.1) becomes problem (1.4.1) with parameters (y,z). On the other hand, we can rewrite (1.4.1) as

$$v(y) = \inf_x \{F(x,u) : u = y\}$$

which is problem $P(0,y)$ with $x = (x,u)$, $C = X \times Y$, $f(x,u) = F(x,u)$, $h(x,u) = u$ and $g(x,u) = 0$. So where we start is a matter of taste and predisposition.

Theorem 1.4.1 (Duality and Lagrange Multipliers) *The followings are equivalent:*

(i) *the primal problem has a Lagrange multiplier λ.*
(ii) *there is no duality gap, i.e. $d = p$ is finite and the dual problem has solution $-\lambda$.*

Proof. If the primal problem has a Lagrange multiplier λ then $-\lambda \in \partial v(0)$. By the Fenchel-Young equality

$$v(0) + v^*(-\lambda) = \langle -\lambda, 0\rangle = 0.$$

Direct calculation yields

$$\begin{aligned} v^*(-\lambda) &= \sup_y \{\langle -\lambda, y\rangle - v(y)\} \\ &= \sup_{y,x} \{\langle -\lambda, y\rangle - F(x,y)\} = F^*(0,-\lambda). \end{aligned}$$

Since

$$-F^*(0,-\lambda) \le v^{**}(0) \le v(0) = -v^*(-\lambda) = -F^*(0,-\lambda), \tag{1.4.4}$$

λ is a solution to the dual problem and $p = v(0) = v^{**}(0) = d$.

On the other hand, if $v^{**}(0) = v(0)$ and λ is a solution to the dual problem then all the quantities in (1.4.4) are equal. In particular,

$$v(0) + v^*(-\lambda) = 0.$$

This implies that $-\lambda \in \partial v(0)$, so that λ is a Lagrange multiplier of the primal problem. ●

Example 1.4.2 (Finite Duality Gap) Consider

$$v(y) = \inf\{|x_2 - 1| : \sqrt{x_1^2 + x_2^2} - x_1 \le y\}.$$

We can easily calculate

$$v(y) = \begin{cases} 0 & y > 0 \\ 1 & y = 0 \\ +\infty & y < 0, \end{cases}$$

and $v^{**}(0) = 0$, i.e. there is a finite duality gap $v(0) - v^{**}(0) = 1$.

In this example neither the primal nor the dual problem has a Lagrange multiplier yet both have solutions. Hence, even in two dimensions, existence of a Lagrange multiplier is only a sufficient condition for the dual to attain a solution and is far from necessary. ♢

1.4.2 Fenchel Duality

Let us specify $F(x, y) := f(x) + g(Ax + y)$, where $A : X \to Y$ is a linear operator. We then get the Fenchel formulation of duality. Now the primal problem is

$$p = v(0) = \inf_x [f(x) + g(Ax)]. \tag{1.4.5}$$

To derive the dual problem we calculate

$$F^*(0, -y^*) = \sup_{x,y} [\langle -y^*, y \rangle - f(x) - g(Ax + y)].$$

Letting $u = Ax + y$ we have

$$\begin{aligned} F^*(0, -y^*) &= \sup_{x,u} [\langle -y^*, u - Ax \rangle - f(x) - g(u)] \\ &= \sup_x [\langle y^*, Ax \rangle - f(x)] + \sup_u [\langle -y^*, u \rangle - g(u)] \\ &= f^*(A^* y^*) + g^*(-y^*). \end{aligned}$$

Thus, the dual problem is

$$d = v^{**}(0) = \sup_{y^*} [-f^*(A^* y^*) - g^*(-y^*)]. \tag{1.4.6}$$

If both f and g are convex functions, then it is easy to see that so is

$$v(y) = \inf_x [f(x) + g(Ax + y)].$$

We have already checked that $\operatorname{dom} v = \operatorname{dom} g - A \ \operatorname{dom} f$. Thus, a sufficient condition for the existence of Lagrange multipliers for the primal problem, i.e., $\partial v(0) \neq \emptyset$, is

$$0 \in \operatorname{ri} \ \operatorname{dom} v = \operatorname{ri}[\operatorname{dom} g - A \ \operatorname{dom} f]. \tag{1.4.7}$$

Figure 1.3 illustrates the Fenchel duality theorem for $f(x) := x^2/2 + 1$ and $g(x) = (x - 1)^2/2 + 1/2$. The upper function is f and the lower one is $-g$. The minimum gap occurs at $1/2$ and, which is $7/4$.

Condition (1.4.7) is often referred to as a *constraint qualification* or a *transversality* condition. Enforcing such constraint qualification conditions we can write Theorem 1.4.1 in the following form:

Theorem 1.4.3 (Strong Duality) *If the lower semicontinuous convex functions f, g and the linear operator A satisfy the constraint qualification conditions (1.4.7) then there is a zero duality gap between the primal and dual problems, (1.4.5) and (1.4.6), and the dual problem has a solution.*

A really illustrative example is the application to entropy optimization.

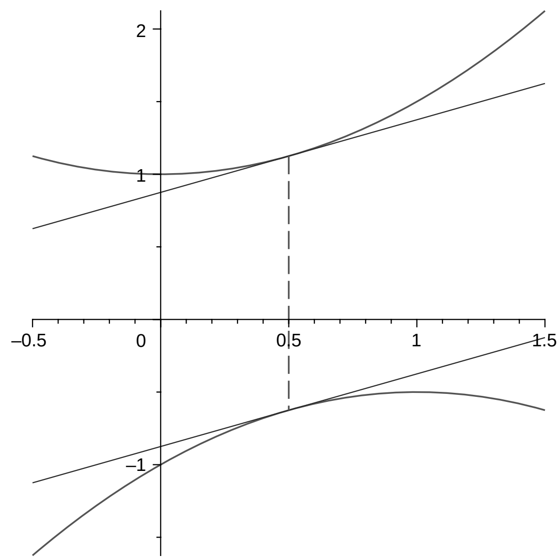


Fig. 1.3. The Fenchel duality sandwich

Example 1.4.4 (Entropy Optimization Problem) Entropy maximization refers to

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & Ax = b \in R^N, \end{aligned} \tag{1.4.8}$$

with the lower semicontinuous convex function f defined on a Banach space of signals, emulating the negative of an entropy and A emulating a finite number of continuous linear constraints representing conditions on some given *moments*. A wide variety of applications can be covered by this model due to its physical relevance.

Applying Theorem 1.4.3 with $g = \iota_{\{b\}}$ we have if $b \in \mathrm{ri}(A \ \operatorname{dom} f)$ then

$$\inf_{x \in X} \{f(x) \mid Ax = b\} = \max_{\phi \in R^N} \{\langle \phi, b \rangle - f^*(A^*\phi)\} = (f^* \circ A^*)^*(b). \tag{1.4.9}$$

When $N < \dim X$ (often infinite) the dual problem is typically much easier to solve then the primal. ♢

Example 1.4.5 (Boltzmann–Shannon Entropy in Euclidean Space) Let

$$f(x) := \sum_{n=1}^{N} p(x_n), \tag{1.4.10}$$

where

$$p(t) := \begin{cases} t \ln t - t & \text{if } t > 0, \\ 0 & \text{if } t = 0, \\ +\infty & \text{if } t < 0. \end{cases}$$

The functions p and f defined above are (negatives of) Boltzmann-Shannon entropy functions on R and R^N, respectively. For $c \in R^N$, $b \in R^M$ and linear mapping $A : R^N \to R^M$ consider the entropy optimization problem

$$\text{minimize } \{f(x) + \langle c, x\rangle : Ax = b\}. \tag{1.4.11}$$

Example 1.4.4 can help us conveniently derive an explicit formula for solutions of (1.4.11) in terms of the solution to its dual problem.

First we note that the sublevel sets of the objective function are compact, thus ensuring the existence of solutions to problem (1.4.11). We can also see by direct calculation that the directional derivative of the cost function is $-\infty$ on any boundary point x of $\operatorname{dom} f = R^N_+$, the domain of the cost function, in the direction of $z - x$. Thus, any solution of (1.4.11) must be in the interior of R^N_+. Since the cost function is strictly convex on int (R^N_+), then the solution is unique.

Let us denote this unique solution of (1.4.11) by $\bar{x}$. Then the duality result in Example 1.4.4 implies that

$$\begin{aligned} f(\bar{x}) + \langle c, \bar{x}\rangle &= \inf_{x \in R^N} \{f(x) + \langle c, x\rangle : Ax = b\} \\ &= \max_{\phi \in R^M} \{\langle \phi, b\rangle - (f + c)^*(A^\top \phi)\}. \end{aligned}$$

Now let $\bar{\phi}$ be a solution to the dual problem, i.e., a Lagrange multiplier for the constrained minimization problem (1.4.11). We have

$$f(\bar{x}) + \langle c, \bar{x}\rangle + (f + c)^*(A^\top \bar{\phi}) = \langle \bar{\phi}, b\rangle = \langle \bar{\phi}, A\bar{x}\rangle = \langle A^\top \bar{\phi}, \bar{x}\rangle.$$

It follows from the Fenchel-Young equality that $A^\top \bar{\phi} \in \partial(f + c)(\bar{x})$. Since $\bar{x} \in$ int (R^N_+) where f is differentiable, we have $A^\top \bar{\phi} = f'(\bar{x}) + c$. Explicit computation shows that $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)$ is determined by

$$\bar{x}_n = \exp(A^\top \bar{\phi} - c)_n, n = 1, \ldots, N. \tag{1.4.12}$$

Indeed, we can use the existence of the dual solution to prove that the primal problem has the given solution without direct appeal to compactness — we deduce the existence of the primal from the duality theory. ♢

Remark 1.4.6 In view of Remark 1.2.6, when both f and g are polyhedral functions the constraint qualification condition (1.4.7) simplifies to

$$\operatorname{dom} g \cap A \operatorname{dom} f \neq \emptyset. \tag{1.4.13}$$

This is very useful in dealing with polyhedral cone programming and, in particular, linear programming problems. One can also similarly handle a subset of polyhedral constraints, see [7, 8].

1.4.3 Lagrange Duality

For problem (1.1.1) define the *Lagrangian*

$$L(\lambda, x; (y, z)) = f(x) + \langle \lambda, (g(x) - y, h(x) - z)\rangle.$$

Then

$$\sup_{\lambda \in K^+ \times Z^*} L(\lambda, x; (y,z)) = \begin{cases} f(x) & \text{if } g(x) \leq_K y, h(x) = z \\ +\infty & \text{otherwise.} \end{cases}.$$

Then problem (1.1.1) can be written as

$$p = v(0) = \inf_{x \in C} \sup_{\lambda \in K^+ \times Z^*} L(\lambda, x; 0). \tag{1.4.14}$$

We can calculate

$$\begin{aligned} v^*(-\lambda) &= \sup_{y,z} [\langle -\lambda, (y,z) \rangle - v(y,z)] \\ &= \sup_{y,z} [\langle -\lambda, (y,z) \rangle - \inf_{x \in C} \{ f(x) : g(x) \leq_K y, h(x) = z \}] \\ &= \sup_{x \in C, y, z} \{ \langle -\lambda, (y,z) \rangle - f(x) : g(x) \leq_K y, h(x) = z \}. \end{aligned}$$

Letting $\xi = y - g(x) \in K$ we can rewrite the expression above as

$$\begin{aligned} v^*(-\lambda) &= \sup_{x \in C, \xi \in K} [\langle -\lambda, (g(x), h(x)) \rangle - f(x) + \langle -\lambda, (\xi, 0) \rangle] \\ &= - \inf_{x \in C, \xi \in K} [L(x, \lambda, 0) + \langle \lambda, (\xi, 0) \rangle] \\ &= \begin{cases} -\inf_x L(x, \lambda, 0) & \text{if } \lambda \in K^+ \times Z^* \\ +\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, the dual problem is

$$d = v^{**}(0) = \sup_{\lambda} -v^*(-\lambda) = \sup_{\lambda \in K^+ \times Z^*} \inf_{x \in C} L(\lambda, x; 0). \tag{1.4.15}$$

We can see that the weak duality inequality $v(0) \geq v^{**}(0)$ is simply the familiar fact that

$$\inf \sup \geq \sup \inf .$$

Example 1.4.7 (Classical Linear Programming Duality) Consider a linear programming problem

$$\begin{aligned} \max \quad & \langle c, x \rangle \\ \text{subject to} \quad & Ax \leq b, x \geq 0 \end{aligned} \tag{1.4.16}$$

where $x \in R^N$, $b \in R^M$, A is a $M \times N$ matrix and $\leq = \leq_{\mathbb{R}^M_+}$. Then by the Lagrange duality, the dual problem is

$$\begin{aligned} \min \quad & \langle b, \lambda \rangle \\ \text{subject to} \quad & A^* \lambda \geq c, \lambda \geq 0. \end{aligned} \tag{1.4.17}$$

In fact, we need to deal with the minimizing problem

$$\min[\langle -c, x \rangle : Ax \leq b, x \geq 0] = -\max[\langle c, x \rangle : Ax \leq b, x \geq 0]$$

We write the Lagrangian

$$L(\lambda, x) = \langle -c, x\rangle + \langle \lambda, Ax - b\rangle$$

Then the primal problem is

$$\inf_{x\geq 0} \sup_{\lambda\geq 0} L(\lambda, x).$$

The dual problem is

$$\sup_{\lambda\geq 0} \inf_{x\geq 0} L(\lambda, x).$$

We can see that

$$\inf_{x\geq 0} L(\lambda, x) = \inf_{x\geq 0} \langle -c + A^*\lambda, x\rangle - \langle \lambda, b\rangle = \begin{cases} -\langle \lambda, b\rangle & \text{if } A^*\lambda \geq c \\ +\infty & \text{otherwise.} \end{cases}$$

So we have

$$\begin{aligned} \max[\langle c, x\rangle : Ax \leq b, x \geq 0] &= -\max_{\lambda\geq 0}[-\langle \lambda, b\rangle : A^*\lambda \geq c] \\ &= \min[\langle \lambda, b\rangle : A^*\lambda \geq c, \lambda \geq 0]. \end{aligned}$$

Clearly all the functions involved here are polyhedral. Applying the constraint qualification condition for polyhedral functions we can conclude that if either the primal problem or the dual problem is feasible then there is no duality gap. Moreover, when the common optimal value is finite then both problems have optimal solutions. ♢

The hard work in Example 1.4.7 was hidden in establishing that the constraint qualification (1.4.13) is sufficient, but unlike many applied developments we have *rigorously* recaptured linear programming duality within our framework.

Note that the primal Lagrange multiplier λ is the dual solution and vice versa. Table 1.3 can help us formulating the dual problem.

Primal constraint	Dual variable	P.variable	D.constraint
$Ax \leq b$	$\lambda \geq 0$	$x \geq 0$	$A^*\lambda \geq c$
$Ax = b$	λ free	x free	$A^*\lambda = c$
$Ax \geq b$	$\lambda \leq 0$	$x \leq 0$	$A^*\lambda \leq c$

Table 1.3. Transformed conjugates.

1.4.4 Generalized Fenchel-Young Inequality

Reexamining the graphic representation of the Fenchel–Young inequality we also realize that the underlying inequality relationship remains valid when the area is weighted by a positive 'density' function $K(s,t)$. Thus, we have

Theorem 1.4.8 (Weighted Fenchel-Young Inequality) *Let $K(x,y)$ be a continuous positive function and let ϕ be a continuous increasing function with $\phi(0)=0$. Then*

$$\int_0^x \int_0^{x^*} K(s,t)dtds \le \int_0^x \int_0^{\phi(s)} K(s,t)dtds + \int_0^{x^*} \int_0^{\phi^{-1}(t)} K(s,t)dsdt$$

and equality holds when $x^* = \phi(x)$ and $x = \phi^{-1}(x^*)$.

Proof. If $\phi(x) \ge x^*$ we have

$$\begin{aligned} &\int_0^x \int_0^{\phi(s)} K(s,t)dtds + \int_0^{x^*} \int_0^{\phi^{-1}(t)} K(s,t)dsdt \\ \ge &\int_0^x \int_0^{x^*} K(s,t)dsdt + \int_{\phi^{-1}(x^*)}^x \int_{x^*}^{\phi(s)} K(s,t)dtds \\ \ge &\int_0^x \int_0^{x^*} K(s,t)dtds. \end{aligned} \tag{1.4.18}$$

Otherwise, $\phi(x) < x^*$ and we have

$$\begin{aligned} &\int_0^x \int_0^{\phi(s)} K(s,t)dtds + \int_0^{x^*} \int_0^{\phi^{-1}(t)} K(s,t)dsdt \\ \ge &\int_0^x \int_0^{x^*} K(s,t)dsdt + \int_x^{\phi^{-1}(x^*)} \int_{\phi(s)}^{x^*} K(s,t)dtds \\ \ge &\int_0^x \int_0^{x^*} K(s,t)dtds. \end{aligned} \tag{1.4.19}$$

Clearly equality holds if and only if $\phi(x) = x^*$. ●

The condition $\phi(0)=0$ merely conveniently locates the lower left conner of the graph to the coordinate origin and is clearly not essential. In general we can always shift this corner to any point $(a, \phi(a))$. More substantively, the requirment that ϕ being a continuous increasing function can be relaxed to nondecresing as long as ϕ^{-1} is replaced appropriately by

$$\phi_{\inf}^{-1}(t) = \inf\{s, \phi(s) \ge t\}.$$

Now we can state a more general Fenchel–Young inequality whose proof is an easy exercise.

Theorem 1.4.9 (Weighted Fenchel-Young Inequality) *Let $K(x,y)$ be a bounded essentially positive measurable function and let ϕ be a nondecreasing function. Then*

$$\int_a^x \int_{\phi(a)}^{x^*} K(s,t)dsdt \le \int_a^x \int_{\phi(a)}^{\phi(s)} K(s,t)dtds + \int_{\phi(a)}^{x^*} \int_a^{\phi_{\inf}^{-1}(t)} K(s,t)dsdt$$

with equality attiained when $x^* \in [\phi(x-), \phi(x+)], x \in [\phi_{\inf}^{-1}(x^*-), \phi_{\inf}^{-1}(x^*+)]$.

The above idea can be further pushed in two different directions in the next two sections.

Multidimensional Fenchel–Young Inequality

It is easier to understand and to formulate n-dimensional Fenchel Young inequality starting by re-examining the graphs presented above with a parameterization (ϕ_1, ϕ_2) of the graph of ϕ in Fig. 1.4–1.5.

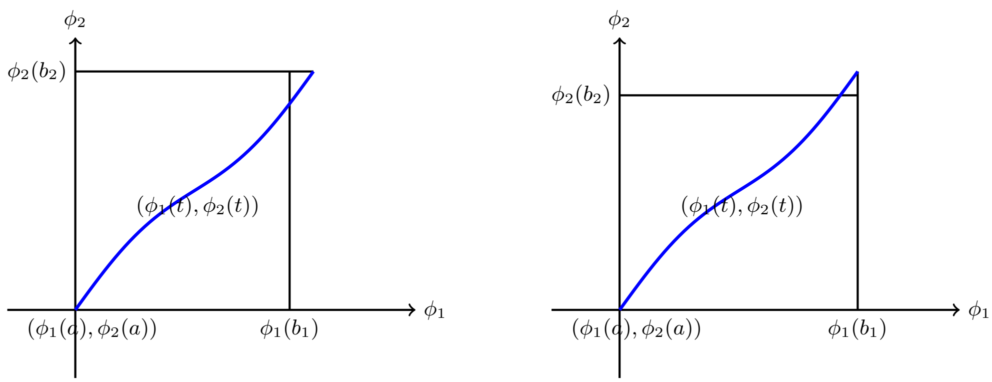


Fig. 1.4. Fenchel-Young inequality

Let $K(s_1, s_2)$ be a nonnegative function and let ϕ_1, ϕ_2 be increasing functions. To avoid technical complication we assume that ϕ_1, ϕ_2 are invertible. Then we can rewrite the Fenchel-Young inequality as

$$\int_{\phi_2(a)}^{\phi_2(b_2)} \int_{\phi_1(a)}^{\phi_1(b_1)} K(s_1,s_2)ds_1ds_2 \tag{1.4.20}$$

$$\le \int_{\phi_1(a)}^{\phi_1(b_1)} \int_{\phi_2(a)}^{\phi_2(\phi_1^{-1}(s_1))} K(s_1,s_2)ds_2ds_1 + \int_{\phi_2(a)}^{\phi_2(b_2)} \int_{\phi_1(a)}^{\phi_1(\phi_2^{-1}(s_2))} K(s_1,s_2)ds_1ds_2$$

with equality attained when $b_1 = b_2$.

This form of the Fenchel-Young inequality can easily be generalized to N-dimention with an induction argument. We will use the following vector notation: $s^N = (s_1, \ldots, s_N), 1^N = (1, 1, \ldots, 1)$ and

$$s_n^N = (s_1, \ldots, s_{n-1}, s_{n+1}, \ldots, s_N).$$

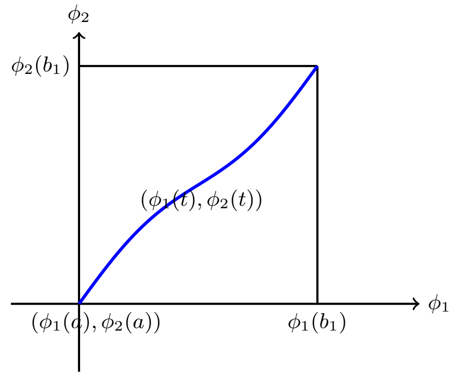


Fig. 1.5. Fenchel-Young inequality

When $\phi^N = (\phi_1, \ldots, \phi_N)$ is a vector valued function we define

$$\phi^N(s^N) = (\phi_1(s_1), \ldots, \phi_N(s_N)).$$

Similarly,

$$\int_{\phi^N(a^N)}^{\phi^N(b^N)} K(s^N)ds^N = \int_{\phi_1(a_1)}^{\phi_N(b_1)} \ldots \int_{\phi_N(a_N)}^{\phi_N(b_N)} K(s_1, \ldots, s_N)ds_N \ldots ds_1.$$

Now we can state and prove the multidimensional Fenchel–Young inequality.

Theorem 1.4.10 (Multidimensional Generalized Fenchel-Young Inequality) *Let $K : \mathbb{R}^N \mapsto \mathbb{R}$ be a nonnegative function and let ϕ^N be a vector function with all the components increasing and invertible. We have*

$$\int_{\phi^N(a \cdot 1^N)}^{\phi^N(b^N)} K(s^N)ds^N \leq \sum_{n=1}^{N} \int_{\phi_n(a)}^{\phi_n(b_n)} \int_{\phi_n^N(a \cdot 1^{N-1})}^{\phi_n^N(\phi_n^{-1}(s_n) \cdot 1^{N-1})} K(s^N)ds_n^N ds_n \quad (1.4.21)$$

with equality attained when $b_1 = b_2 = \ldots = b_N$.

Proof. We prove by induction. The case $N = 2$ has already been established. We focus on the induction step. By separating the integration with respect to ds_{N+1}, we can write the left hand side of the inequality as

$$\begin{aligned} LHS &= \int_{\phi^{N+1}(a \cdot 1^{N+1})}^{\phi^N(b^{N+1})} K(s^{N+1})ds^{N+1} \\ &= \int_{\phi_{N+1}(a)}^{\phi_{N+1}(b_{N+1})} \int_{\phi^N(a \cdot 1^N)}^{\phi^N(b^N)} K(s^{N+1})ds^N ds_{N+1} \end{aligned}$$

Applying the induction hypothesis to the inner layer of the integration we have

$$
\begin{aligned}
LHS &\le \int_{\phi_{N+1}(a)}^{\phi_{N+1}(b_{N+1})} \sum_{n=1}^{N} \int_{\phi_n(a)}^{\phi_n(b_n)} \int_{\phi_n^N(a\cdot 1^{N-1})}^{\phi_n^N(\phi_n^{-1}(s_n)\cdot 1^{N-1})} K(s^{N+1}) ds_n^N ds_n ds_{N+1} \\
&= \sum_{n=1}^{N} \int_{\phi_{N+1}(a)}^{\phi_{N+1}(b_{N+1})} \int_{\phi_n(a)}^{\phi_n(b_n)} \left(\int_{\phi_n^N(a\cdot 1^{N-1})}^{\phi_n^N(\phi_n^{-1}(s_n)\cdot 1^{N-1})} K(s^{N+1}) ds_n^N \right) ds_n ds_{N+1}.
\end{aligned}
$$

The last equality groups the two out layers of the integration together. Now Applying the Fenchel-Young inequality with $N = 2$ to get

$$
\begin{aligned}
LHS \le& \sum_{n=1}^{N} \int_{\phi_n(a)}^{\phi_n(b_n)} \int_{\phi_{N+1}(a)}^{\phi_{N+1}(\phi_n^{-1}(s_n))} \int_{\phi_n^N(a\cdot 1^{N-1})}^{\phi_n^N(\phi_n^{-1}(s_n)\cdot 1^{N-1})} K(s^{N+1}) ds_n^N ds_{N+1} ds_n \\
&+ \int_{\phi_{N+1}(a)}^{\phi_{N+1}(b_{N+1})} \sum_{n=1}^{N} \int_{\phi_n(a)}^{\phi_n(\phi_{N+1}^{-1}(s_{N+1}))} \int_{\phi_n^N(a\cdot 1^{N-1})}^{\phi_n^N(\phi_n^{-1}(s_n)\cdot 1^{N-1})} K(s^{N+1}) ds_n^N ds_n ds_{N+1}
\end{aligned}
$$

Combining the inner layers of the integration in the first sum and applying the equality part of the induction hypothesis for the second sum we arrive at

$$
\begin{aligned}
LHS &\le \sum_{n=1}^{N} \int_{\phi_n(a)}^{\phi_n(b_n)} \int_{\phi_n^{N+1}(a\cdot 1^N)}^{\phi_n^{N+1}(\phi_n^{-1}(s_n)\cdot 1^N)} K(s^{N+1}) ds_n^{N+1} ds_n \\
&+ \int_{\phi_{N+1}(a)}^{\phi_{N+1}(b_{N+1})} \int_{\phi_n^N(a\cdot 1^N)}^{\phi^N(\phi_{N+1}^{-1}(s_{N+1})\cdot 1^N)} K(s^{N+1}) ds^N ds_{N+1} \\
&= \sum_{n=1}^{N+1} \int_{\phi_n(a)}^{\phi_n(b_n)} \int_{\phi_n^{N+1}(a\cdot 1^N)}^{\phi_n^{N+1}(\phi_n^{-1}(s_n)\cdot 1^N)} K(s^{N+1}) ds_n^{N+1} ds_n = RHS.
\end{aligned}
$$

●

A three dimensional graphical illuatration of the multidimentional Fenchel–Young inequality is presented in Fig. 1.6. In this figure we illustrate the simple case where $K(s_1, s_2, s_3) = 1$ so that the left hand side of the inequality (1.4.21) is the volume of a rectangular region. We set $(\phi_1(t), \phi_2(t), \phi_3(t)) = (t, t^2, t)$, $(a_1, a_2, a_3) = (0, 0, 0)$ and $(b_1, b_2, b_3) = (0.9, 1, 0.8)$. The light lines are the edges of the rectangular region and the dark lines outlines the boundaries of the three regions corresponding to the three integrals on the right hand side of Fenchel–Young inequality (1.4.21).

Remark 1.4.11 We also have the following alternative form of estimations by changing the way of integration. Let $K(s_1, s_2)$ be a nonnegative function and let ϕ_1, ϕ_2 be non-decreasing functions.

$$
\begin{aligned}
&\int_{\phi_2(a)}^{\phi_2(b_2)} \int_{\phi_1(a)}^{\phi_1(b_1)} K(s_1, s_2) ds_1 ds_2 \\
&\le \int_{\phi_1(a)}^{\phi_1(b_2)} \int_{\phi_2(\phi_1^{-1}(s_1))}^{\phi_2(b_2)} K(s_1, s_2) ds_2 ds_1 + \int_{\phi_2(a)}^{\phi_2(b_1)} \int_{\phi_1(\phi_2^{-1}(s_2))}^{\phi_1(b_1)} K(s_1, s_2) ds_1 ds_2
\end{aligned}
$$

with equality attained when $b_1 = b_2$.

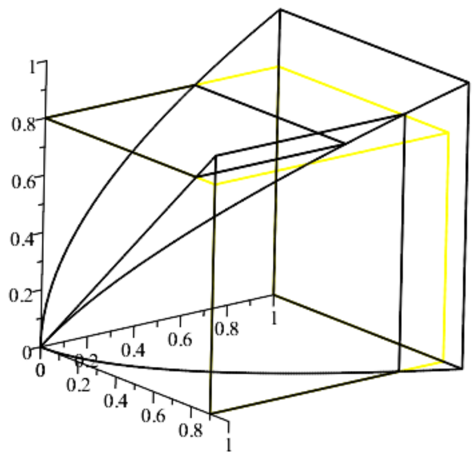


Fig. 1.6. Three Dimensional Fenchel–Young Inequality

1.5 Generalized Convexity, Conjugacy and Duality

Note that the graphic illustrations in section 1.4.3 only works when $x, x^* \in \mathbb{R}$. When, in general, $(x, x^*) \in X \times X^*$ we can immitate the general defintion of the Fenchel conjugate. In such a generalization a nonlinear function $c(x, x^*)$ replaces the role of $\langle x^*, x\rangle$ just as in Theorem 1.4.8 $\int_0^x \int_0^{x^*} K(s,t)dsdt$ replacing the product x^*x. In fact, x^* does not even have to be in X^*. This is a more significant generalization. To implement this idea, one needs to first revise the concept of convexity.

Definition 1.5.1 (Generalized Convexity) *Let Φ be a set of extended real valued functions. We say f is $\Phi-$ convex if*

$$f(x) = \sup\{\phi(x) : \phi \in \Phi, f \geq \phi\}.$$

It is easy to verify that $\Phi-$ convex functions are closed under supremum. Thus, every function has a largest $\Phi-$convex minorant called its $\Phi-$convex hull. Moreover, if f is $\Phi-$ convex then it is coincide with its $\Phi-$convex hull. By setting Φ to be the class of affine functions we get the usual convexity with in the class of lower semicontinuous functions.

Similar to Fenchel conjugate we define:

Definition 1.5.2 (Generalized Fenchel Conjugate) *Let c be a function on $X \times Y$. We define*

$$f^{c(1)}(y) = \sup_x [c(x,y) - f(x)] \text{ and } g^{c(2)}(x) = \sup_y [c(x,y) - g(y)].$$

They are generalizations of Fenchel conjugate. When the function c is not symmetric with respect to its two variables, the $c(1)$ and $c(2)$ conjugate are different. It is easy to see that the generalized Fenchel conjugate also has the order reversing property. Define $\Phi_{c(1)} = \{c(\cdot, y) - b : y \in Y, b \in \mathbb{R}\}$ and $\Phi_{c(2)} = \{c(x, \cdot) - b : x \in X, b \in \mathbb{R}\}$. Then $f^{c(1)}$ is $\Phi_{c(2)}-$convex and $g^{c(2)}$ is $\Phi_{c(1)}-$convex.

Next we discuss some basic properties of generalized Fenchel conjugate.

Theorem 1.5.3 (Fenchel Inequality and Duality) *Let $f : X \to \mathbb{R} \cup \{+\infty\}$ and $g : Y \to \mathbb{R} \cup \{+\infty\}$. Then*

(i) (Fenchel inequality) $f^{c(1)}(y) \geq c(x,y) - f(x), g^{c(2)}(x) \geq c(x,y) - g(y)$,
(ii) (Convex hull) The $\Phi_{c(1)(c(2))}-$convex hull of $f(g)$ is $f^{c(1)c(2)}(g^{c(2)c(1)})$,
(iii) (Duality) $f^{c(1)} = f^{c(1)c(2)c(1)}, g^{c(2)} = g^{c(2)c(1)c(2)}$.

Proof. (i) follows directly from the definitions.

To prove (ii) we observe that by (i) $f(x) \geq c(x,y) - f^{c(1)}(y)$. Taking sup over y we get $f \geq f^{c(1)c(2)}$. On the other hand if for some y, b, $f(x) \geq c(x,y) - b$ for all x, then $b \geq c(x,y) - f(x)$. Taking sup over x we have $b \geq f^{c(1)}(y)$. Thus,

$$f(x) \geq f^{c(1)c(2)}(x) \geq c(x,y) - f^{c(1)}(y) \geq c(x,y) - b$$

establishing $f^{c(1)c(2)}$ as the largest $\Phi_{c(1)}-$ convex function dominated by f. The proof that $g^{c(2)c(1)}$ is the $\Phi_{c(2)}-$ convex hull of g is similar.

(iii) follows from (ii) since $f^{c(1)}$ is $\Phi_{c(2)}-$ convex and $g^{c(2)}$ is $\Phi_{c(1)}-$ convex. ●

Remark 1.5.4 We see from the discussion about generalized Fenchel conjugate that what is essential in dealing with conjugate operation is the closedness with respect to the sup operation. For simple convexity the key link is that a convex function is the sup of all the affine functions it dominates. It is a fact based on the fundamental convex separation theorem.

The generalized convexity can characterize many class of functions. The followings are a few examples showcase the potent of this concept.

Example 1.5.5 Let $\langle\cdot,\cdot\rangle$ be the dual pairing between X and X^*. Define $c(x,x^*) = \ln\langle x,x^*\rangle$, with $\ln t = -\infty$ for $t \le 0$. Then a function $f : X \to \mathbb{R}\cup\{+\infty\}$ is $\Phi_{c(1)}-$ convex if and only if e^f (with the convention $e^{-\infty} = 0$) is sublinear.

Example 1.5.6 Let $X = Y = [0,+\infty]$ and define $c(x,y) = xy$, with the convention $a(+\infty) = +\infty$. Then a function $f : X \to \mathbb{R}\cup\{+\infty\}$ is $\Phi_{c(1)}-$ convex if and only if it is convex and nondecreasing.

Example 1.5.7 Let X be a Hilbert space and $Y = \mathbb{R}_+ \times X$. Define $c(x,(\rho,y)) = -\rho\|x-y\|^2$. Then $f : X \to \mathbb{R}\cup\{+\infty\}$ is $\Phi_{c(1)}-$ convex if and only if it is lower semicontinuous and has a finite minorant $\phi \in \Phi_{c(1)}$.

The concept of subdifferential and its relationship with Fenchel conjugate can also be generalized.

Definition 1.5.8 (Generalized Subdifferential) *Let c be a function on $X \times Y$. We say $y_0(x_0)$ is a $c(1)(c(2))-$subdifferential of $f(g)$ at $x_0(y_0)$ if*

$$f(x) - c(x,y_0)(g(y) - c(x_0,y))$$

attains minimum at $x_0(y_0)$.

Notation $y_0 \in \partial_{c(1)}f(x_0)(x_0 \in \partial_{c(2)}g(y_0))$.

Theorem 1.5.9 (Generalized Fenchel–Young Equality)

(i) (Fenchel equality) $y_0 \in \partial_{c(1)}f(x_0)$ iff $f(x_0) + f^{c(1)}(y_0) = c(x_0,y_0)$.
(ii) (Symmetry) $y_0 \in \partial_{c(1)}f^{c(1)c(2)}(x_0)$ iff $x_0 \in \partial_{c(2)}f^{c(1)}(y_0)$.
(iii)(Φ convexity) $\partial_{c(1)}f(x_0) \ne \emptyset$ implies that f is $\Phi_{c(1)}$ convex at x_0. On the other hand f is $\Phi_{c(1)}$ convex at x_0 implies that $\partial_{c(1)}f(x_0) = \partial_{c(1)}f^{c(1)c(2)}(x_0)$.

Proof. The argument for proving Fenchel equality applies to (i) with $\langle y_0,x_0\rangle$ replaced by $c(x_0,y_0)$. The rest follows from this generalized Fenchel equality. Details are left as an exercise. ●

Similar to the usual subdifferential we have

Theorem 1.5.10 (Cyclical Monotonicity) *Subdifferential $\partial_{c(1)}f$ is $c(1)-$* cyclically monotone *that is for any m pairs of points $y_i \in \partial_{c(1)}f(x_i)$ we have*

$$\begin{aligned}&(c(x_1,y_0) - c(x_0,y_0)) + (c(x_2,y_1) - c(x_1,y_1)) + \\ &\quad \ldots + (c(x_0,y_m) - c(x_m,y_m)) \le 0.\end{aligned}$$

Proof. Adding the following inequalities:

$$f(x_1) - f(x_0) \geq c(x_1, y_0) - c(x_0, y_0)$$
$$f(x_2) - f(x_1) \geq c(x_2, y_1) - c(x_1, y_1)$$
$$\dots \quad \dots$$
$$f(x_0) - f(x_m) \geq c(x_0, y_m) - c(x_m, y_m).$$

and noticing all the terms on the left hand side are cancelled. ●

Next we look at an axiomatic approach to the $c-$conjugate.

Theorem 1.5.11 (Characterization of $c-$conjugate) *Define an operator Δ that maps an extended valued function f on X to an extended valued function Δf on Y. Then Δ is a $c-$conjugate if and only if*

(i) (Duality) $\Delta \inf_\alpha f_\alpha = \sup_\alpha \Delta f_\alpha$
(ii) (Shift reversing) $\Delta(f + d) = \Delta(f) - d,$ *for all* $d \in \mathbb{R}$

where

$$c(x, y) = \Delta(\iota_{\{x\}})(y).$$

Proof. The "if " part: The two properties can be derived from direct computation. For property (i)

$$\begin{aligned}
(\inf f_\alpha)^{c(1)}(y) &= \sup_x [c(x, y) - \inf_\alpha f_\alpha(x)] \\
&= \sup_x \sup_\alpha [c(x, y) - f_\alpha(x)] \\
&= \sup_\alpha \sup_x [c(x, y) - f_\alpha(x)] = \sup_\alpha f_\alpha^{c(1)}(y).
\end{aligned}$$

For property (ii)

$$\begin{aligned}
(f + d)^{c(1)}(y) &= \sup_x [c(x, y) - (f(x) - d)] \\
&= \sup_x [c(x, y) - f(x)] + d = f^{c(1)}(y) + d.
\end{aligned}$$

The "only if" part: The key is the representation

$$f(\cdot) = \inf_x [\iota_{\{x\}}(\cdot) + f(x)].$$

Applying the Δ operator to the above representation we have

$$\begin{aligned}
(\Delta f)(y) &= \Delta \left(\inf_x [\iota_{\{x\}} + f(x)] \right) (y) \\
&= \sup_x \left(\Delta[\iota_{\{x\}} + f(x)] \right) (y) \\
&= \sup_x [\Delta(\iota_{\{x\}})(y) - f(x)] = f^{c(1)}(y)
\end{aligned}$$

where

$$c(x, y) = \Delta(\iota_{\{x\}})(y).$$

●

Rockafellar Duality

Consider the *bi-conjugate* setting again. The primal problem is

$$p = v(0) = \inf_{x \in X} F(x, 0) \tag{1.5.22}$$

as one of the family $v(y) = \inf_x F(x, y)$ on the perturbation space Y. Let Z be the 'dual parameter space' and let $c(y, z)$ be a coupling function. Define the *dual* problem as

$$d = v^{c(1)c(2)}(0) = \sup_{z \in Z}\{c(0, z) - v^{c(1)}(z)\}. \tag{1.5.23}$$

This definition is the same as the Rockafellar duality. However, since now $c(0, z)$ is not necessarily 0 the problem is more involved.

Theorem 1.5.12 (Dual Solution Set) *If* $d = v^{c(1)c(2)}(0) < \infty$ *then the optimal solution set to the dual problem is* $\partial_c v^{c(1)c(2)}(0)$.

Proof. It follows directly from definition and is left as an exercise. ●

Also similar to the Rockafellar duality we have

Theorem 1.5.13 (Weak and Strong Duality) *We always have the weak duality* $d = v^{c(1)c(2)}(0) \leq v(0) = p$. *Equality holds if and only if* v *is* $\Phi_{c(1)}-$*convex at* 0. *In this case if* $d = p$ *is finite then the optimal solution set to the dual problem is* $\partial_{c(1)} v(0)$.

Proof. As before the weak duality follows easily from the Fenchel -Young inequality. To prove strong duality notice that v is $\Phi_{c(1)}-$convex at 0 implies that $\partial_{c(1)} v(0) \neq \emptyset$. Then we can check each element of $\partial_{c(1)} v(0)$ is a solution to the dual problem. ●

Lagrange Duality

Define *Lagrangian* for the primal problem as

$$L(x, z) = c(0, z) - F_x^c(z)$$

where $F_x(y) := F(x, y)$. Then we have the Lagrange form of the primal: If $F_x(y)$ is Φ_c- convex for all $x \in X$ at $y = 0$ then

$$\sup_z L(x, z) = \sup_z\{c(0, z) - F_x^{c(1)}(z)\} = F_x^{c(1)c(2)}(0) = F_x(0) = F(x, 0).$$

Thus, the primal problem becomes

$$\inf_x \sup_z L(x, z).$$

Next we consider the Lagrange form of the dual If $c < +\infty$ we have

$$\begin{aligned}\inf_x L(x,z) &= \inf_x \{c(0,z) - F_x^{c(1)}(z)\}\\ &= \inf_x \{c(0,z) - \sup_y (c(y,z) - F_x(y))\}\\ &= c(0,z) - \sup_y \{c(y,z) - \inf_x F(x,y)\}\\ &= c(0,z) - \sup_y \{c(y,z) - v(y)\} = c(0,z) - v^{c(1)}(z).\end{aligned}$$

Therefore, the dual problem becomes

$$\sup_z \inf_x L(x,z).$$

We see that the primal and dual value equal if and only if

$$\inf_x \sup_z L(x,z) = \sup_z \inf_x L(x,z).$$

2

Financial Models in One Period Economy

Summary. This chapter focuses on financial models in a one period economy with a finite sample space. Mathematically, these models involve only finite dimensional spaces yet they still illustrate the main patterns.

In modeling the behavior of agents in a financial market, we usually use concave utility functions and convex risk measure to characterize their attitude towards risk. These agents are subject to various constraints ranging from the availability of capital, contractual obligation to clients to mandates from regulators. Thus, the theory regarding constrained (convex) optimization discussed in the previous chapter is most relevant. The Lagrange multipliers in such financial models often carry a special financial meaning and are worthy of attention. Moreover, as illustrated in the previous chapter, they also provide the key link between the primal and the dual problems.

2.1 Portfolio

Portfolio theory considers the one period financial model in which transaction can only take place at either the beginning of the period or the end of the period represented by $t = 0$ or 1, respectively. We use probability space $(\Omega, \mathcal{F}, P)$ to represent an economy where the σ-algebra $\mathcal{F}$is generated by finitely many atoms $\mathcal{F} = \sigma(\{B_1, \dots, B_N\})$. We use $RV(\Omega, \mathcal{F}, P)$ to denote the Hilbert space of all $\mathcal{F}$-measurable random variables endowed with the inner product

$$\langle x, y \rangle = \mathbf{E}^P[xy] = \sum_{\omega \in \Omega} x(\omega) y(\omega) P(\omega) = \sum_{i=1}^{N} x(B_i) y(B_i) P(B_i),$$

where $x(B_i)$ and $y(B_i)$ signify the common value of $\mathcal{F}$-measurable random variables x and y on atom B_i, respectively. Elements in $RV(\Omega, \mathcal{F}, P)$ represent the price or payoff of assets. In a one perod economy we may think the sample space is simply consists of the atoms of $\mathcal{F}$. Denoting $\omega_i = B_i$, then $\Omega = \{\omega_1, \dots, \omega_N\}$, $P(\omega_i) = P(B_i)$ and $\mathcal{F}$ contains all subsets of Ω.

A *financial market* is modeled by random vectors $S_t = (S_t^0, S_t^1, \dots, S_t^M), t = 0, 1$ on Ω in which S_t^0 represent the price of a *risk free asset* and for simplicity is assumed to be cash here so that $S_t^0 = 1$ for $t = 0, 1$, and $\hat{S}_t = (S_t^1, \dots, S_t^M)$ represent the prices of *risky assets*at time t. For each asset $i > 0$, we also assume that its price S_0^i is a constant and S_1^i is a $\mathcal{F}$-measurable random variable.

Definition 2.1.1 (Portfolio) *A* portfolio *is a vector* $\Theta = (\theta^0, \theta^1, \dots, \theta^M) \in \mathbb{R}^{M+1}$ *whose* i*th component* θ^i *signifies the share of the* i*th asset (with price at* t *represented by* S_t^i*) in the portfolio. The value of a portfolio* Θ *at time* t *is* $\Theta \cdot S_t$*, where notation "$\cdot$" signifies the dot product in* $\mathbb{R}^{M+1}$.

The question is what is the best portfolio. Since different agents have different preferences there is no unique answer to this question.

2.1.1 Markowitz Portfolio

Markowitz proposed his pioneering portfolio theory in his thesis and later published it in his book [36]. Markowitz consider only risky assets. The idea is that for a fixed expected *return* one should choose portfolios with minimum variation, which serves as a measure for the risk. In general, a portfolio with a higher expected return also accompanied with a higher variation (risk). The tradeoff is left to the individual agent.

Use $\hat{S} = (S^1, \dots, S^M)$ to denote the price process of the risky assets and $\hat{\Theta} = (\theta_1, \dots, \theta_M)$ to denote the portfolio. For a given expected payoff r_0 and an initial wealth w_0 we can formulate the problem as

$$\begin{aligned} \text{minimize} \quad & \mathrm{Var}(\hat{\Theta} \cdot \hat{S}_1) \\ \text{subject to} \quad & \mathbf{E}[\hat{\Theta} \cdot \hat{S}_1] = r_0 \\ & \hat{\Theta} \cdot \hat{S}_0 = w_0. \end{aligned} \tag{2.1.1}$$

Regarding $\hat{S}$ as a row vector of random variables and $\hat{\Theta}$ as a row vector, denoting $\mathbf{E}[\hat{S}_1] = [\mathbf{E}[\hat{S}_1^1], \dots, \mathbf{E}[\hat{S}_1^M]]$,

$$A = \begin{bmatrix} \mathbf{E}[\hat{S}_1] \\ \hat{S}_0 \end{bmatrix}, \text{ and } b = \begin{bmatrix} r_0 \\ w_0 \end{bmatrix},$$

we can rewrite (2.1.1) as an entropy maximization problem

$$\begin{aligned} \text{minimize} \quad & f(x) := \frac{1}{2} x^\top \Sigma x \\ \text{subject to} \quad & Ax = b. \end{aligned} \tag{2.1.2}$$

Here $x = \hat{\Theta}^\top$ and

$$\begin{aligned} \Sigma &= \mathbf{E}[(\hat{S}_1 - \mathbf{E}(\hat{S}_1))^\top (\hat{S}_1 - \mathbf{E}(\hat{S}_1))] \\ &= (\mathbf{E}[(S_1^i - \mathbf{E}(S_1^i))(S_1^j - \mathbf{E}(S_1^j))])_{i,j=1,\dots,M}. \end{aligned} \tag{2.1.3}$$

The coefficient $1/2$ is added to the risk function to make the computation easier. Clearly, Σ is a symmetric positive semidefinite matrix. We will assume that it is in fact positive definite. Then

$$f^*(y) = \frac{1}{2} y^\top \Sigma^{-1} y. \tag{2.1.4}$$

The constraint qualification condition for strong duality here is $b \in \text{range} A$ which is to say (r_0, w_0) is feasible for the constraint. Assuming that this constraint qualification condition is satisfied, it follows from Theorem 1.4.3 on the strong duality that the value of problem (2.1.2) equals to that of its dual:

$$\begin{aligned} \text{maximize} \;\; & b^\top y - \frac{1}{2} y^\top A \Sigma^{-1} A^\top y \\ & = \frac{1}{2} b^\top (A \Sigma^{-1} A^\top)^{-1} b. \end{aligned} \tag{2.1.5}$$

Here the optimal solution to the dual is

$$\bar{y} = (A \Sigma^{-1} A^\top)^{-1} b. \tag{2.1.6}$$

Denote σ the minimum *standard deviation* of portfolios with expected return r_0, we have

$$\sigma^2 = b^\top (A \Sigma^{-1} A^\top)^{-1} b. \tag{2.1.7}$$

Let $\bar{x}$ be the solution of (2.1.2). Decoupling tells us that

$$f(x) + f^*(A^\top y) - \langle y, Ax \rangle = 0 \tag{2.1.8}$$

and

$$g(Ax) + g^*(-y) + \langle y, Ax \rangle = 0 \tag{2.1.9}$$

The equality (2.1.9) is an identity. The equality (2.1.8) via Fenchel equality tells us

$$\bar{x} = (f^*)'(A^\top \bar{y}) = \Sigma^{-1} A^\top \bar{y}.$$

Thus the optimal portfolio is

$$\bar{x} = \Sigma^{-1} A^\top (A \Sigma^{-1} A^\top)^{-1} b. \tag{2.1.10}$$

Define $\alpha = \mathbf{E}[\hat{S}_1] \Sigma^{-1} \mathbf{E}[\hat{S}_1]^\top$, $\beta = \mathbf{E}[\hat{S}_1] \Sigma^{-1} \hat{S}_0^\top$ and $\gamma = \hat{S}_0 \Sigma^{-1} \hat{S}_0^\top$. We have

Theorem 2.1.2 (Markowitz Portfolio Theorem) *For given initial wealth w_0 and expected payoff r_0, the minimum risk in terms of variation σ and the corresponding minimum risk portfolio Θ are determined by*

$$\sigma(r_0, w_0) = \sqrt{\frac{\gamma r_0^2 - 2\beta r_0 w_0 + \alpha w_0^2}{\alpha\gamma - \beta^2}} \tag{2.1.11}$$

and

$$\Theta(r_0, w_0) = \frac{\mathbf{E}(\hat{S}_1)(\gamma r_0 - \beta w_0) + \hat{S}_0(\alpha w_0 - \beta r_0)}{\alpha\gamma - \beta^2} \Sigma^{-1} \tag{2.1.12}$$

Proof. Rewriting (2.1.7) and (2.1.10) in terms of α, β and γ defined above.

●

Note that both $\sigma(r_0, w_0)$ and $\Theta(r_0, w_0)$ are positive homogeneous functions we have

Corollary 2.1.3 *Use μ to denote the expected return on unit initial wealth and let $\sigma = \sigma(\mu, 1)$ and $\Theta = \Theta(\mu, 1)$. Then*

$$\sigma = \sqrt{\frac{\gamma\mu^2 - 2\beta\mu + \alpha}{\alpha\gamma - \beta^2}} \tag{2.1.13}$$

and

$$\Theta = \frac{\mathbf{E}[\hat{S}_1](\gamma\mu - \beta) + \hat{S}_0(\alpha - \beta\mu)}{\alpha\gamma - \beta^2}\Sigma^{-1} \tag{2.1.14}$$

Moreover, $\sigma(\mu w_0, w_0) = w_0\sigma$ and $\Theta(\mu w_0, w_0) = w_0\Theta$.

We now turn to a graphical interpretation of the Markowitz portfolio theory. Note that (2.1.13) also determines μ as a function of σ. Draw this function on the $\sigma\mu$-plan we get the following curve called a *Markowitz bullet* because of its shape. It is also often referred to as the *Markowitz frontier* .

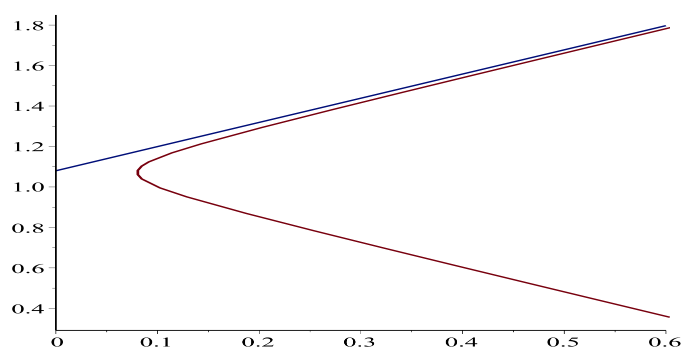


Fig. 2.1. Markowitz Bullet

Every point inside the Markowitz bullet represents a portfolio that can be moved horizontally to the left to a point on the boundary of the bullet. This point on the boundary represents a portfolio with the same expected return but less risk. For every point on the lower half of the boundary of the Markowitz bullet, one can find a corresponding point on the upper half of the boundary with the same variation and a higher expected return. Thus, preferred portfolios are represented by points on the upper boundary of the Markowitz bullet. We note that the upper boundary of the Markowitz bullet has a asymptote whose slope can be determined by

$$\lim_{\sigma\to\infty}\frac{\mu}{\sigma}=\sqrt{\frac{\alpha\gamma-\beta^2}{\gamma}}. \tag{2.1.15}$$

By taking the limit of the tangent line of points on the boundary of the Markowitz bullet one can show that the μ-intercept of this asymptote is at β/γ. This number will play an important role in our discussion of the capital asset pricing model. In fact, the asymptote for the upper boundary of the Markowitz bullet passes through this point.

Although the Markowitz bullet is nonlinear, the Markowitz portfolio is an affine function of the return. This leads to

Theorem 2.1.4 (Two Fund Theorem) *Given two distinct portfolios on the Markowitz bullet, then any portfolio on the Markowitz bullet can be represented as their linear combination.*

Proof. This follows directly from the affine structure of the Markowitz optimal portfolio (2.1.12). ●

Remark 2.1.5 In pointing out that all portfolios on the Markowitz frontier are generated by just two such portfolios, the two fund theorem has great practical significance. One can often use two broad based indices to approximate the two basic generating portfolios for the Markowitz frontier. This can be viewed as a theoretical foundation for the passive investment strategy of buy and hold broad based indices.

If our sole goal is to minimize the risk then our problem becomes

$$\begin{aligned}\text{minimize}\quad & f(x):=\frac{1}{2}x^\top\Sigma x\\ \text{subject to}\quad & \hat{S}_0^\top x=w_0.\end{aligned} \tag{2.1.16}$$

Using a similar argument one can show

Theorem 2.1.6 (Minimum Risk Portfolio) *The* minimum risk portfolio *is*

$$\Theta_{\min}=\gamma^{-1}w_0\hat{S}_0\Sigma^{-1}$$

and its standard deviation is

$$\sigma_{\min}=\gamma^{-1/2}w_0.$$

2.1.2 Capital Asset Pricing Model

Capital asset pricing model (CAPM) works as follows. First it generalizes the Markowitz portfolio theory by allowing risk free asset in the portfolio. It turns out that the optimal portfolios this sense all lies on a straight line in the $\sigma\mu$-plane called the capital market line. Then the model prices a risky asset according to the principle that adding it to the market does not change the capital market line.

We derive the capital market line using convex duality first. Similar to (2.1.1) we now face the problem of

$$\begin{aligned} \text{minimize} \quad & \mathrm{Var}(\Theta \cdot S_1) \\ \text{subject to} \quad & \mathbf{E}[\Theta \cdot S_1] = \mu \\ & \Theta \cdot S_0 = 1. \end{aligned} \tag{2.1.17}$$

Here we standardized the initial wealth to 1 and μ is the expected return. Since $\mathrm{Var}(S_1^0) = 0$ one can show that

$$\mathrm{Var}(\Theta \cdot S_1) = \mathrm{Var}(\hat{\Theta} \cdot \hat{S}_1). \tag{2.1.18}$$

Relation (2.1.18) suggests a strategy of solving problem (2.1.17) in two steps. First, for a portfolio with $\theta = \theta_0 \geq 0$, denote $R = S_1^0/S_0^0$, the return on the risk free asset, we solve problem

$$\begin{aligned} \text{minimize} \quad & \mathrm{Var}(\hat{\Theta} \cdot \hat{S}_1) \\ \text{subject to} \quad & \mathbf{E}[\hat{\Theta} \cdot S_1] = \mu - \theta R \\ & \hat{\Theta} \cdot \hat{S}_0 = 1 - \theta. \end{aligned} \tag{2.1.19}$$

Then, we minimize the minimum variation of (2.1.19) as a function of θ.

By Theorem 2.1.2 the minimum variation corresponding to problem (2.1.19) as a function of θ is determined by

$$\begin{aligned} f(\theta) &= [\sigma(\mu - \theta R, 1 - \theta)]^2 \\ &= \frac{\gamma(\mu - \theta R)^2 - 2\beta(\mu - \theta R)(1 - \theta) + \alpha(1 - \theta)^2}{\alpha\gamma - \beta^2} \end{aligned} \tag{2.1.20}$$

Clearly, the solution of problem (2.1.17) corresponds to the minimum of function f, if it exists. Since f is a quadratic function of θ, the minimum attains at

$$\bar{\theta} = \frac{\alpha - \beta(\mu + R) + \gamma\mu R}{\alpha - 2\beta R + \gamma R^2}, \tag{2.1.21}$$

the solution to the equation $f'(\theta) = 0$. Denote $\Delta := \alpha - 2\beta R + \gamma R^2 > 0$. It is easy to see that the share invested in the risky assets is

$$1 - \bar{\theta} = (\beta - \gamma R)\frac{\mu - R}{\Delta} \tag{2.1.22}$$

We observe that only $\mu > R$ makes sense because by including risky assets we always expect to get a higher return than the risk free assets. Note that the risky assets are involved in the minimum variance portfolio only when $1 - \bar{\theta} > 0$. This implies

$$R < \beta/\gamma \tag{2.1.23}$$

by (2.1.22). Let us focus on the case when R satisfies (2.1.23). We can calculate

$$\mu - \bar{\theta}R = (\alpha - \beta R)\frac{\mu - R}{\Delta}. \tag{2.1.24}$$

By the positive homogeneous property of σ we have

$$\sigma = \sigma(\mu - \bar{\theta}R, 1 - \bar{\theta}) = \sigma(\alpha - \beta R, \beta - \gamma R)\frac{\mu - R}{\Delta}. \tag{2.1.25}$$

It is easy to verify that $\sigma(\alpha - \beta R, \beta - \gamma R) = \sqrt{\Delta}$. Thus, all the optimal portfolios lie on the line

$$\mu = R + \sqrt{\Delta}\sigma. \tag{2.1.26}$$

This line on the $\sigma\mu$-plane is usually referred to as the *capital market line*. This linear structure of the optimal portfolios suggests that we can derive all the optimal portfolios as the linear combinations of two distinct portfolios. Taking the risk free bond and a portfolio of pure risky assets we have the following

Theorem 2.1.7 (Two Fund Separation Theorem) *All the optimal portfolios on the capital market line can be represented as the linear combination of the riskless bond and the* capital market portfolio

$$\Theta_M = \frac{\mathbf{E}[S_1] - RS_0}{\beta - \gamma R}\Sigma^{-1} = \frac{\mathbf{E}[S_1] - RS_0}{(\mathbf{E}[\hat{S}_1] - R\hat{S}_0)\Sigma^{-1}\hat{S}_0^{\top}}\Sigma^{-1}, \tag{2.1.27}$$

whose corresponding coordinates in the $\sigma\mu$-plane is

$$(\sigma_M, \mu_M) = \left(\frac{\sqrt{\Delta}}{\beta - \gamma R}, \frac{\alpha - \beta R}{\beta - \gamma R}\right). \tag{2.1.28}$$

Proof. Clearly the riskless bond is on the capital market line and can be represented in the $\sigma\mu$-plane as $(0, R)$. We now seek a portfolio on the capital market line that contains only risky asset. We denote its coordinates by (σ_M, μ_M). Note such a portfolio corresponding to $\bar{\theta} = 0$. It follows from (2.1.22) that

$$\mu_M = R + \frac{\Delta}{\beta - \gamma R} = \frac{\alpha - \beta R}{\beta - \gamma R}. \tag{2.1.29}$$

Thus, we can find risky part of the capital market portfolio by solving

$$\begin{aligned} &\text{minimize} \quad \mathrm{Var}(\hat{\Theta} \cdot \hat{S}_1) \\ &\text{subject to} \quad \mathbf{E}[\hat{\Theta} \cdot S_1] = \frac{\alpha - \beta R}{\beta - \gamma R} \\ &\qquad\qquad \hat{\Theta} \cdot \hat{S}_0 = 1. \end{aligned} \tag{2.1.30}$$

By Theorem 2.1.2, we derive the optimal portfolio of (2.1.30) to be

$$\hat{\Theta}_M = \frac{\mathbf{E}[\hat{S}_1] - R\hat{S}_0}{\beta - \gamma R}\Sigma^{-1}. \tag{2.1.31}$$

Noting that the weight on the riskless bond is 0 for the capital market portfolio we arrive at the representation in (2.1.27): $\Theta_M = (0, \hat{\Theta}_M)$.

Finally, comparing (2.1.26) and (2.1.29), we derive

$$\sigma_M = \frac{\sqrt{\Delta}}{\beta - \gamma R}. \tag{2.1.32}$$

●

Clearly, the point (σ_M, μ_M) lies on the boundary of the Markowitz bullet. Moreover, since the capital market line represents optimal portfolio, the Markowitz frontier must lie below it. Thus, the capital market line must tangent to the Markowitz frontier at (σ_M, μ_M) (see Fig. 2.2). As a result, if $R \geq \beta/\gamma$, there is no capital market line (see Fig. 2.3), which confirms what has been derived analytically in (2.1.23).

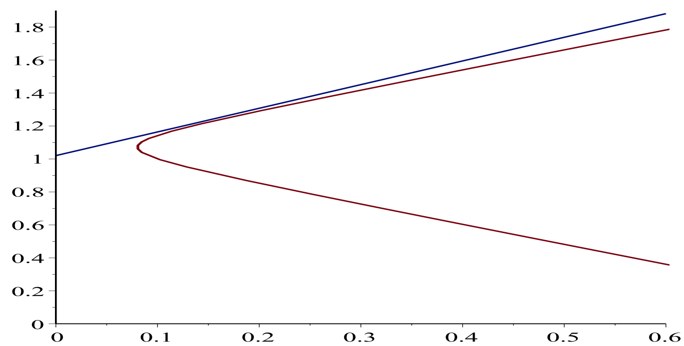


Fig. 2.2. Capital market line

Using the fact that both $(0, R)$ and (σ_M, μ_M) belong to the capital market line we can rewrite the capital market line as

$$\mu = \frac{\mu_M - R}{\sigma_M}\sigma + R. \tag{2.1.33}$$

The theorem below tells us how to use this capital market line to price a risky asset in terms of its expected return.

Theorem 2.1.8 (Capital Asset Pricing Model) *Suppose that we know a financial market S with a riskless bond returning R. Let a^i be a fair priced risky asset with expected percentage return μ_i. Then*

$$\mu_i = R + \beta_i(\mu_M - R). \tag{2.1.34}$$

Here $\beta_i = \sigma_{iM}/\sigma_M^2$ is called the beta *of a^i, where $\sigma_{iM} = cov(a^i, \hat{\Theta}_M \cdot \hat{S}_1)$ is the covariance of a^i and the market portfolio.*

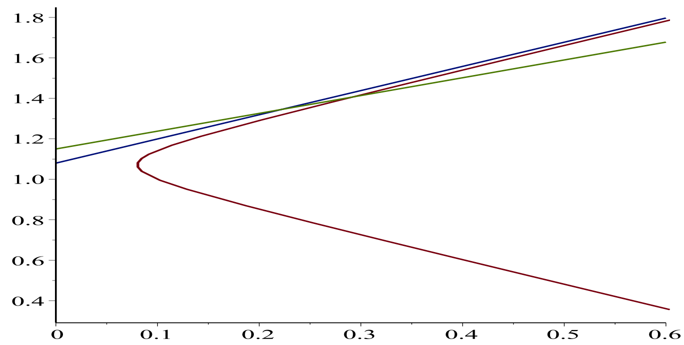


Fig. 2.3. No capital market line

Proof. Consider a portfolio relies on the parameter α that consists the risky asset a^i and the capital market portfolio:

$$p(\alpha) = \alpha a^i + (1-\alpha)\hat{\Theta}_M \cdot \hat{S}. \tag{2.1.35}$$

Denote the expected return and the standard variation of $p(\alpha)$ by μ_α and σ_α, respectively, we have

$$\mu_\alpha = \alpha\mu_i + (1-\alpha)\mu_M, \tag{2.1.36}$$

and

$$\sigma_\alpha^2 = \alpha^2\sigma_i^2 + 2\alpha(1-\alpha)\sigma_{iM} + (1-\alpha)^2\mu_M^2, \tag{2.1.37}$$

where μ_i and σ_i are the expected return and standard deviation of asset a^i, respectively. The parametric curve $(\sigma_\alpha, \mu_\alpha)$ must lie below the capital market line because the latter consists of optimal portfolios. On the other hand it is clear that when $\alpha = 0$ this curve coincide with the capital market line. Thus, the capital market line is an tangent line of the parametric curve $(\sigma_\alpha, \mu_\alpha)$ at $\alpha = 0$. It follows that

$$\frac{\mu_M - R}{\sigma_M} = \left[\frac{d\mu_\alpha}{d\sigma_\alpha}\right]_{\alpha=0} = \frac{\sigma_M(\mu_i - \mu_M)}{\sigma_{iM} - \sigma_M^2}. \tag{2.1.38}$$

Solving for μ_i we derive

$$\mu_i = R + \beta_i(\mu_M - R). \tag{2.1.39}$$

●

2.1.3 Sharpe Ratio

Think a little bit more we will realize that to construct the capital market portfolio, theoretically, we need to use every available risky asset available to us. Given the huge number of available equities, constructing the capital market portfolio is practically impossible even if we have accurate probability distribution information on all the available risky assets (which is another impossible task). Thus, we have to deal with suboptimal situation. What happens if we mix risk free asset with an arbitrary portfolio of risky assets (not necessarily the capital market portfolio)? Let $\hat{\Theta} = (\theta_1, \ldots, \theta_M)$ be such a portfolio corresponding to risky assets $(a^1, \ldots, a^M)$ with price random vector $\hat{S} = (S^1, \ldots, S^M)$. Again we standardize the portfolio so that $\hat{\Theta} \cdot \hat{S}_0 = 1$. Denote $\mu^* = \mathbf{E}[\hat{\Theta} \cdot \hat{S}_1]$ and $\sigma^* = \sqrt{\mathrm{Var}(\hat{\Theta} \cdot \hat{S}_1)}$. Then any mix of this portfolio with a risk free asset having return R will produce a portfolio whose expected return μ and standard deviation σ lies on the line

$$\mu = \frac{\mu^* - R}{\sigma^*}\sigma + R. \tag{2.1.40}$$

Portfolios of risky assets with larger $\frac{\mu^*-R}{\sigma^*}$ have the potential of generating higher return for a fixed level of risk (see Fig. 2.4). Sharpe proposes the formula to compare risky portfolios such as those maintained by mutual funds using this idea. As an illustration, suppose that $R_1, \ldots, R_N$ are the monthly returns of a mutual fund a in the past N months and the monthly return of the risk free asset is R. Define a random variable X with finite values $\{R_n - R \mid n = 1, \ldots, N\}$ and $\mathrm{prob}(X = R_n - R) = 1/N$. Then the Sharpe ratio of a is defined as

$$s(a) = \frac{\mathbf{E}[X]}{\sqrt{Var(X)}}. \tag{2.1.41}$$

We can see that the Sharpe ratio is, in fact, a statistical estimate of $\frac{\mu^*-R}{\sigma^*}$.

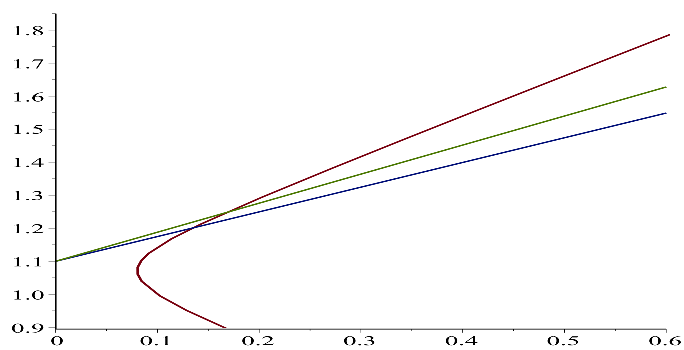


Fig. 2.4. Sharpe ratio

2.2 Utility Functions

In financial problems maximizing utilities and minimizing risks are constant themes. In the Markowitz portfolio theory, one uses expected return to measure performance and the variance to measure the risk. They are among the simplest of such measures. In general, utility functions most of the time are concave and risk measures are convex. Hence convex analysis is a natural tool in dealing with financial modeling.

2.2.1 Utility Functions

In 1738, while working in St. Petersburg, Daniel Bernoulli posted the following problem later known as the St. Petersburg Wager paradox:

"Peter tosses a coin and continues to do so until it should land 'heads' when it comes to the ground. He agrees to give Paul one ducat if he gets 'heads' on the very first throw, two ducats if he gets it on the second, four if on the third, eight if the on the fourth, and so on, so that with each additional throw the number he must pay is doubled. Suppose we seek to determine the value of Paul's expectation."

Of course assuming a fair coin we can easily calculate the expectation to be

$$\sum_{n=1}^{\infty} 2^{n-1} \cdot P(getting\ the\ first\ head\ on\ the\ nth\ throw)$$
$$= \sum_{n=1}^{\infty} 2^{n-1}\frac{1}{2^n} = \sum_{n=1}^{\infty}\frac{1}{2} = \infty.$$

The paradox lies in according to this computation the value of the rights of playing such a game would be infinity. In other words, one would be willing to pay any cost to play it, which is obviously absurd.

One way to resolve this is based on Peter only has limited amount of money so that he must post a limit on the payoff and, therefore, restricts the number of throwing of coins.

This looks more like playing a lottery. However, the expectation alone does not explain the fact that people keep playing the lottery even if it is not a fair game (since lottery makes money the expectation must be less than the price of tickets). Daniel Bernoulli himself suggested a solution which became highly influential later. Observing that an extra 100 ducat maybe considered a small fortune to a poor it may mean little to a rich, Daniel Bernoulli argued that people intuitively value money not according to its face value but its relative usefulness. Mathematically, he introduced *utility function* to capture this. For the St. Petersburg Wager problem, Bernoulli suggested to use $u(x) = \ln(x)$ as the utility function.

Bernoulli choose the ln as a utility function because of two of the properties of this function. First the ln function is increasing signaling the more the better. Second the derivative of the ln function is $1/x$ which is decreasing. This matches the intuition that the more you have the less you care about additional money. Abstractly, let us denote a utility function by $u(x)$. For convenience let us assume u is twice differentiable. Then we can characterize the above two properties as $u'(x) \geq 0$ and $u''(x) \leq 0$. Alternatively, without assuming differentiability of u we can also coding the intuition above mathematically by requiring a utility function to be an increasing *concave* function. We say a function $f : R \to R$ is concave if and only if $-f$ is convex. If $-f$ is concave we say f is convex. Usually we assume rational agents maximizing their expected utility when making decisions. Thus, convex optimization becomes important in analyzing financial problems.

There are many increasing concave functions. A few are listed below.

- Power utility: $(x^{1-\gamma} - 1)/(1-\gamma), \gamma > 0$.
- Log utility: $\ln(x)$.
- Exponential utility: $-e^{-\alpha x}, \alpha > 0$.

In dealing with a particular application problem the choice of the utility function is often based on economic or tractability considerations. Different agents can have different utility functions that reflect their own attitude towards rewards and risks of various degree.

For our mathematical model, it is important to know what kind of general conditions we should impose on a utility function. We consider a general extended valued upper semicontinuous utility function u. The following is a collection of additional conditions that are often used in financial models to accommodate different levels of tolerance to risk:

(u1) (Risk aversion) u is strictly concave,
(u2) (Profit seeking) u is strictly increasing and $\lim_{t\to+\infty} u(t) = +\infty$,
(u3) (Bankruptcy forbidden) For any $t < 0$, $u(t) = -\infty$,

2.2.2 Measuring Risk Aversion

Comparing tendency of risk aversion by directly examining the utility functions is difficult. The following tools are useful.

Definition 2.2.1 (Arrow-Pratt Absolute Risk Aversion Coefficient (ARA)) *The* coefficient of absolute risk aversion *is defined as*

$$A(x) = -\frac{u''(x)}{u'(x)}.$$

Constant absolute risk aversion (CARA) refers to $A(x) = \alpha$ is a constant, e.g. $u(x) = 1 - e^{-\alpha x}$ *Hyperbolic absolute risk aversion (HARA)* refers to $A(x) = 1/(ax + b)$ is a hyperbolic function, e.g.

$$u(x) = \frac{(x - x_0)^{1-\gamma}}{1-\gamma}$$

where $\gamma = 1/a$, $x_0 = -b/a$.

Definition 2.2.2 (Relative Risk Aversion Coefficient (RRA)) *The* coefficient of relative risk aversion *is defined as*

$$R(x) = -\frac{xu''(x)}{u'(x)}.$$

When ARA decreases the investor will increase risky investment in absolute amount. Similarly, when RRA decreases the investor will increase risky investment in percentage.

The property that an utility function has bounded ARA and RRA can be characterized by generalized convexity. We showcase the proof for RRA.

Theorem 2.2.3 (Characterization of Bounded Relative Risk Aversion) *Let* $u : \mathbb{R}_+ \to \mathbb{R}$ *be an increasing (decreasing) function with continuous second order derivative. Then, for any* $p \in \mathbb{R}$, *u has a coefficient of relative risk aversion* $R(x) \leq (\geq) 1 - p$ *if and only if u is* $\Phi_{(x^p y)(1)}$*-convex.*

Proof. We focus on the case that u is increasing and the case of decreasing is similar. The "If" part. Assume u is $\Phi_{(x^p y)(1)}$-convex. Then, for any $x > 0$ we can find $y(x), b(x)$ such that

$$u(z) \geq y(x)z^p - b(x), \text{ for all } z > 0$$

with equality holds at $z = x$. Let

$$z \to f(z) := u(z) - y(x)z^p + b(x).$$

We have $f'(x) = 0, f''(x) \geq 0$, which give us

$$R(x) = -\frac{xu''(x)}{u'(x)} \leq 1 - p.$$

The "Only if" part. Write the $R(x) \leq 1 - p$ condition as

$$\frac{u''(s)}{u'(s)} \geq \frac{p-1}{s}.$$

Then solving for u on $[x, z]$. Details are left as an exercise.

●

Similarly, we have

Theorem 2.2.4 (Characterization of Bounded Absolute Risk Aversion) *Let* $u : \mathbb{R}_+ \to \mathbb{R}$ *be an increasing (decreasing) function with continuous second order derivative. Then, for any* $p \in \mathbb{R}$, *u has a coefficient of absolute risk aversion* $A(x) \leq (\geq) p$ *if and only if u is* $\Phi_{(e^{-px} y)(1)}$*-convex.*

Remark 2.2.5 It is not hard to see that the above two theorems are also valid for functions with piecewise continuous second order derivatives.

2.2.3 Growth Portfolio Theory

Maximizing the expected log utility leads to the growth portfolio theory. The convex optimization problem can be stated as

$$\text{maximizing } \mathbf{E}[\ln(\Theta \cdot S_1)] \tag{2.2.1}$$

$$\text{subject to } \Theta \cdot S_0 = w_0. \tag{2.2.2}$$

To make the theory fit to various application situations one often standardize it by assume $w_0 = 1$, $S_0 = \vec{1}$, $S_1^0 = 1$ and $\Theta^0 = 1$. Then $g = \hat{S}_1 - \hat{S}_0$ represents the vector of percentage return of the risky assets in the market. We can then write problem (2.2.1) as

$$\text{maximizing } \mathbf{E}[\ln(1 + \hat{\Theta} \cdot g)]. \tag{2.2.3}$$

Note that

$$\mathbf{E}[\ln(1 + \hat{\Theta} \cdot g)] = \sum_{n=1}^{N} \ln(1 + \hat{\Theta} \cdot g(B_n))P(B_n) = \ln[\Pi_{n=1}^{N}(1 + \hat{\Theta} \cdot g(B_n))^{P(B_n)}].$$

Problem (2.2.3) has the same effect as to maximizing

$$\Pi_{n=1}^{N}(1 + \hat{\Theta} \cdot g(B_n))^{P(B_n)},$$

the compounded return or "growth".

A growth optimal portfolio has the theoretical advantage of maximum rate of growth of one's wealth. However, in practice it often suffers the drawback of being too risky. To understand this risk let us look at a simple financial market with only one risk asset. In this case $s = \hat{\Theta}$ is just one real number. For the simplicity of the notation we denote $g_n = g(B_n)$ and $p_n = P(B_n)$. Then the growth portfolio optimization problem becomes

$$\text{maximizing} \quad f(s) = \sum_{n=1}^{N} p_n \ln(1 + s g_n). \tag{2.2.4}$$

We will call $f(s)$ a log return function.

Theorem 2.2.6 (Compute the Optimal Leverage) *Assume without loss of generality that $g_1 < g_2 < \ldots < g_N$. Then the optimal leverage $\bar{s}$ is determined by the unique solution of the $(N-1)$th order polynomial equation*

$$0 = \Pi_{n=1}^{N}(1 + s g_n) \left(\sum_{n=1}^{N} \frac{p_n g_n}{1 + s g_n} \right) \tag{2.2.5}$$

on the interval $(-\frac{1}{g_N}, -\frac{1}{g_1})$.

Proof. Since the log return function,

$$f(s) = \sum_{n=1}^{N} p_n \ln(1 + s g_n),$$

is a strictly concave function on $(-\frac{1}{g_N}, -\frac{1}{g_1})$, its derivative is strictly decreasing. Moreover, it is easy to see that $\lim_{s \to (-1/g_N)+} f'(s) = \infty$ and $\lim_{s \to (-1/g_1)-} f'(s) = -\infty$. Thus, there is a unique solution $\bar{s}$ to the equation

$$0 = f'(s) = \sum_{n=1}^{N} \frac{p_n g_n}{1 + s g_n} \tag{2.2.6}$$

on $(-\frac{1}{g_N}, -\frac{1}{g_1})$ which is the optimal leverage.

Finally, observing that the polynomial $\Pi_{n=1}^{N}(1 + sg_n)$ has no solution in the interval $(-\frac{1}{g_N}, -\frac{1}{g_1})$, which shows that $\bar{s}$ must be the unique solution of the $(N-1)$th polynomial equation

$$0 = \Pi_{n=1}^{N}(1 + sg_n) \left(\sum_{n=1}^{N} \frac{p_n g_n}{1 + sg_n} \right)$$

on the inverval $(-\frac{1}{g_N}, -\frac{1}{g_1})$. ●

When the market has only two or three states explicit solutions are not hard to derive. Those results are very useful for analyzing betting on games and, therefore, presented below.

Proposition 2.2.7 (Two States) *Consider a market with two distinct states represented by $g_1 < g_2$ corresponding to probabilities p_1 and p_2, respectively. Then the best investment size is*

$$\bar{s} = -\frac{p_1 g_1 + p_2 g_2}{g_1 g_2}. \tag{2.2.7}$$

Proof. The log return function for such an investment system is $f(s) = p_1 \ln(1 + sg_1) + p_2 \ln(1 + sp_2)$. By Theorem 2.2.6, the best investment size $\bar{s}$ is the solution of equation

$$0 = (1 + sp_1)(1 + sp_2) \left(\frac{p_1 g_1}{1 + sp_1} + \frac{p_2 g_2}{1 + sg_2} \right).$$

Solving this equation produces equation (2.2.7). ●

Proposition 2.2.8 (Three States) *Consider a market with three distinct states represented by $g_1 < g_2 < g_3$ corresponding to probabilities p_1, p_2 and p_3, respectively. Then the best investment size $\bar{s}$ is given by*

$$\bar{s} = \begin{cases} 0 & \text{if } C = 0 \\ -\frac{p_1 g_1 + p_3 g_3}{(p_1 + p_3) g_1 g_3} & \text{if } g_2 = 0 \\ \frac{-B + \sqrt{B^2 - 4AC}}{2A} & \text{if } C < 0, g_2 \neq 0 \\ \frac{-B - \sqrt{B^2 - 4AC}}{2A} & \text{if } C > 0, g_2 \neq 0. \end{cases} \tag{2.2.8}$$

Here $A = g_2 g_2 g_3$, $B = g - 2[p_3 g_3 + p_1 g_1 + p_2(g_1 + g_3)] + (p_1 + p_3) g_1 g_3$ and $C = p_1 g_1 + p_2 g_2 + p_3 g_3$.

Proof. The proof is similar to that of Proposition 2.2.7 and is left as an exercise. ●

Remark 2.2.9 (The Kelly Criterion and the Shannon Information Rate) In Proposition 2.2.7 if $-g_1 = g_2 = 1$ are symmetric and standardized then at the best leverage size

$$\bar{s} = p_2 - p_1$$

the value of the log return function is

$$f(\bar{s}) = p_1 \ln p_1 + p_2 \ln p_2 + \ln 2.$$

This is Shannon's information rate for a communication channel with noise [50]. Note that when $g_2 = -1$ and $g_1 = 1$ our portfolio is equivalent to a game with symmetric payoffs. This is exactly what Kelly observed in [27], which is that Shannon's information rate can be explained as the best possible outcome of using communication channel with noise when the signal is used for a game with symmetric payoffs.

Let us apply Proposition 2.2.7 to a simplified Blackjack game.

Example 2.2.10 (Money Management in Blackjack) In play a certain version of the Blackjack we know with counting cards a skilled player has a winning probability of 51% over the house. We simplify the problem by assuming the win and loss are always equal to the bet and apply Proposition 2.2.8 to determine the best betting size s as a percentage of all the bankroll of the player. In this case $g_2 = 1$ (wining 100% of the bet), $g_1 = -1$ (losing 100% of the bet), $p_2 = 51\%$ and $p_1 = 49\%$. Thus, the best betting size is

$$\bar{s} = -\frac{p_1 g_1 + p_2 g_2}{g_1 g_2} = 2\%.$$

This is actually recommended by Ed Thorp an expert in the Blackjack game and a pioneer in applying the Kelly method to investment management in his classical book [56]. ●

The game of Blackjack has changed a lot and the player's advantage has mostly slipped away due to the use of multiple deck of cards and frequent shuffling. However, even if the assumption in Example 2.2.10 were correct, the optimal betting size $\bar{s}$ is too aggressive as explained in the next example.

Example 2.2.11 Now consider playing a game with symmetric payoff $t = -c = 1$ with the wining probability of 90%. We can easily calculate that the best betting size $\bar{s} = 80\%$. Putting 80% of your wealth on the line is clearly too aggressive no matter how favorable the game is to you. ●

2.2.4 Efficiency Index

Despite the short comings of the growth portfolio theory, similar to the Markowitz portfolio theory the idea can also be used to construct a criterion for evaluating investment performance. The key is to realize by examining e.g. Propositions 2.2.7 that the effectiveness of an investment strategy must be evaluated with appropriate leverage level.

Example 2.2.12 We consider two simplified investment strategies each with ten trades whose percentage gain (loss) is listed in the first two columns of Table 1. The effects of the two systems are tested using an investment capital of \$100 with two different investment sizes: 100% and 30% of the available capital for each trade, respectively. The results show that with an investment size of 100% of the available capital for each trade, System 2 is better than System 1, but with an investment size of 30% System 1 becomes better. ●

trades	S1 %gain	S2 %gain	100% S1	100% S2	30% S1	30% S2
1	13%	6%	113.00	106.00	103.90	101.80
2	−25%	6%	84.75	112.36	96.11	103.63
3	13%	−5%	95.77	106.74	99.86	102.08
4	−25%	−5%	71.83	101.40	92.37	100.55
5	−25%	6%	53.87	107.49	85.44	102.36
6	13%	−5%	60.87	102.11	88.77	100.82
7	13%	−5%	68.79	97.00	92.23	99.31
8	13%	−5%	77.73	92.16	95.83	97.82
9	13%	6%	87.83	97.69	99.57	99.58
10	13%	6%	99.25	103.55	103.45	101.37

Table 2.1. Effects of investment systems under different investment sizes.

How to put them on a leveled field? One way to do it is to compare them using their best investment size respectively. This leads to the following definition.

Definition 2.2.13 (Efficiency Index) *Suppose an investment strategy is characterized by its returns $g \in RV(\Omega, \mathcal{F}, P)$. we define its* efficiency index γ *as*

$$\gamma = \max_{s \in [-1/\max(g_n), -1/\min(g_n)]} \sum_{n=1}^{N} p_n \ln(1 + s g_n), \tag{2.2.9}$$

where $g_n = g(B_n)$ *and* $p_n = P(B_n)$.

If $g_n \geq 0, n = 1, \ldots, N$ or $g_n \leq 0, n = 1, \ldots, N$ then we have an arbitrage signaled by $\gamma = +\infty$. Otherwise the efficiency index γ is the log return of the portfolio of cash and the given investment strategy under the best leverage level. In view of Remark 2.2.9 the efficiency index gauges the useful information contained in an investment strategy.

Example 2.2.14 Let us re-examine Example 2.2.12 using the efficiency index. Both investment strategies in Example 2.2.12 have two distinct gains and their profiles are summarized below.

	g_1	p_1	g_2	p_2
Strategy 1	13%	0.7	−25%	0.3
Strategy 2	6%	0.5	−5%	0.5

Drawing the log return functions of these two investment strategies simultaneously in Figure 2.5 we can understand the reasons behind the phenomenon observed in Example 2.2.12. Moreover, we see that neither strategy was tested in Example 2.2.12 under the best investment size. Using Theorem 2.2.7 we can calculate that, for

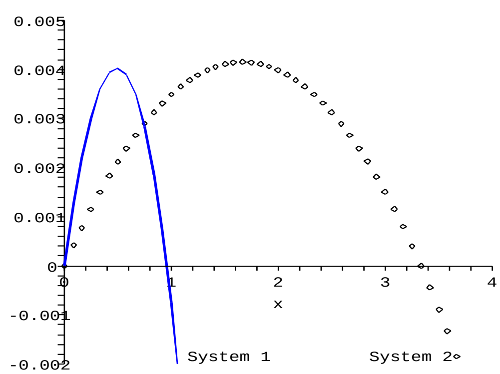


Fig. 2.5. Log return functions

Strategy 1, $\bar{s} = 49\%$, $\gamma = 0.040$ and for Strategy 2, $\bar{s} = 167\%$, $\gamma = 0.041$. If we compare the efficiency indices only then the two investment strategies are almost the same while Strategy 2 is slightly better. Yet this fact is hard to unveil without the help of the efficiency index. However, if margin is not allowed then Strategy 1 is the better choice even though Strategy 2 has a slightly higher efficiency index. ●

Remark 2.2.15 While the Kelly criterion, as a special case of the GOP, can help us calculating a theoretical best betting size for any game, in practice such an optimal strategy is often too risky as illustrated in Example 2.2.11. Various fractional Kelly betting scheme, often ad hoc where proposed to limiting the risk of GOP. Recently Vince and Zhu [63] and Lopez de Prado, Vince and Zhu [33] provided theoretical justification for such more conservative betting strategies. They use more realistic finite investment horizon and select betting size based on risk adjusted returns. The analysis involves, however, nonconvex functions.

2.3 Fundamental Theorem of Asset Pricing

We turn to consider optimizing a general utility of the payoff of a portfolio $\Theta \in \mathbb{R}^{M+1}$. We wish to endow a norm on the space of portfolios that can reflect the size of a portfolio. Intuitively, the magnitude of Θ as a vector in $\mathbb{R}^{M+1}$ in a sense indicates the level of capital commitment or leverage level of a portfolio. However, one needs to be careful here. Holding a portfolio, an investor's goal is to derive a risk adjusted gain represented by the random variable

$$\Theta \cdot (S_1 - S_0) \in RV(\Omega, \mathcal{F}, P). \tag{2.3.10}$$

We can see that increasing or reducing the share of cash in the portfolio clearly swings the leverage level as measured by the magnitude of Θ, yet does nothing to the gain (2.3.10). The following example shows that even if we fix the share of the cash, such a phenomenon can still happen.

Example 2.3.1 (Infinitely Many Portfolio with Equivalent Gain) Consider a state space $\Omega = \{0, 1\}$ and with a financial market with three risky assets whose prices at times $0, 1$ are given by $S_0 = (1, 1, 1, 1)$, $S_1(0) = (1, 0.8, 0.9, 1)$ and $S_1(1) = (1, 1.1, 1.2, 1.1)$. We can easily verify that for portfolio $\bar{\Theta} = (1, 1, -2, 3)$, $\bar{\Theta} \cdot (S_1 - S_0)(i) = 0$ for both $i = 0$ and $i = 1$. It follows that for any $r \in \mathbb{R}$, all the portfolios $\Theta + r\bar{\Theta}$ have the same gain. ●

Notice that as $|r| \to \infty$, the magnitude of $\Theta + r\bar{\Theta} \in \mathbb{R}^{M+1}$ also goes to infinity. This example demonstrates that the magnitude of a portfolio in $\mathbb{R}^{M+1}$ is not an appropriate measure for the leverage level of the portfolio. Moreover, it clearly does not make sense in practice to use a portfolio of the form $\Theta + r\bar{\Theta}$ with large $|r|$. This is because doing so will greatly increase the risk (as the price of assets in a financial market is not deterministic) without benefit to the gain. These considerations lead to the following definitions:

Definition 2.3.2 (Equivalent Portfolios) *We say two portfolios Θ^1 and Θ^2 are equivalent in market S if they have the same initial value and the gain, that is to say,*

$$\Theta^1 \cdot S_0 = \Theta^2 \cdot S_0 \tag{2.3.11}$$

and, as random variables,

$$\Theta^1 \cdot (S_1 - S_0) = \Theta^2 \cdot (S_1 - S_0).$$

We will use $S[\Theta]$ to denote all the portfolios that are equivalent to Θ in market S.

Since all the portfolio in $S[\Theta]$ are equivalent we prefer those that have low leverages as measured by $\|\Theta\|$. The following lemma provides us with an optimally leveraged portfolio in each equivalent class.

Lemma 2.3.3 *For any portfolio Θ in S, the optimization problem*

$$\min\{\|x\| : x \in S[\Theta]\}. \tag{2.3.12}$$

has a unique solution $\underline{\Theta}$. Moreover, there exists a constant $K = K(S)$ depending only on S such that, for any portfolio Θ,

$$\|\underline{\Theta}\| \le K\|\Theta \cdot (S_1 - S_0)\|_{RV}. \tag{2.3.13}$$

Here $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^{M+1}$ and $\|\cdot\|_{RV}$ is the norm on $RV(\Omega, \mathcal{F}, P)$ induced by the inner product $\langle\cdot,\cdot\rangle$.

Proof. Note that problem (2.3.12) and the following problem (2.3.14) has the same solution

$$\min\{\|x\|^2 : x \in S[\Theta]\}. \tag{2.3.14}$$

Denote

$$A = \begin{bmatrix} S_1(B_1) - S_0 \\ S_1(B_2) - S_0 \\ \dots \\ S_1(B_N) - S_0 \end{bmatrix},$$

where $\{B_1, \dots, B_N\}$ are the set of atoms of the probability space $(\Omega, \mathcal{F}, P)$. Then A is a $N \times (M+1)$ matrix.

We observe that $x \in S[\Theta]$ amounts to requiring

$$Ax = \Theta \cdot (S_1 - S_0). \tag{2.3.15}$$

We first consider the special case when $\text{rank}(A) = \min(M+1, N)$ If $\text{rank}(A) = M+1$, the constraint uniquely determines $\underline{\Theta} = x = (A^\top A)^{-1} A^\top \Theta \cdot (S_1 - S_0)$. Otherwise, $\text{rank}(A) = N$ and the quadratic function $\|x\|^2$ attains a minimum on the affine set characterized by the linear constraint. It is easy to calculate this solution to be $\underline{\Theta} = x = A^\top(AA^\top)^{-1}\Theta \cdot (S_1 - S_0)$. In both cases $\underline{\Theta}$ is unique. Moreover, defining

$$K = K(S) = \max(\|A^\top(A^\top A)^{-1}\|, \|(AA^\top)^{-1}A^\top)\|,$$

we have (2.3.13).

If $\operatorname{rank}(A) < \min(M+1, N)$ then we can first remove the rows or columns in A that are dependent on others and then apply the above special case to the reduced matrix A. ●

Definition 2.3.4 (Portfolio Space) *We call the quotient space of* $\mathbb{R}^{M+1}$ *with respect the portfolio equivalent relationship in market* S *the* portfolio space *on* S *and denote it* $port[S]$. *For* $\Theta \in port[S]$ *we define its norm by*

$$\|\Theta\|_p = \|\underline{\Theta}\|.$$

The portfolio space $(port[S], \|\cdot\|_p)$ *is a finite dimensional Banach space.*

2.3.1 Fundamental Theorem of Asset Pricing

Gain without risk is what every investor desires. Such opportunities arguably will not last as when everyone tries to chase it. Based on this observation, in a financial market a guiding principle is that arbitrage should not exist. The following is a formal definition.

Definition 2.3.5 (Arbitrage) *We say that a portfolio* Θ *is an arbitrage if it involves no risk,* $\Theta \cdot (S_1 - S_0) \geq 0$ *and has opportunity to gain something* $\Theta \cdot (S_1 - S_0) \neq 0$.

A rational investor with a utility function u satisfying conditions (u1)-(u3) will try to maximize the expected utility of the final wealth among all portfolios in $port[S]$. In other words, if $w_0 > 0$ is the initial wealth of the investor, he wants to solve the following portfolio utility maximization problem. Find:

$$\sup\{\mathbf{E}[u(w_0 + \Theta \cdot (S_1 - S_0)] : \Theta \in port[S]\}. \tag{2.3.16}$$

It turns out that an arbitrage opportunity is exactly characterized by the optimal value for problem (2.3.16) to be $+\infty$.

Theorem 2.3.6 (Characterizing Arbitrage with Utility Optimization) *The portfolio space* $port[S]$ *contains an arbitrage if and ony if the optimal value of the utility optimization problem is* $+\infty$

Proof. The "only if" part is easy: if $\Theta \in port[S]$ is an arbitrage then so is $r\Theta$ for any $r > 0$. Then it is easy to see that $\mathbf{E}[u(w_0 + r\Theta \cdot (S_1 - S_0)] \to +\infty$ as $r \to +\infty$.

To prove the "if part" assume the optimal value for problem (2.3.16) is $+\infty$. Then there exists a sequence $\Theta^n \in port[S]$ such that $\mathbf{E}(u(w_0 + \Theta^n \cdot (S_1 - S_0)) \to +\infty$ as $n \to +\infty$. Necessarily, $t_n = \|\Theta^n \cdot (S_1 - S_0)\|_{RV} \to +\infty$ as n goes to ∞. By Lemma 2.3.3 there exists a constant $K = K(S)$ such that $\|\underline{\Theta^n}/t_n\| \le K$. Without loss of generality we may assume that $\underline{\Theta^n}/t_n$ converges to some $\Theta^* \in port[S]$. Note that, for any n, $\Theta^n \cdot (S_1 - S_0) \ge -w_0$ by property (u3) of the utility function. Thus, $\Theta^* \cdot (S_1 - S_0) \ge 0$. Also,

$$\|\Theta^* \cdot (S_1 - S_0)\| \ge \liminf_{n\to\infty} \|\Theta^n \cdot (S_1 - S_0)/t_n\| = 1.$$

Therefore, Θ^* is an arbitrage. ●

The fundamental theorem of asset pricing (FTAP) links no arbitrage with the existence of risk neutral or martingale measures defined below:

Definition 2.3.7 (Equivalent Martingale Measure) *We say that Q is an* equivalent martingale measure (EMM) *on economy $(\Omega, \mathcal{F}, P)$ for financial market S provides that, for any atom B_i of $\mathcal{F}$, $Q(B_i) \neq 0$ if and only if $P(B_i) \neq 0$, and*

$$\mathbf{E}^Q[S_1] = S_0.$$

Given an initial wealth $w_0 > 0$, the set of all achievable wealth outcomes at the end of the one period economy $t = 1$ using all possible portfolios is

$$w_0 + \{\Theta \cdot (S_1 - S_0) : \Theta \in port[S]\} \subset RV(\Omega, \mathcal{F}, P).$$

We denote the set of gains

$$W := \{\Theta \cdot (S_1 - S_0) : \Theta \in port[S]\} \subset RV(\Omega, \mathcal{F}, P).$$

In fact, W is a subspace of $RV(\Omega, \mathcal{F}, P)$. It is not hard to see that if Θ is an arbitrage portfolio then $\Theta \cdot (S_1 - S_0) \in RV(\Omega, \mathcal{F}, P)^+ \backslash \{0\}$, where $RV(\Omega, \mathcal{F}, P)^+$ is the cone of nonnegative random variables. Thus, no arbitrage can be described as

$$W \cap RV(\Omega, \mathcal{F}, P)^+ \backslash \{0\} = \emptyset.$$

Traditional proof of the FTAP relies on applying an appropriate version of the cone separation theorem to ensure that there is a hyperplane separating W and $RV(\Omega, \mathcal{F}, P)^+$. Then, a scaling of the normal vector of such a separation hyperplane gives us an equivalent martingale measure. This geometric picture is often interpreted as the no arbitrage price being independent of investors preferences. However, we will give a proof of the FTAP below based on portfolio utility optimization (2.3.16). We show that the equivalent martingale measure can be viewed as a scaling of the solution to the dual problem or equivalently the Lagrange multiplier related to such a utility optimization problem. As a result, a pricing martingale measure does depend on the utility function of the investor when the market is incomplete.

Theorem 2.3.8 (Refined Fundamental Theorem of Asset Pricing) *Let S be a financial market, let u be a utility function that satisfies properties (u1),(u2) and (u3) and let $w_0 \geq 0$ be a given initial endowment. Then TFAE:*

(i) *$port[S]$ contains no arbitrage.*
(ii) *The optimal value of the portfolio utility optimization problem (2.3.16) is finite and attained.*
(iii) *There is an equivalent S-martingale measure proportional to a subgradient of $-u$ at the optimal solution of (2.3.16).*

Proof. First observe that the utility optimization problem (2.3.16) can be written equivalently as

$$\begin{aligned} \max \quad & \mathbf{E}[u(y)] \\ \text{subject to} \quad & y \in w_0 + W. \end{aligned} \tag{2.3.17}$$

Define $f(y) = -\mathbf{E}[u(y)]$ and $g(y) = \iota_{w_0+W}(y)$. Then we can rewrite problem (2.3.17) as

$$-\min_{y}\{f(y) + g(y)\} \tag{2.3.18}$$

The dual problem of (2.3.18) is,

$$\begin{aligned} &-\max\{-f^*(-z) - g^*(z)\} \\ &= \min\{\mathbf{E}[(-u)^*(-z)] + \langle w_0, z\rangle + \sigma_W(z)\} \end{aligned} \tag{2.3.19}$$

Since we can check that the constraint qualification condition

$$w_0 \in \mathrm{ri}[\mathrm{dom}\, g - \mathrm{dom}\, f] = \mathrm{ri}[w_0 + W - RV(\Omega, \mathcal{F}, P)^+ \backslash \{0\}] \tag{2.3.20}$$

(corresponding to (1.4.7)) holds, Fenchel strong duality implies (2.3.18) and its dual (2.3.19) have the same value.

By Theorem 2.3.6, $port[S]$ contains no arbitrage if and only if the optimal values of problem (2.3.16) are finite and, therefore, the dual problems (2.3.18) and (2.3.19) are all finite. Since W is a subspace, the optimal value of (2.3.19) is not $-\infty$ implies that its solution $z \perp W$. Moreover, $\mathbf{E}[(-u)^*(-z)] > -\infty$ implies that $z(B_i) > 0$ for all $P(B_i) \neq 0$. Thus, $Q = z/\mathbf{E}[z]$ is a risk neutral measure equivalent to P. That is, (i) implies (ii).

On the other hand the existence of an equivalent S-martingale measure implies that the constraint qualification condition for (2.3.19) holds. In fact, problem (2.3.19) can be viewed as minimizing the convex function $z \to \mathbf{E}[(-u)^*(-z)] + \langle w_0, z\rangle$ over the entire subspace $W^{\perp}$ ($z > 0$ is merely a consequence of the domain of $\mathbf{E}[(-u)^*(\cdot)]$ being a subset of int $-RV(\Omega, \mathcal{F}, P)^+$ and, therefore, is not a separate constraint). Thus, the constrain qualification condition for (2.3.19) satisfies (see e.g. [65, Theorem 2.7.1]). It follows that problem (2.3.16) which equivalent to (2.3.18) as the dual of (2.3.19) has a finite value and attains its solution, which is to say (ii) implies (iii).

Finally, if (iii) is true then there cannot be any arbitrage in $port[S]$ because adding an arbitrage to the optimal solution of (2.3.16) will improve it. Thus, (iii) implies (i) and we have completed a cyclic proof of the equivalence of (i), (ii) and (iii). ●

An equivalent martingale measure can also be viewed as a scaling of a Lagrange multiplier for the portfolio utility optimization problem (2.3.16) due to the relationship between Lagrange multipliers and dual solutions highlighted in [9]. To see this let us rewrite problem (2.3.16) as a constrained minimization problem

$$\begin{aligned} \text{minimize} \quad & \mathbf{E}[(-u)(x)] \\ \text{subject to} \quad & x - \Theta \cdot (S_1 - S_0) - w_0 = 0. \end{aligned} \tag{2.3.21}$$

We have already known from the proof of the Theorem 2.3.8 that this problem has a solution (x^*, Θ^*). Moreover, since we know strong duality holds and the dual problem has a solution, which implies that problem (2.3.21) has a Lagrange multiplier. Let λ be the Lagrange multiplier of problem (2.3.21). Then the Lagrangian is

$$\begin{aligned} L((x, \Theta), \lambda) &= \mathbf{E}[(-u)(x)] + \langle \lambda, x - \Theta \cdot (S_1 - S_0) - w_0 \rangle \\ &= \mathbf{E}[(-u)(x)] + \langle \lambda, x - w_0 \rangle - \langle \lambda, \Theta \cdot (S_1 - S_0) \rangle \\ &= \mathbf{E}[(-u)(x) + \lambda(x - w_0)] - \langle \lambda, \Theta \cdot (S_1 - S_0) \rangle. \end{aligned}$$

It attains mininum at (x^*, Θ^*). Thus, we have $\langle \lambda, S_1 - S_0 \rangle = 0$ and $-\lambda(B_i) \in \partial(-u)(x^*(B_i)), i = 1, 2, \ldots, N$ for $P(B_i) > 0$. Since $-u$ is strictly decreasing we have $\lambda(B_i) > 0$ whenever $P(B_i) > 0$. Moreover, dividing $\langle \lambda, S_1 - S_0 \rangle = \mathbf{E}[\lambda(S_1 - S_0)] = 0$ by $\mathbf{E}[\lambda]$ and noticing that S_0 is a constant vector we get

$$\mathbf{E}[(\lambda/\mathbf{E}[\lambda])S_1] = S_0.$$

This is to say that $Q = (\lambda/\mathbf{E}[\lambda])P$ is a martingale measure equivalent to P. We can see that this martingale measure is indeed a scaling of the Lagrange multiplier.

Condition (u3) can be removed from Theorem 2.3.8 to derive a generalization of the version of FTAP in [17].

Theorem 2.3.9 (Refined Fundamental Theorem of Asset Pricing) *Let S be a market. Then the following are equivalent:*

(i) *There exists no arbitrage trading strategy in port$[S]$;*
(ii) *There is an equivalent S-martingale measure.*
(iii) *There exists a utility function u with properties (u1) and (u2), such that the finite optimal value of the trading strategy utility optimization problem (2.3.16) is attained.*

Proof. Implication (i) → (ii) → (iii) follows from Theorem 2.3.8. If the finite optimal value of the trading strategy utility optimization problem (2.3.16) is attained then there can be no arbitrage because superposition of an arbitrage to the optimal solution will improve it. Thus (iii) also implies (i) completing a cyclic proof. ●

Remark 2.3.10 Although the fundamental result of no arbitrage is equivalent to existence of an equivalent martingale measure is well known, as pointed out in [67] the proof of Theorem 2.3.8 using a class of utility functions says more: when the martingale measure is not unique, the dual problem actually points to one particular martingale measure. Thus, in principle, every choice of risk neutral measure (corresponding to a particular price of the contingent claim) can be viewed as a particular portfolio optimization problem with a corresponding concave utility function.

The useful perspective we can get from this exercise is that pricing contingent claims either by a replicating portfolio or by using a martingale measure can be viewed as a special case of portfolio optimization with respect to a certain utility function. Moreover, when the market is not complete there are many possibilities in selecting the utility functions. Thus, the pricing of contingent claims do rely on the trader's preference. There can exist many different reasonable prices as a result of the differences in trader's risk-reward preferences.

2.3.2 Pricing Contingent Claims

Suppose a contingent claim's payoff at $t = 1$ is $\phi(S_1)$, a function of the price of the assets at $t = 1$. To find the arbitrage free price ϕ_0 of this contingent claim we form a portfolio holding one such contingent claim along with a portfolio of other assets in the market scaled to the initial wealth of the investor and then (as in the previous section) consider the portfolio optimization problem of maximizing the utility of the final wealth:

$$\begin{aligned} \text{maximizing} \quad & \mathbf{E}[u(\beta(\phi(S_1) + \Theta \cdot S_1))] \\ \text{subject to} \quad & \beta(\phi(S_0) + \Theta \cdot S_0) = w_0. \end{aligned}$$

Equivalently we can write this portfolio optimization problem as

$$\begin{aligned} \text{minimizing} \quad & \mathbf{E}[(-u)(x)] \\ \text{subject to} \quad & x - \beta(\phi(S_1) - \phi_0 + \Theta \cdot (S_1 - S_0)) - w_0 = 0. \end{aligned} \tag{2.3.22}$$

Assume there is no arbitrage then Theorem 2.3.6 implies that the optimal value of problem 2.3.22 is finite and is attained at (x^*, β^*, Θ^*). As in the previous section that we can check that the constraint qualification condition for problem (2.3.22) is satisfied and, therefore, problem (2.3.22) has a Lagrange multiplier $\lambda \in RV(\Omega, \mathcal{F}, P)$ such that the Lagrangian

$$\begin{aligned} L((x, \beta, \Theta), \lambda) &= \mathbf{E}[(-u)(x)] + \langle \lambda, x - \beta[\phi(S_1) - \phi_0 + \Theta \cdot (S_1 - S_0)] - w_0 \rangle \\ &= \mathbf{E}[(-u)(x)] + \langle \lambda, x - w_0 \rangle - \langle \lambda, \beta[\phi(S_1) - \phi_0 + \Theta \cdot (S_1 - S_0)] \rangle \\ &= \mathbf{E}[(-u)(x) + \lambda(x - w_0)] - \langle \lambda, \beta[\phi(S_1) - \phi_0 + \Theta \cdot (S_1 - S_0)] \rangle, \end{aligned}$$

attains mininum at (x^*, β^*, Θ^*). Thus, we have $-\lambda(B_i) \in \partial(-u)(x^*(B_i)), i = 1, 2, \ldots, N$ for $P(B_i) > 0$. Since $-u$ is strictly decreasing we have $\lambda(B_i) > 0$ whenever $P(B_i) > 0$. Moreover, $\langle \lambda, S_1 - S_0 \rangle = 0$, which is $\mathbf{E}[\lambda(S_1 - S_0)] = 0$. Dividing by $\mathbf{E}[\lambda]$ and noticing that S_0 is a constant vector we get

$$\mathbf{E}[(\lambda/\mathbf{E}[\lambda])S_1] = S_0.$$

This is to say that $Q = (\lambda/\mathbf{E}[\lambda])P$ is a P-equivalent martingale measure. Finally, $\langle \lambda, \phi(S_1) - \phi_0 \rangle = 0$. That is

$$\phi_0 = E^Q[\phi(S_1)],$$

in other words, if there is no arbitrage then the price of the contingent claim must be the expectation of its payoff under one of the martingale measures that are equivalent to P.

We note that unless the market is complete, martingale measures are not unique. We can see from above that martingale measures and, therefore, the resulting prices of the contingent claim depend on the choice of utility functions. We now give a simple example that explicitly calculates the martingale measures in terms of a class of utility functions.

Example 2.3.11 Consider a market S contains only one risky asset. Assume that the market has N states $\Omega = \{\omega_1, \ldots, \omega_N\}$ and state ω_n happens with probability p_n. Assume for simplicity that $S_0 = 1$ and denote $x_n := S_1(\omega_n) - S_0$. In this case a trading strategy H is simply a constant h indicating the share of S that the trader holds. Given a utility function u satisfying properties (u1)–(u3) the utility maximization problem (2.3.16) takes the following concrete form:

$$\max \mathbf{E}[u(1 + h \cdot (S_1 - S_0))] = \sum_{n=1}^{N} p_n u(1 + hx_n). \tag{2.3.23}$$

Rewrite (2.3.23) as a constrained minimization problem

$$\begin{aligned} \min \quad & -\sum_{n=1}^{N} p_n u(y_n) \\ \text{subject to} \quad & y_n - 1 - hx_n = 0, n = 1, \ldots, N. \end{aligned} \tag{2.3.24}$$

Let's write the Lagrangian

$$L((y,h),\lambda) = -\sum_{n=1}^{N} p_n[u(y_n) + \lambda_n(y_n - 1 - hx_n)].$$

Setting $\nabla_{y,h} L = 0$ we derive, at the optimal solution,

$$\sum_{n=1}^{N} p_n \lambda_n x_n = 0, \tag{2.3.25}$$

and

$$\lambda_n = u'(y_n) = u'(1 + hx_n). \tag{2.3.26}$$

Equation (2.3.25) clearly shows that a scaled λ gives us the martingale measure. To solve for h so as to derive the solution to the utility optimization problem (2.3.23) we can substitute (2.3.26) into (2.3.25) to get the following equation for h,

$$\sum_{n=1}^{N} p_n u'(1 + hx_n) x_n = 0. \tag{2.3.27}$$

Equation (2.3.26) clearly shows that the martingale measure depends on the choice of utility function. ●

We continue this example by considering a concrete family of utility functions.

Example 2.3.12 (Risk Aversion) Let us consider a class of utility function that depend on parameter $c > 0$,

$$u_c(x) = \begin{cases} \ln x + cx & x > 0 \\ -\infty & x \leq 0, \end{cases}$$

and set $N = 3$, $p_1 = p_2 = p_3 = 1/3$ and $x_1 = 1$,$x_2 = 0.5$ and $x_3 = -0.5$.

In this case the Lagrangian is

$$L((y, h), \lambda) = -\sum_{n=1}^{N} p_n[\ln(y_n) + cy_n + \lambda_n(y_n - 1 - hx_n)].$$

At the optimal solution $(\bar{y}, \bar{h})$, equation (2.3.26) determines the Lagrange multiplier as

$$\lambda = (\lambda_1, \lambda_2, \lambda_3) = \left(\frac{1}{1+\bar{h}} + c, \frac{1}{1+0.5\bar{h}} + c, \frac{1}{1-0.5\bar{h}} + c \right). \tag{2.3.28}$$

The optimal portfolio $\bar{h}$ can be determined by (2.3.27) that is

$$\left(\frac{1}{1+\bar{h}} + c \right) + \left(\frac{1}{1+0.5\bar{h}} + c \right) 0.5 - \left(\frac{1}{1-0.5\bar{h}} + c \right) 0.5 = 0. \tag{2.3.29}$$

Numerically solving (2.3.28) and (2.3.29) and scaling the Lagrange multipliers yield the following table that relates c to optimal portfolio $\bar{h}$ and risk neutral measure π:

c	$\bar{h}$	π_1	π_2	π_3
0.0	0.868	.178	.232	.589
0.2	1.023	.183	.226	.591
0.4	1.154	.185	.222	.593
0.6	1.258	.189	.219	.593

Martingale measures when $w_0 = 1$.

w_0	θ^*	π_1	π_2	π_3
1	1.024	.183	.226	.591
3	3.777	.188	.218	.594
6	8.830	.192	.212	.596

Martingale measures when $c = 0.2$.

We can see that fixing w_0 when c increases so does $\bar{h}$, which is a fact that is not hard to verify to be true in general from equation (2.3.29). Note that in our family of utility functions depend on the parameter c, decreasing of c corresponding to increasing of risk aversion. On the other hand, fixing a utility function (by fixing c) decreasing of w_0 corresponds to increasing of risk aversion. This is consistent with an intuitive explanation of the change in the martingale measure: increasing in the weight in the middle (π_2) while decreasing the weight on both extremes (π_1 and π_3).

Example 2.3.13 (Pricing Contingent Claims) We now turn to pricing contingent claims. We consider the same financial market as in Example 2.3.12 defined by $S_0 = (1, 1)$ and

$$S_1(\omega_1) = (1, 2), S_1(\omega_2) = (1, 1.5), S_1(\omega_3) = (1, 0.5)$$

the payoff of a call option with strike 1 is

$$C(\omega_1) = 1, C(\omega_2) = 0.5, C(\omega_3) = 0.$$

Fixing a utility $\ln(x) + 0.2x$, pricing C using the equivalent martingale measure from the previous example gives the following results:

w_0	price	π_1	π_2	π_3
1	**0.296**	.183	.226	.591
3	**0.297**	.188	.218	.594
6	**0.298**	.192	.212	.596

Prices of a call option when $c = 0.2$.

Fixing $u(x) = \ln(x) + 0.2x, w_0 = 3$ from the table $p = 0.297$. This is the *private* price of the agent corresponding to his/her risk aversion. The meaning of this private price is that the agent should buy (long) when the market price is lower than $p = 0.297$ and sell (short) when the market price is higher to improve his/her utility as shown in Fig 2.6.

Remark 2.3.14 We can see that when market price differs from the agent's private price an opportunity of improving utility arises. However, this does not mean opportunity for arbitrage. In fact, from the graph we can see that buying (or shorting) too much will actually reduce the utility. Market price equals the agent's private price means no opportunity of improving utility. In this case the agent should take no position.

The utility optimization point of view also explains that trading will happen between agents with different risk aversion determined by utility and initial endowment. For example, if market price is 0.297, then agents with $w_0 = 1$ will sell, agents with $w_0 = 6$ will buy while agent with $w_0 = 3$ will take no action.

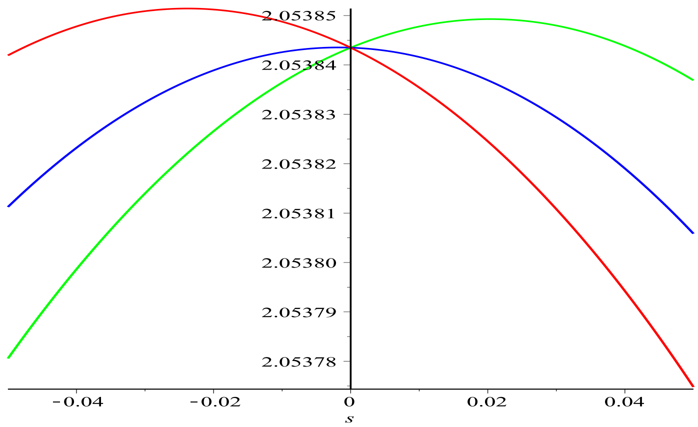


Fig. 2.6. Green $p = 0.296$, Blue $p = 0.297$ and Red $p = 0.298$

2.3.3 Complete Market

We have seen that in general the risk neutral measure is not unique and they are related to the investor's utility function. One exception is when the financial market is complete:

Definition 2.3.15 (Complete Market) *We say a financial market S is complete if*

$$\{\Theta \cdot S_1 \mid \Theta \in port[S]\} = RV(\Omega, \mathcal{F}, P),$$

or equivalently

$$\{1_B : B \in \mathcal{F}\} \subset \{\Theta \cdot S_1 \mid \Theta \in port[S]\}.$$

When a financial market is complete, it is a simple fact in linear algebra that there is only one unique equivalent martingale measure.

Proposition 2.3.16 (Unique Martingale Measure) *Let S be a complete financial market. Then there is only one unique equivalent martingale measure.*

Proof. Since $W = \{\Theta \cdot S_1 \mid \Theta \in port[S], \Theta \cdot S_0 = 0\}$, dim$W$=dim $\{\theta \cdot S_1 \mid \Theta \in port[S]\} - 1$. Thus, for a complete market dim $W^\perp$=1. Hence, in a complete market equivalent martingale measure is unique. ●

If we focus only on complete markets then utility functions are irrelevant to asset pricing. But, of course, most markets are incomplete. In a complete market the search for optimal portfolio can also be simplified.

Suppose that (x^*, Θ^*) is the solution to the constrained minimization problem (2.3.21) then it is also the solution to the problem of minimizing the Lagrangian

$$L((x, \Theta), \lambda) = \mathbf{E}[(-u)(x) + \lambda(x - w_0)] - \langle \lambda, \Theta \cdot (S_1 - S_0) \rangle.$$

which implies that $Q = \lambda/\mathbf{E}[\lambda]P$ is the unique risk neutral measure. Moreover, since x^* satisfies the constaint $x^* - \Theta^* \cdot (S_1 - S_0) - w_0 = 0$ we also know that$\langle \lambda, x^* - w_0 \rangle = \mathbf{E}^Q[x^* - w_0] = 0$. Thus, x^* is also a solution to the constrained minimization problem

$$\begin{aligned} \text{minimize} \quad & \mathbf{E}[(-u)(x)] \\ \text{subject to} \quad & \mathbf{E}^Q[x] = w_0. \end{aligned} \tag{2.3.30}$$

On the other hand, since $-u$ is strictly convex, the solution to (2.3.30) is unique and, therefore, must be x^*. Thus, problem (2.3.21) and (2.3.30) have the same solution.

Remark 2.3.17 1. Problem (2.3.30) only provides a solution x^*. To get the optimal portfolio one has to do additional work using the constraint.

2. The equivalence of the solutions of the two problem breaks down if martingale measures are not unique and, therefore the above result only holds in a complete market.

2.3.4 Use Linear Programming Duality

If we set $w_0 = 0$ then the utility optimization problem becomes

$$\sup\{\mathbf{E}[u(x)] : x \in W\}.$$

Importantly, property (u2) of the utility function forces $x \in RV(\Omega, \mathcal{F}, P)^+$ so that the problem is, in fact,

$$\sup\{\mathbf{E}[u(x)] : x \in W \cap RV(\Omega, \mathcal{F}, P)^+\}.$$

Note that no arbitrage is equivalent to

$$W \cap RV(\Omega, \mathcal{F}, P)^+ = \{0\}.$$

Thus, for the purpose of characterizing no arbitrage, the problem is trivial.

What do we get from our theory then? We still see that no arbitrage implies the existence of an equivalent martingale measure. Moreover, we still have the martingale measure is proportional to a subdifferential of the negative of the utility function at the optimal portfolio. This is where we can derive more from our approach. In this trivial problem the only solution is 0 for all economic states $\omega \in \Omega$. Since $u(t) = -\infty, t < 0$, the subdifferential of $-u$ at 0 is determined by the right directional derivative:

$$k := \lim_{t \downarrow 0} \frac{u(t) - u(0)}{t} > 0.$$

In fact,

$$-\partial(-u)(0) = [0, k]. \tag{2.3.31}$$

Since this is true for all states $\omega \in \Omega$, it tells us the equivalent martingale measure is proportional to an vector in $[0, k]^N$, $N =$ number of states in Ω. This amounts to constraint in the martingale measure. We also note that in this case nothing is lost by picking the utility function $u(t) = t - \iota_{(-\infty,0)}(t)$ so that the utility maximization problem becomes a linear programming problem. This way one can use the more widely known linear programming duality instead of Fenchel duality.

In a finite dimensional space, linear programming duality is equivalent to the Krep-Yan cone separation theorem which is used by Harrison and Kreps [23], Harrison and Pliska [22], Delbaen and Schachermayer [15] and many others in their proofs of FTAP in different settings. These approaches, however, lose the information relating to the agent's risk aversion.

2.4 Risk Measures

We have discussed variance –standard deviation and drawdown as risk measures. There are many other risk measures. To be systematic, in this section, we take an axiomatic approach: list desired properties of risk measures. We focus on coherent risk measures. It is proposed by Artzner et al. in [2] partly motivated by margin rules developed by Chicago Mercantile Exchange (CME) and Security and Exchange Commission (SEC) and have attracted much attention. From a mathematical point of view coherent risk measure is sublinear, a special type of convex function. As a result many tools in convex analysis and duality theory are applicable.

2.4.1 Coherent Risk Measure

Definition 2.4.1 (Coherent Risk Measure) *Let $RV(\Omega, \mathcal{F}, P)$ represent the payoff space. We say a lower semicontinuous function $\rho : RV(\Omega, \mathcal{F}, P) \to \mathbb{R} \cup \{+\infty\}$ is a* coherent risk measure *if, for any $x, y \in RV(\Omega, \mathcal{F}, P)$, ρ has the following properties:*

(r1) (Positive homogeneity) $\rho(rx) = r\rho(x)$ *for any* $r > 0$,
(r2) (Subadditivity) $\rho(x + y) \leq \rho(x) + \rho(y)$,
(r3) (Translation property) $\rho(x + c\vec{1}) = \rho(x) - c$ *for all* $x \in RV(\Omega, \mathcal{F}, P)$ *and* $c \in \mathbb{R}$.
(r4) (Monotonicity) $\rho(x) \leq \rho(y)$ *for any* $x \geq y$,

Propoerty (r1) says that the risk measure is proportional to scaling; Propoerty (r2) reflects the belief that diversification reduces risk; The idea of (r3) is that one may measure the risk of x by the minimum amount of additional capital reserve to ensure that there is no risk of bankruptcy; And finally (r4) says that a dominant payoff is less risky.

Not all commonly used risk measures satisfy all the requirements of a coherent risk measure. The axioms of coherent risk measures provide a set of desired properties for further discussions and comparisons. A coherent risk measure as defined above has a simple structure and afford several equivalent characterization which we will discuss below.

2.4.2 Equivalent Characterization of Coherent Risk Measures

Dual Representation

Coherent risk measure is convex. Any l.s.c. convex function on a finite dimensional Banach space has the dual representation

$$\rho(x) = \sup_{y \in RV(\Omega,\mathcal{F},P)} [\langle x, y\rangle - \rho^*(y)], \tag{2.4.1}$$

where $\langle x, y\rangle = \mathbf{E}[xy]$. What is interesting here is that ρ^* for any risk measure ρ satisfying (r1) and (r2) must be an indicator function. Properties (r3) and (r4) further restrict the support of this indicator function.

Proposition 2.4.2 (Conjugate of a sublinear risk measure) *Let ρ be a risk measure satisfying axioms (r1) and (r2). Then*

$$\rho^* = \iota_M,$$

where

$$M = \{y : \langle x, y\rangle \le \rho(x), \text{ for all } x \in RV(\Omega,\mathcal{F},P)\}.$$

Proof. Clearly, for any $y \in RV(\Omega,\mathcal{F},P)$, we have

$$\rho^*(y) = \sup_{x \in RV(\Omega,\mathcal{F},P)} [\langle x, y\rangle - \rho(x)] \ge \langle 0, y\rangle - \rho(0) = 0.$$

For any $y \in M$, $\rho^*(y)$ cannot exceeds 0 so that it must be equal to 0.

On the other hand, for any $y \notin M$, there exists $x \in RV(\Omega,\mathcal{F},P)$ such that $\langle x, y\rangle - \rho(x) \ge 0$. Since the function $x \to \langle x, y\rangle - \rho(x)$ is positive homogeneous, we must have

$$\rho^*(y) \ge \sup_{r>0} [\langle rx, y\rangle - \rho(rx)] = \sup_{r>0} r[\langle x, y\rangle - \rho(x)] = +\infty.$$

Thus,

$$\rho^* = \iota_M.$$

●

We note that the characterization of M in Proposition 2.4.2 depends on ρ. Thus we cannot use it to describe ρ. Information leads to ρ independent restriction is useful. The axioms (r3) and (r4) provide such information.

Proposition 2.4.3 (Effect of the Translation Property) *Let ρ be a risk measure satisfying (r1), (r2) and (r3). Then there exists a closed convex subset*

$$M \subset \{y \in RV(\Omega,\mathcal{F},P) : \mathbf{E}[-y] = 1\},$$

such that

$$\rho^* = \iota_M.$$

Proof. By Proposition 2.4.2 $M = \{y : \langle x, y\rangle \le \rho(x),\ \text{for all } x \in RV(\Omega, \mathcal{F}, P)$. If ρ also satisfies (r3), choose $x = \vec{1}$ and $x = -\vec{1}$, respectively we have $\mathbf{E}[y] \le -1$ and $\mathbf{E}[-y] \le 1$, respectively. Thus, $\mathbf{E}[-y] = 1$ as was to be shown. ●

Proposition 2.4.4 (Effect of Monotonicity) *Let ρ be a risk measure satisfying (r1), (r2) and (r4). Then there exists a closed convex subset*

$$M \subset -RV(\Omega, \mathcal{F}, P)^+,$$

such that

$$\rho^* = \iota_M.$$

Proof. By Proposition 2.4.2 $M = \{y : \langle x, y\rangle \le \rho(x),\ \text{for all } x \in RV(\Omega, \mathcal{F}, P)$. If ρ also satisfies (r4), then for any $y \in M$ and $x \in RV(\Omega, \mathcal{F}, P)^+$ we have $\langle x, y\rangle \le 0$ so that $y \in -RV(\Omega, \mathcal{F}, P)^+$. ●

Since the conjugate of an indicator function is a support function we derived the following characterization of a coherent risk measure.

Theorem 2.4.5 (Dual Characterization of Coherent Risk Measure) *Let ρ be a coherent risk measure. Then there exists a closed convex subset*

$$M \subset \{y \in -RV(\Omega, \mathcal{F}, P)^+ : \mathbf{E}[-y] = 1\},$$

such that

$$\rho = \sigma_M.$$

Remark 2.4.6 Coherent risk measure is directly related to cash reserve. It is a way to gauge how much cash reserve one needs to have for investing in a certain risky asset. The set $\{y \in -RV(\Omega, \mathcal{F}, P)^+ : \mathbf{E}[-y] = 1\}$ represents standardized losses because $\mathbf{E}[y] = -1$. Theorem 2.4.5 tells us a coherent risk measure is in essence picking a particular 'test' set of typical losses represented by the set M to determine the level of cash reserve for a certain investment. There are infinitely many possibilities in choosing the set M and thus determining particular coherent risk measures. The larger the set M, the more conservative the risk measure (requiring higher cash reserves). In fact, this is the original motivation for the definition of the coherent risk measure. The CME margin system is an example of using this method with a finite set M. The idea is rather similar to 'stress' test. In implementation, it is clear that what is important is not how many elements one includes in M but how 'diversified' the elements in M are.

Coherent Acceptance Cone

A second characterization for a coherent risk measure is the acceptance cone defined by

$$A_\rho = \{x \in RV(\Omega, \mathcal{F}, P) \mid \rho(x) \leq 0\}. \tag{2.4.2}$$

It is easy to check the following properties for A_ρ.

Proposition 2.4.7 *Let ρ be a coherent risk measure. Then the related acceptance cone A_ρ has the following properties:*

(a1) *A_ρ is a closed convex cone,*
(a2) *$\vec{1} \in A_\rho$,*
(a3) *$RV(\Omega, \mathcal{F}, P)^+ \subset A_\rho$.*

Proof. We merely note that (a1) is a consequence of (r1) and (r2), (a2) follows from the transitive property (r3) and (a3) is the result of monotone property (r4). Details are left as an exercise. ●

What is interesting is that and set has properties (a1)–(a3) must be the acceptance set of some coherent risk measure. This leads to the following definition.

Definition 2.4.8 (Coherent Acceptance Cone) *We say a set $A \subset RV(\Omega, \mathcal{F}, P)$ is a* coherent acceptance cone *provided that it has the following properties:*

(a1) *A is a closed convex cone,*
(a2) *$\vec{1} \in A$,*
(a3) *$RV(\Omega, \mathcal{F}, P)^+ \subset A$.*

Theorem 2.4.9 (Coherent Risk and Acceptance Cone) *Let $A \subset RV(\Omega, \mathcal{F}, P)$ be a coherent acceptance cone. Then there exists a coherent risk measure ρ_A such that*

$$A = \{x \in RV(\Omega, \mathcal{F}, P) \mid \rho_A(x) \leq 0\}.$$

Proof. The way to construct ρ_A is

$$\rho_A(x) = \inf\{t \in \mathbb{R} \mid x + t\vec{1} \in A\}.$$

All the desired properties then follow naturally. We leave checking the details as an exercise. ●

It is natural to ask the relationship between the acceptance cone and the generating set of a coherent risk measure.

Theorem 2.4.10 (Acceptance Cone and the Generating Set) *Let ρ be a coherent risk measure with a generating set M, i.e. $\rho = \sigma_M$. Let A_ρ be its acceptance cone. Then*

$$A_\rho = -(\text{cone } M)^+,$$

where cone M *is the cone generated by M, i.e. the smallest cone containing M.*

Proof. We only need to observe $x \in -(\text{cone } M)^+$ if and only if $\langle x, m\rangle \leq 0$, for all $m \in M$ iff $\rho(x) = \sigma_M(x) \leq 0$, i.e. $x \in A_\rho$. ●

Fig. 2.7 provides a graphic illustration of the relationship between M and A_ρ.

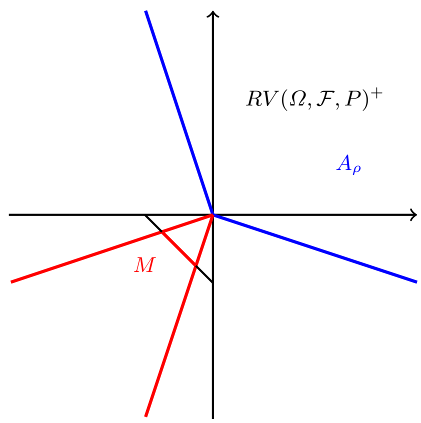


Fig. 2.7. Generating set M and acceptance set A_ρ

The coherent acceptance cone provides a dual representation of a coherent risk measure. It provides a different implementation of margin rules that are essentially the SEC methods adopted by National Association of Security Dealers (NASD). The way they implement is to consider a portfolio as consisting of a list of component securities and for each of these securities there is a corresponding margin requirement. In the language of coherent acceptance cone, this amounts to specify a set of generating elements of the cone.

Coherent Preference

We know that any closed convex cone induces a continuous partial order. Let $\leq_A$ be the linear partial order defined by an coherent acceptance cone A, that is $x \leq_A y$ if and only if $y - x \in A$.

Proposition 2.4.11 *Let A be a coherent acceptance cone and define partial order $\leq_A$ by $x \leq_A y$ if and only if $y - x \in A$. Then $\leq_A$ has the following properties:*

(o1) (Positive homogeneous) $0 \leq_A x$ *implies* $0 \leq_A tx$ *for any* $t > 0$,
(o2) (Additive) $x \leq_A y$ *and* $u \leq_A v$ *implies* $x + u \leq y + v$,
(o3) (Reflexive) $x \leq_A x$,
(o4) (Monotone) $0 \leq x$ *for any* $x \in RV(\Omega, \mathcal{F}, P)^+$.

Proof. Exercise. ●

Properties (o1)–(o4) also characterize partial order generated by a coherent acceptance set.

Definition 2.4.12 (Coherent Partial Order) *We say $\leq$ is a* coherent partial order *provided that it has the following properties:*

(o1) (Positive homogeneous) $0 \leq x$ *implies* $0 \leq tx$ *for any* $t > 0$,
(o2) (Additive) $x \leq y$ *and* $u \leq v$ *implies* $x + u \leq y + v$,
(o3) (Reflexive) $x \leq x$,
(o4) (Monotone) $0 \leq x$ *for any* $x \in RV(\Omega, \mathcal{F}, P)^+$.

Theorem 2.4.13 (Coherent Partial Order and Acceptance Cone) *Let $\leq$ be a coherent partial order. Then there exists a coherent acceptance cone A such that $x \leq y$ if and only if $y - x \in A$.*

Proof. The coherent acceptance cone can be identified as

$$A = \{x \in RV(\Omega, \mathcal{F}, P) \mid 0 \leq x\}.$$

Verifying the properties of A is not hard and is left as an exercise. ●

Valuation Bounds and Price System

Definition 2.4.14 (Valuation Bounds) *Let $\leq$ be a coherent partial order. We define the related* coherent valuation bounds, *for $x \in RV(\Omega, \mathcal{F}, P)$ by*

$$\overline{\pi}(x) = \inf\{r : x \leq r\vec{1}\} \text{ and } \underline{\pi}(x) = \sup\{r : r\vec{1} \leq x\}.$$

Definition 2.4.15 (Admissible Price) *Let $\leq$ be a coherent partial order. We say $\pi \in RV(\Omega, \mathcal{F}, P)^* = RV(\Omega, \mathcal{F}, P)$ is an* admissible price operator *if, for all $0 \leq x$,*

$$\langle \pi, x \rangle \geq 0.$$

We say π is normalized if $\pi(1) = 1$.

Definition 2.4.16 (Consistent Price) *Consider a one period financial market S on $RV(\Omega, \mathcal{F}, P)$. We say $\pi \in RV(\Omega, \mathcal{F}, P)^* = RV(\Omega, \mathcal{F}, P)$ is an* consistent price operator *for S, provided that*

$$\langle \pi, S_1 \rangle = \langle \pi, S_0 \rangle.$$

Viewing price operators as elements in the dual space is consistent with the one price principle. The definition of admissible price operators recognizes the value of any payoff $0 \leq x$, or $x \in A$ where A is the coherent acceptance cone generating the partial order $\leq$. Normalized price is consistent with the value of cash implied in the translation property of the coherent risk measure. Consistent price operator is, in fact, looking at martingale measures from the perspective of pricing system. The next proposition explains the meaning of valuation bounds and follows directly from the definition.

Proposition 2.4.17 (Bounds for Normalized Price) *Let π be a normalized admissible price operator. Then, for any $x \in RV(\Omega, \mathcal{F}, P)$,*

$$\underline{\pi}(x) \leq \langle \pi, x \rangle \leq \overline{\pi}(x).$$

Proof. Exercise. ●

While the concepts of valuation bounds and prices provide different perspectives they are closely related to the coherent risk and its equivalent description in terms of its coherent acceptance cone and coherent partial order as evidenced in the theorem below.

Theorem 2.4.18 (Valuation Bounds and Coherent Risk Measure) *Let $\leq$ be the coherent partial order generated by the coherent risk measure ρ and let $\overline{\pi}$ and $\underline{\pi}$ be the price bounds induced by the partial order $\leq$. Then, for any $x \in RV(\Omega, \mathcal{F}, P)$,*

$$\rho(x) = \overline{\pi}(-x) = -\underline{\pi}(x).$$

Proof. Consider $r \in \mathbb{R}$ with $-x \leq r\vec{1}$. We have $0 \leq x + r\vec{1}$ so that $\rho(x) - r = \rho(x + r\vec{1}) \leq 0$ or $\rho(x) \leq r$. Taking infimum over all such r we have

$$\rho(x) \leq \overline{\pi}(-x).$$

On the other hand, $\rho(x + \rho(x)\vec{1}) = \rho(x) - \rho(x) = 0$ implies that

$$\rho(x) \geq \overline{\pi}(-x).$$

The equality $\overline{\pi}(-x) = -\underline{\pi}(x)$ follows directly from definition. ●

2.4.3 Good Deal

Having a risk measure and related acceptance cone, we can consider 'good deal' opportunities of making money with some acceptable risks instead completely focusing on arbitrage.

Definition 2.4.19 (Good Deal) *Consider a one period financial market S on $RV(\Omega, \mathcal{F}, P)$. Let $port[S]$ be the portfolio space and let $W = \{\Theta \cdot (S_1 - S_0) : \Theta \in port[S]\}$ be the gain space. For a coherent acceptance cone A we say that $x \in W$ is a* good deal *with respect to A if there exists $r > 0$ such that*

$$x - r\vec{1} \in A.$$

We note that a good deal with respect to $A = RV(\Omega, \mathcal{F}, P)^+$ is an arbitrage. Thus, good deal is a relaxation of arbitrage. We have the following characterization of the existence (or absence) of a good deal.

Proposition 2.4.20 (Existence of Good Deals) *Portfolio on S contains a good deal with respect to A if and only if $\vec{1} \in W - A$. Equivalently, $port[S]$ contains no good deal with respect to A if and only if $\vec{1} \notin W - A$.*

Proof. If $\vec{1} \in W - A$ we can find $x \in W$ and $a \in A$ such that $x - \vec{1} = a \in A$. In other words, x is a good deal. One the other hand if x is a good deal then $x - r\vec{1} = a$ for some $r > 0$ and $a \in A$. Now $\vec{1} = x/r - a/r \in W - A$ as was to be shown. ●

The above characterization for the existence of good deal is from the perspective of payoffs. We now relate it to price and price bounds. Mathematically, it is a process of scalarization. What we do here is to consider the potential price of a payoff z in the market. First we discuss price bounds for a good deal.

Definition 2.4.21 (Good Deal Bounds) *Let A be a coherent acceptance cone and let $z \in W$ the gain space of financial market S. We define the upper and lower good deal bounds with respect to A by*

$$\overline{\pi}_W(z) = \inf_{r \in \mathbb{R}, x \in W} \{r : x + r\vec{1} - z \in A\}$$

and

$$\underline{\pi}_W(z) = \sup_{r \in \mathbb{R}, x \in W} \{r : x - r\vec{1} + z \in A\}.$$

As the name suggests, good deal bounds reveal prices for good deals. The interval $[\underline{\pi}_W(z), \overline{\pi}_W(z)]$ is the interval of nomalized admissible prices that consistent with the absence of a good deal. In fact, if z has a normalized admissible price $P > \overline{\pi}_W(z)$ then there exists $x = \Theta \cdot (S_1 - S_0) \in W$ and $0 < r < P$ such that $x + r\vec{1} - z \in A$, then we can sell short z at price P and assemble portfolio $\Theta \cdot S_0$ at time $t = 0$. When $t = 1$ the value of the portfolio gives us $y = x + P\vec{1} - z$. Since $y - (P - r)\vec{1} = x + r\vec{1} - z \in A$, it is a good deal.

The good deal bounds are actually coherent valuation bounds.

Proposition 2.4.22 (Good Deal Bounds as Valuation Bounds) *The upper and lower good deal bounds $\overline{\pi}_W(z)$ and $\underline{\pi}_W(z)$ defined in Definition 2.4.21 are actually coherent valuation bounds.*

Proof. It is easy to check that $\overline{\pi}_W(-z) = -\underline{\pi}_W(z)$. Moreover, rewrite $-\underline{\pi}_W(z)$ as

$$\begin{aligned} -\underline{\pi}_W(z) &= - \sup_{r \in \mathbb{R}, x \in W} \{r : x - r\vec{1} + z \in A\} \\ &= \inf_{-r \in \mathbb{R}} \{r : -r\vec{1} + z \in A - W\} \\ &= \inf_{r \in \mathbb{R}} \{r : z + r\vec{1} \in A - W\}. \end{aligned}$$

Since $A - W$ is a cone containing $RV(\Omega, \mathcal{F}, P)^+$, we can see that $-\underline{\pi}_W(z) = \rho_{A-W}(z)$ is the coherent risk measure corresponding to the coherent acceptance cone $A - W$. ●

Actually, one can show that $\rho_{A-W}(z) = \inf_{x \in W} \rho_A(x + z)$ (Exercise).

Note that the fundamental theorem of asset pricing is essentially based on the separation of W and $RV(\Omega, \mathcal{F}, P)^+$. The same argument can be applied to yield a similar result regarding good deal.

Theorem 2.4.23 (Fundamental Theorem of Asset Pricing for Good Deal) *Let A be a coherent acceptance cone and let $W = \{\Theta \cdot (S_1 - S_0) : \Theta \in port[S]\}$ be the gain space of financial market S. Then $port[S]$ contains no good deal iff there exists an admissible consistent normalized price operator.*

Proof. The portfolio space $port[S]$ contains no good deal if and only if W not intersect with the interior of A if and only if there exists $y \in RV(\Omega, \mathcal{F}, P)^* = RV(\Omega, \mathcal{F}, P)$ such that

$$\langle x, y \rangle \leq \langle a, y \rangle, \text{ for all } x \in W \text{ and } a \in A.$$

Since $0 \in W$, we have, for all $a \in A$, $\langle a, y \rangle \geq 0$. Thus, y is an admissible price. Since $0 \in A$, we have, for all $x \in W$, $\langle x, y \rangle \leq 0$. Since W is a subspace $\langle x, y \rangle = 0$ for all $x \in W$. This is equivalent to $\pi = y / \langle \vec{1}, y \rangle$ is an admissible consistent normalized price operator. ●

2.4.4 Several Commonly Used Risk Measures

Standard Deviation

Variance or equivalently standard deviation has been used as a risk measure since Markowitz proposed the modern portfolio theory. It satisfies (r1) and (r2) but fails (r3) and (r4). The standard deviation does not satisfy axiom (r4) has long been criticized as unreasonable. Some remedies have been suggested such as count the deviation only on losses. It turns out that

$$\rho_s(x) = \sqrt{\mathbf{E}[((x - \mathbf{E}[x])^-)^2)} - \mathbf{E}[x]$$

is actually a coherent risk measure that is faithful to the idea of using downside deviation as a measure for risk.

Both implementations suggested by the dual representation Theorem 2.4.5 and the acceptance cone formulation in Theorem 2.4.9 are viable. For example, if one uses the acceptance cone to implement then each security is paired with a margin requirement equals to its modified standard deviation if that can be estimated.

Drawdown

The maximum absolute drawdown, denoted $dd(x)$ in a given period of time is often used by traders. This risk measure also satisfies axioms (r1) and (r2) but fails (r3) and (r4).

As in the case of standard deviation we can also subtract $\mathbf{E}(x)$ to make it satisfy (r3). One way to adjust it so that it has property (r4) is to make the reference point for maximum down move to the fixed beginning wealth. But this completely distorts the intention of drawdown as a risk measure.

Both implementations suggested by the dual representation Theorem 2.4.5 and the acceptance cone formulation in Theorem 2.4.9 are viable without axiom (r4). The only difference is that the acceptance cone may not contain the entire cone $RV(\Omega, \mathcal{F}, P)^+$. This is not unreasonable in practice.

Value at Risk

The value at risk of a portfolio in a given period is a gauge for the risk of the portfolio that is important for both portfolio managers and regulators. It is defined on the random variable of loss $L = -x$.

Definition 2.4.24 (Value at Risk) *Let L be the random variable representing the loss of a portfolio in a given period. The value at risk with confidence level $\alpha \in (0, 1)$, denoted by VaR_α is defined as*

$$VaR_\alpha(L) = \inf\{l \in \mathbb{R} \mid P(L > l) \le 1 - \alpha\}.$$

In other words, VaR_α is a minimum level of loss which has a probability of happening $1-\alpha$. The following is an illustration.

Example 2.4.25 (VaR of a Discrete Loss Distribution) Suppose that the loss L is discretely distributed as in the following table

L	Prob
600	0.02
50	0.03
40	0.05
30	0.10
20	0.10
10	0.05
0	0.65

Table 3. A discrete loss distribution.

Then $VaR_{0.95}(L) = 50$, $VaR_{0.9}(L) = 40$, and $VaR_{0.8}(L) = 30$.

Let $F_L(l) := P(L \le l)$ be the cumulative distribution function of L. Then

$$VaR_\alpha(L) = \inf\{l \in \mathbb{R} \mid F_L(l) \ge \alpha\}.$$

We define the quantile function of L by

$$Q_L(p) = \inf\{l \in \mathbb{R} \mid p \le F_L(l)\}.$$

When F_L is an invertible function, $Q_L = F_L^{-1}$.

Value at risk as a risk measure satisfies axioms (r1) and (r4). Similar to the maximum drawdown one can adjust the cash position and define a revised version that also meets the requirement of (r3). However, missing (r2) is a big drawback for VaR as a risk measure and the remedy is complicated.

Conditional Value at Risk

Rockafellar and Uryasev [44, 45] proposed *conditional value at risk* as a remedy for VaR does not satisfy (r2).

Definition 2.4.26 (Conditional Value at Risk) *Let L be the random variable that represents the loss of a portfolio in a given period. The* conditional value at risk *with confidence level $\alpha \in (0,1)$, denoted by $CVaR_\alpha$ is defined as*

$$CVaR_\alpha(L) = \frac{1}{1-\alpha}\int_\alpha^1 VaR_s(L)ds.$$

We can see that $CVaR_\alpha$ is the expected or average loss that has a probability $1-\alpha$ of happening.

Example 2.4.27 (CVaR of a Discrete Loss Distribution) Suppose again that the loss L is discretely distributed as in Table 3. Then $CVaR_{0.95}(L) = (50 \cdot 0.03 + 600 \cdot 0.02)/0.05 = 270$, $VaR_{0.9}(L) = (40 \cdot 0.05 + 50 \cdot 0.03 + 600 \cdot 0.02)/0.1 = 155$, and $VaR_{0.8}(L) = (30 \cdot 0.1 + 40 \cdot 0.05 + 50 \cdot 0.03 + 600 \cdot 0.02)/0.2 = 92.5$.

Table 4 Compares VaR and CVaR.

L	Prob	α	VaR	CVaR
600	0.02			
50	0.03	0.95	50	270
40	0.05	0.9	40	155
30	0.10	0.8	30	92.5
20	0.10			
10	0.05			
0	0.65			

Table 4. Comparing VaR and CVaR.

We can see that VaR has the effect of give unreasonable incentive to insurance writers in general and Credit Default Swap (CDS) writers in particular.

It is not hard to see that both $VaR_\alpha(L)$ and $CVaR_\alpha(L)$ are increasing functions of α and $VaR_\alpha(L)$ is dominated by $CVaR_\alpha(L)$.

The following representation reveals that the conditional value at risk is convex with respect to L.

Theorem 2.4.28 (Representation as an Expectation)

$$\begin{aligned} CVaR_\alpha(L) &= \min_{r\in\mathbb{R}} \left\{ r + \frac{1}{1-\alpha}\mathbf{E}[(L-r)^+] \right\} \qquad (2.4.3) \\ &= VaR_\alpha(L) + \frac{1}{1-\alpha}\mathbf{E}[(L - VaR_\alpha(L))^+]. \end{aligned}$$

Proof. Note that for any r,

$$\begin{aligned}\frac{1}{1-\alpha}\mathbf{E}[(L-r)^+] &= \frac{1}{1-\alpha}\int_\Omega (L(\omega)-r)^+ P(d\omega) \\ &= \frac{1}{1-\alpha}\int_\Omega \int_r^\infty 1_{[t,\infty)}(L(\omega))dtP(d\omega) \\ &= \frac{1}{1-\alpha}\int_r^\infty \int_\Omega 1_{[t,\infty)}(L(\omega))P(d\omega)dt \\ &= \frac{1}{1-\alpha}\int_r^\infty P(L\geq t)dt.\end{aligned} \tag{2.4.4}$$

In particular (see Fig. 2.8 in which the shaded area represents $\mathbf{E}[(L-r_\alpha)^+]$), let $r = r_\alpha = VaR_\alpha(L)$ we have

$$\begin{aligned}\frac{1}{1-\alpha}\mathbf{E}[(L-r_\alpha)^+] &= \frac{1}{1-\alpha}\int_{r_\alpha}^\infty P(L\geq t)dt \\ &= \frac{1}{1-\alpha}\int_\alpha^1 (VaR_t(L)-r_\alpha)dt \\ &= \frac{1}{1-\alpha}\int_\alpha^1 VaR_t(L)dt - r_\alpha \\ &= CVaR_\alpha(L) - r_\alpha.\end{aligned} \tag{2.4.5}$$

This proves

$$CVaR_\alpha(L) = VaR_\alpha(L) + \frac{1}{1-\alpha}\mathbf{E}[(L-VaR_\alpha(L))^+].$$

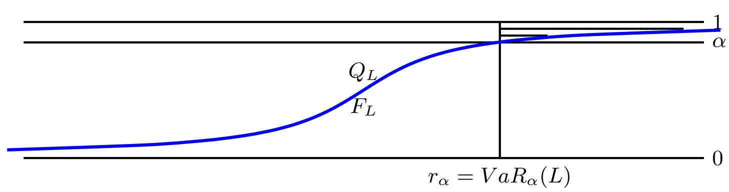


Fig. 2.8. Represent CVaR

To show that the min with respect to r is attained at $r = r_\alpha$ we define

$$\begin{aligned}D &= [r + \frac{1}{1-\alpha}\mathbf{E}[(L-r)^+]] - [r_\alpha + \frac{1}{1-\alpha}\mathbf{E}[(L-r_\alpha)^+]] \\ &= r - r_\alpha + \frac{1}{1-\alpha}\int_r^{r_\alpha} P(L\geq t)dt,\end{aligned} \tag{2.4.6}$$

and we need only to show the easy fact that, for any r,

$$(1-\alpha)D = (1-\alpha)(r-r_\alpha) + \int_r^{r_\alpha} P(L\geq t)dt \geq 0. \tag{2.4.7}$$

The intuition is illustrated in Fig. 2.9 in which the short vertial bars siginify $r < r_\alpha$ and $r > r_\alpha$, respectively. ●

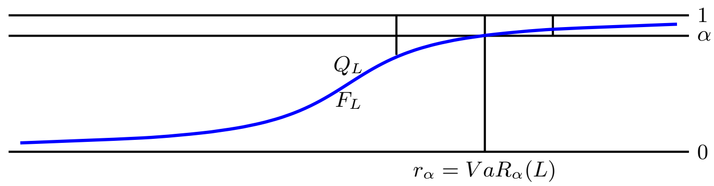


Fig. 2.9. Inequality (2.4.7).

The representation (2.4.3) can actually be written as a linear programming which yields the following dual representation.

Theorem 2.4.29 (Dual Representation)

$$CVaR_\alpha(L) = \max\left\{\langle v, -L\rangle : \mathbf{E}[-v] = 1, \vec{0} \le -v \le \frac{1}{1-\alpha}\vec{1}\right\}. \tag{2.4.8}$$

Proof. We can write the conditional value at risk with confidence level α as the value function of the following linear programming problem:

$$CVaR_\alpha(L) = \inf_{r\in\mathbb{R}, u\in RV(\Omega,\mathcal{F},P)} \{r + \frac{1}{1-\alpha}\mathbf{E}[u] : u \ge 0, u + r\vec{1} \ge L\}.$$

The Lagrangian of this linear programming problem is,

$$L((r,u),(s,v)) = r + \langle \frac{1}{1-\alpha}\vec{1}, u\rangle + \langle s, u\rangle + \langle v, u + r\vec{1} - L\vec{1}\rangle,$$

where $s, v \le 0$. For linear programming problem as long as both primal and dual problems are feasible strong duality holds. Thus, we have

$$\begin{aligned}
CVaR_\alpha(L) &= \inf_{r,u} \sup_{s\le 0, v\le 0} L((r,u),(s,v)) \\
&= \sup_{s\le 0, v\le 0} \inf_{r,u} L((r,u),(s,v)) \\
&= \sup_{s\le 0, v\le 0} \inf_{r,u} \left[r(1 + \langle v, \vec{1}\rangle) + \langle \frac{1}{1-\alpha}\vec{1} + s + v, u\rangle + \langle v, -L\rangle \right] \\
&= \sup_{s\le 0, v\le 0} \left[\langle v, -L\rangle : \langle -v, \vec{1}\rangle = 1, \frac{1}{1-\alpha}\vec{1} + s + v \ge 0 \right] \\
&= \sup \left[\langle v, -L\rangle : \mathbf{E}[-v] = 1, \frac{1}{1-\alpha}\vec{1} \ge -v \ge 0 \right].
\end{aligned}$$

Since the dual solution exists the sup is, in fact, a max. ●

As a corollary we see that $CVaR$ is essentially a coherent risk measure.

Corollary 2.4.30 *Define $\rho(x) = CVaR_\alpha(-x)$. Then ρ is a coherent risk measure.*

Estimating CVaR

The dual representation in Theorem 2.4.29 provides a method of estimating the conditional value at risk. Consider a portfolio Θ. Its corresponding gain is $\Theta \cdot R$ where $R = S_1 - S_0$ is the vector of gains of the assets in the financial market. The loss is then represented by $-\Theta \cdot r$. Now suppose $R^1, \ldots, R^m$ is a sample of the gain vector of size m, then we can estimate the expectation of the return of the portfolio Θ by

$$\mathbf{E}[\Theta \cdot R] \approx \frac{1}{m} \sum_{k=1}^{m} \Theta \cdot R^k.$$

It follows that

$$CVaR_\alpha(\Theta \cdot R) \approx \min_{r \in \mathbb{R}} \left\{ r + \frac{1}{(1-\alpha)m} \sum_{k=1}^{m} (\Theta \cdot R^k - r)^+ \right\} \tag{2.4.9}$$

Thus, by discretizing the dual representation we can estimate

$$\begin{aligned} CVaR_\alpha(\Theta \cdot R) \approx \max \Big\{ & \sum_{k=1}^{m} -v_k \Theta \cdot R^k \\ & 0 \le v_k \le \frac{1}{(1-\alpha)m}, k = 1, \ldots, m, \sum_{k=1}^{m} v_k = 1 \Big\}. \end{aligned} \tag{2.4.10}$$

We can view v_k as an alternative probability measure on the sample space $\{R^1, R^2, \ldots, R^m\}$.

3

Finite Period Financial Models

Summary. We now expand our discussion to a multi-period economy with finite status. This setting models trading in the real world quite well, where we always only deal with finite number of transactions and finite number of possible scenarios. On the technical side, both payoffs and trading strategies are still belong to finite dimensional vector spaces. The first three sections show that the key results in one period economy also holds in the more general setting of a multi-period economy. Section 4 discusses super and sub-hedging from the perspective of duality. Section 5 discusses how to model the more practical financial markets with bid and ask spreads.

3.1 The Model

3.1.1 An Example

Consider the game of bet on flipping a fair coin.

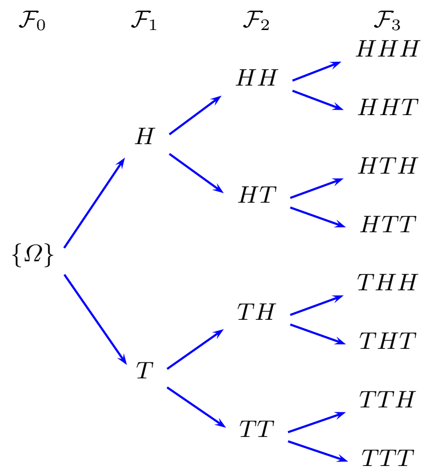
- Head: the house will double your bet.
- Tail: you lose your bet to the house.

Play the game i times and always bet 1 unit. Denote the outcome of the ith game by X_i. Then X_i is a random variable and $P(X_i = 1) = P(X_i = -1) = 1/2$. If we start with an initial endowment of w_0 then our total wealth after the ith game is

$$w_i = w_0 + X_1 + \ldots + X_i. \tag{3.1.1}$$

Now $(w_i)_{i=1}^n$ is an example of a discrete stochastic process.

We turn to consider the available information at each stage. Suppose we know $X_1, \ldots, X_i$. Does this help us to play the $(i+1)$th game? In this case we have no reason to believe so. How do we clearly describe this conclusion? Let us look at the game with $n = 3$ to get some feeling. We use H to represent a head and T, tail. The information we can get at each stage can be illustrated with the following binary tree.



In this example all the information are represented by $\mathcal{F}_3 = 2^{\Omega}$,where

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Similarly, after 2 tosses $\mathcal{F}_2 = 2^{\{HH,HT,TH,TT\}}$, where

$$\{HH, HT, TH, TT\} = \{\{HHH, HHT\}, \{HTH, HTT\}, \{THH, THT\}, \{TTH, TTT\}\}.$$

$\mathcal{F}_2$ has less information than $\mathcal{F}_3$. Similarly, $\mathcal{F}_1 = 2^{\{H,T\}}$, where

$$\{H, T\} = \{\{HHH, HHT, HTH, HTT\}, \{THH, THT, TTH, TTT\}\}.$$

At the beginning $\mathcal{F}_0 = \{\emptyset, \{\Omega\}\}$.

A random variable such as w_i relies only on information up to time i. Then, for any a, $(w_i < a) \in \mathcal{F}_i$. In other words, w_i is $\mathcal{F}_i$-measurable. We say a stochastic process $X = (X_i)$ is $\mathcal{F}$-adapted if, for each i, X_i is $\mathcal{F}_i$-measurable. The random process (w_i) in the coin toss example is $\mathcal{F}$-adapted.

3.1.2 A General Model

We continue using probability space $(\Omega, \mathcal{F}, P)$ to represent an economy where the sample space Ω is finite. Transactions now can happen in a finite set of times $\{0, 1, \ldots, T\}$ instead of only $\{0, 1\}$. Involving transactions at multiple stages requires us to be more elaborative about the information available at each of the stags. An *information structure* is a finite chain of σ-algebras of Ω: $\mathbb{F} = \{\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_T = \mathcal{F}\}$. It represents the gradually revealing information as illustrated in the previous subsection. Since Ω is finite, each $\mathcal{F}_t$ is generated by a finite number of atoms $\mathbb{B}_t = \{B_t^n, n = 1, \ldots, N_t\}$. We model a financial market with a $M + 1$-dimensional $\mathbb{F}$-adapted stochastic process $S = (S_0, S_1, \ldots, S_T)$ where $S_t = (S_t^0, S_t^1, \ldots, S_t^M)$ represents the prices of $M + 1$ assets at time t and is $\mathcal{F}_t$-measurable. Again we assume the risk free rate is 0 so that $S_t^0 = 1$.

Definition 3.1.1 (Trading Strategies) *A trading strategy $\Theta = (\Theta_0, \Theta_1, \ldots, \Theta_{T-1})$ is a $\mathbb{F}$-adapted process of $M+1$ dimensional random vectors. Each Θ_t can be viewed as a portfolio that the trader holds in the time interval $[t, t+1)$. Restricting this portfolio Θ_t to each of the atoms $B_t^n, n = 1, \ldots, N_t$ of $\mathcal{F}_t$, $\Theta_t|_{B_t^n}$ is a constant vector in $\mathbb{R}^{M+1}$ as in Definition 2.3.2 of portfolios. Two portfolios Θ_t^1 and Θ_t^2 are equivalent on market S if their restriction on all the atoms $B_t^n, n = 1, 2, \ldots, N_t$ of $\mathcal{F}$ are equivalent on S. Similarly two trading strategies are equivalent on S if all of their corresponding portfolios are equivalent on S. The quotient space of all trading strategies with respect to this equivalent relationship is called the trading strategy space on market S and is denoted by $ts[S]$. We define the norm of the portfolio Θ_t by*

$$\|\Theta_t\|_p = \sqrt{\sum_{n=1}^{N_t} \|\Theta_t|_{B_t^n}\|_p^2}$$

and the norm of a trading strategy $\Theta \in ts[S]$ by

$$\|\Theta\|_{ts} = \sqrt{\sum_{t=0}^{T-1} \|\Theta_t\|_p^2}.$$

Then $(ts[S], \|\cdot\|_{ts})$ is a finite dimensional Banach space.

In a real world of investing, the investors often face scenarios in which not all the trading strategies in $ts[S]$ are available. For example

- If short selling is not allowed, then the set of admissible trading strategies is defined by
$$ts[S]_+ = \{\Theta \in ts[S] \mid \Theta_t \geq 0, t = 0, 1, \ldots, T-1\}.$$
- If for a particular investor only a subset of the assets $\{S^0, S^1, \ldots, S^k\}$ is available, then the set of admissible trading strategies becomes
$$ts[\{S^0, S^1, \ldots, S^k\}] = \{\Theta \in ts[S] \mid \Theta_t^m = 0, m = k+1, \ldots, M, t = 0, 1, \ldots, T-1\}.$$
- Suppose a subset of the assets $S^{k+1}, \ldots, S^M$ can only be traded at $t = 0$ and $t = T$. Then the set of admissible trading strategies is defined by
$$\{\Theta \in ts[S] \mid \Theta_t^m = \Theta_0^m, m = k+1, \ldots, M, t = 1, \ldots, T-1\}.$$

By choosing different subset of $ts[S]$ we can conveniently handle different scenarios of the finite period financial model over economy $(\Omega, \mathcal{F}, P)$. We can view various questions related to these scenarios as to find suitable admissible trading strategies to obtain preferred risk adjusted gains. However, the preference will depend on the agent who is usually risk avert. By and large, there are two ways of modeling the risk aversion: using concave utility functions and using convex risk or loss functions . As a result, problems related to these financial models will be handled in the framework of maximizing expected utility functions or minimizing convex risk functions. Thus, tools in convex analysis again play essential roles.

We say a trading strategy is self-financing if

$$\Theta_{t-1} \cdot S_t = \Theta_t \cdot S_t, t = 1, 2, \ldots, T-1.$$

We use $\mathcal{T}$ to denote all self-financing trading strategies on market S. Clearly $\mathcal{T}$ is a subspace of $ts[S]$. The gain of a self-financing trading strategy Θ up to time t is the cumulative gains of portfolios $\Theta_s, s = 0, 1, \ldots, t-1$:

$$G_t(\Theta) := \sum_{s=1}^{t} \Theta_{s-1} \cdot (S_s - S_{s-1}) = \Theta_{t-1} \cdot S_t - \Theta_0 \cdot S_0.$$

We can verify that $G_t(\Theta) \in RV(\Omega, \mathcal{F}, P)$ for all $t = 1, 2, \ldots, T$.

The norm of a trading strategy is a good proxy for its *leverage level* which is very important for many purposes. As a corollary of Lemma 2.3.3 we have

Corollary 3.1.2 *There exists a constant $K = K(S)$ that depends only on market S such that for any self-financing trading strategy $\Theta \in \mathcal{T}$,*

$$\|\Theta\|_{ts} \le K \max\{\|G_t(\Theta)\|_{RV}, t = 1, 2, \ldots, T\}.$$

3.2 Arbitrage and Admissible Trading Strategies

We extend the definition of arbitrage in Definition 2.3.5 to trading strategies.

Definition 3.2.1 (Arbitrage Trading Strategy) *We say that a self-financing trading strategy Θ on market S is an arbitrage if $G_t(\Theta) \ge 0, t = 1, \ldots, T$ and $G_T(\Theta) \ne 0$.*

In every practical trading there is always a limit in how much one can lose. This leads to the concept of admissible trading strategies described below.

Definition 3.2.2 (Admissible Trading Strategy) *Let $a > 0$ be a constant. We say that a self-financing trading strategy $\Theta \in \mathcal{T}$ is* a-admissible *if, for all $t = 1, 2, \ldots, T$,*

$$G_t(\Theta) \ge -a. \tag{3.2.2}$$

We use $\mathcal{A}(a)$ to denote the (convex) set of all $a-$admissible trading strategies.

An arbitrage trading strategy is a- admissible for any $a > 0$. Thus, we have

Lemma 3.2.3 *For $a > 0$, $\mathcal{T}$ contains no arbitrage if and only if $\mathcal{A}(a)$ contains no arbitrage.*

The next lemma shows that when $\mathcal{T}$ contains no arbitrage to show Θ is a-admissible we need only to check condition (3.2.2) at $t = T$. The proof is an adaptation of the argument used in [22].

Lemma 3.2.4 *If $\mathcal{T}$ contains no arbitrage then $\Theta \in \mathcal{T}$ is a-admissible if and only if*

$$G_T(\Theta) \geq -a. \tag{3.2.3}$$

Proof. The "only if" part is obvious.

To prove the "if" part observe first that without loss of generality we may assume that the initial endowment $\Theta_0 \cdot S_0 = 0$ so that $G_t(\Theta) = \Theta_{t-1} \cdot S_t, t = 1, 2, \ldots, T$. Now assume that (3.2.3) holds and Θ is not a-admissible. Then there exist $t \leq T$ and $A \in \mathcal{F}_t$ such that on A,

$$\Theta_{t-1} \cdot S_t = b < -a$$

and $\Theta_{s-1} \cdot S_s \geq -a$ on A for all $s \geq t$.

Define a trading strategy $\bar{\Theta}$ as follows: for all $s \leq t-1$, $\bar{\Theta}_s = 0$. For $\omega \notin A$, $\bar{\Theta}_t(\omega) = 0$ and for $\omega \in A$,

$$\bar{\Theta}_t^n(\omega) = \begin{cases} \Theta_t^0(\omega) - b & for\ n = 0 \\ \Theta_t^n(\omega) & for\ n = 1, 2, \ldots, M. \end{cases} \tag{3.2.4}$$

For $s > t$ define

$$\bar{\Theta}_s^n = \begin{cases} \bar{\Theta}_t \cdot S_{t+1} & for\ n = 0 \\ 0 & for\ n = 1, 2, \ldots, M. \end{cases} \tag{3.2.5}$$

We can see that $\bar{\Theta}$ is $\mathbb{F}$-adapted. Moreover, for $\omega \in A$,

$$\begin{aligned} \bar{\Theta}_t \cdot S_t &= \Theta_t^0 - b + \sum_{n=1}^{M} \Theta_t^n S_t^n \\ &= \Theta_t \cdot S_t - b = \Theta_{t-1} \cdot S_t - b = 0 = \bar{\Theta}_{t-1} \cdot S_t. \end{aligned} \tag{3.2.6}$$

For $\omega \notin A$, $\bar{\Theta}_t \cdot S_t = 0 = \bar{\Theta}_{t-1} \cdot S_t$ by definition. For $s > t$, $\bar{\Theta}_{s-1} \cdot S_s = \bar{\Theta}_t \cdot S_{t+1}$ are pure cash and, therefore, $\bar{\Theta}$ is a self-financing trading strategy.

Finally, for all $s > t$,

$$\begin{aligned}\bar{\Theta}_{s-1} \cdot S_s &= \bar{\Theta}_t \cdot S_{t+1} \\ &= \Theta_t^0 - b + \sum_{n=1}^{M} \Theta_t^n S_{t+1}^n \\ &= \begin{cases} \Theta_t \cdot S_{t+1} - b > -a - b > 0 & \text{for } \omega \in A \\ 0 & \text{for } \omega \notin A. \end{cases} \end{aligned} \tag{3.2.7}$$

This implies that $\bar{\Theta}$ is an arbitrage, which leads to a contradiction. ●

We can also show that when there is no arbitrage the set of admissible trading strategies $\mathcal{A}(a)$ is compact.

Lemma 3.2.5 *For any $a > 0$, if $\mathcal{A}(a)$ contains no arbitrage then it is bounded and compact.*

Proof. We first show that $\mathcal{A}$ is bounded. For $t = 1, 2, \ldots, T$, let us denote $\mathcal{A}_t = \{\Theta \in \mathcal{A} : \Theta_s$ contains only cash position for $s > t - 1\}$. We note that $\mathcal{A}_T = \mathcal{A}$ and prove by induction on t. Again without loss of generality we assume the initial endowment is always 0.

For $t = 1$, assume that there is no arbitrage but $\mathcal{A}_1$ is unbounded. By Corollary 3.1.2 there exists a sequence of trading strategies $\Theta(m) \in \mathcal{A}_1$ such that $\|\Theta(m)_0 \cdot S_1\|$ is unbounded. Without loss of generality we may assume that, for all m, $\|\Theta(m)_0 \cdot S_1\| > 1$ and $\|\Theta(m)_0 \cdot S_1\| \to +\infty$ then $\Theta(m)/\|\Theta(m)_0 \cdot S_1\| \in \mathcal{A}_1$ and is bounded by Corollary 3.1.2. Selecting a subsequence if necessary we may assume that $\Theta(m)/\|\Theta(m)_0 \cdot S_1\|$ converges to $\Theta^* \in \mathcal{A}_1$. Since $\Theta(m)_0 \cdot S_1 \geq -a$, taking limit we have

$$\lim_{m\to\infty} \Theta(m)_1 \cdot S_1/\|\Theta(m)_1 \cdot S_1\|) = \Theta_1^* \cdot S_1 \geq 0.$$

On the other hand we also know from the above limiting process that $\|\Theta_1^* \cdot S_1\| = 1$. This means Θ^* is an arbitrage, a contradiction.

Now under the induction hypothesis of $\mathcal{A}_s, s = 1, 2, \ldots, t-1$ are all bounded, we show that $\mathcal{A}_t$ is bounded. Assume that the contrary holds. Then there exists a sequence of trading strategies $\Theta(m) \in \mathcal{A}_t$ such that $\|\Theta(m)_{t-1} \cdot S_t\|$ is unbounded. Since all $\mathcal{A}_s, s = 1, 2, \ldots, t-1$ are bounded, the portfolio $\Theta_{t-1}(m)$ must be unbounded. Then the same argument as in the case of $t = 1$ will yield a contradiction. This completes the induction proof and, therefore, $\mathcal{A}$ is bounded.

Since $\Theta_t \cdot S_t$ is continuous in Θ_t, $\mathcal{A}$ defined by constraint (3.2.2) is also closed and, therefore, it is compact.

●

3.3 Fundamental Theorem of Asset Pricing

Now we turn to prove the FTAP in multiperiod market model and discuss related applications.

3.3.1 Fundamental Theorem of Asset Pricing

As in the case of $T = 1$, we prove the FTAP by considering a pair of dual convex programming problems in which the primal is maximizing utility among admissible trading strategies:

$$\sup\{\mathbf{E}[u(\Theta_{T-1} \cdot S_T)] : \Theta_0 \cdot S_0 = w_0, \Theta \in ts[S]\}. \tag{3.3.8}$$

We show that a solution to the dual of (3.3.8) when scaled gives us a martingale measure and, thus, linking the fundamental theorem of asset pricing to utility maximization problem (3.3.8).

Theorem 3.3.1 *Let S be a financial market. Then the following are equivalent:*

(i) *There exists no arbitrage trading strategy in $ts[S]$;*
(ii) *For every utility function u with properties (u1),(u2), and (u3), the finite optimal value of the trading strategy utility optimization problem (3.3.8) is attained.*
(iii) *There is an equivalent S-martingale measure proportional to an element of the subdifferential of the utility function at the optimal portfolio.*

Proof. First observe that the utility optimization problem (3.3.8) can be written equivalently as

$$\begin{aligned} \max \quad & \mathbf{E}[u(y)] \\ \text{subject to} \quad & y \in w_0 + W, \end{aligned} \tag{3.3.9}$$

where $W = \{G_T(\Theta) : \Theta \in \mathcal{T}\}$ is the linear subspace of all achievable gains using self-financing trading strategies.

Defining $f(y) = -\mathbf{E}[u(y)]$ and $g(y) = \iota_{w_0+W}(y)$, we can rewrite problem (3.3.9) as

$$-\min_y\{f(y) + g(y)\} \tag{3.3.10}$$

The dual problem of (3.3.10) is,

$$\begin{aligned} & -\max\{-f^*(-z) - g^*(z)\} \\ = & \min\,\{\mathbf{E}[(-u)^*(-z)] + \langle z, w_0\rangle + \sigma_W(z)\} \end{aligned} \tag{3.3.11}$$

Since we can check that the constraint qualification condition

$$w_0 \in \text{int cont} f \cap \text{dom} g = RV(\Omega, \mathcal{F}, P)^+ \cap (w_0 + W) \tag{3.3.12}$$

holds , (3.3.10) and its dual (3.3.11) have the same value.

When $\mathcal{T}$ contains no arbitrage, by property (u2) of the utility function, $\mathbf{E}[u(\Theta_{T-1} \cdot S_T)] > -\infty$ implies $\Theta_{T-1} \cdot S_T \geq 0$ or $G_T(\Theta) \geq -w_0$. By Lemma 3.2.4, we must have $\Theta \in \mathcal{A}(w_0)$. Thus, the utility maximization problem (3.3.8) is equivalent to

$$\sup\{\{\mathbf{E}[u(\Theta_{T-1} \cdot S_T)] : \Theta_0 \cdot S_0 = w_0, H \in \mathcal{A}(w_0)\}. \tag{3.3.13}$$

By Lemma 3.2.5 problem (3.3.13) and, therefore, (3.3.9) has a finite solution. By the strong duality, the dual problem (3.3.11) has a finite optimal value and attains its solution. Condition (u2) forces the domain of $\mathbf{E}[(-u)^*(\cdot)]$ to be a subset of int $(-RV(\Omega, \mathcal{F}, P)_+)$. Thus, we only need to consider $z > 0$ in the dual problem (3.3.11). Moreover, we must have $\langle z, G_T(\Theta)\rangle = 0$ in (3.3.11) since $\sigma_W(z) < \infty$ and W is a subspace of $RV(\Omega, \mathcal{F}, P)$. Hence we can write problem (3.3.11) as

$$\min \{\mathbf{E}[(-u)^*(-z)] + \langle w_0, z\rangle \mid z > 0,\ \langle z, G_T(\Theta)\rangle = 0,\ \text{for all } \Theta \in \mathcal{T}\}. \tag{3.3.14}$$

Let $\bar{z}$ be a solution to (3.3.14) it is easy to check that $Q = (\bar{z}/\mathbf{E}[\bar{z}])P$ is an equivalent S-martingale measure. Thus, (i) implies (ii).

On the other hand, the existence of an equivalent S-martingale measure implies that the dual problem (3.3.11) has a finite value and, therefore is equivalent to problem (3.3.14) whose dual is the utility maximization problem (3.3.8). Problem (3.3.14) can be viewed as minimizing the convex function $z \to \mathbf{E}((-u)^*(-z)) + \langle w_0, z\rangle$ over the entire subspace $\{z : \langle z, G_T(\Theta)\rangle = 0,\ \text{for all } \Theta \in \mathcal{T}$ ($z > 0$ is merely a consequence of the domain of $\mathbf{E}[(-u)^*(\cdot)]$ being a subset of int $-RV(\Omega, \mathcal{F}, P)_+$ and, therefore, is not a separate constraint). Thus, the constrain qualification condition for (3.3.14) satisfies (see e.g. [65, Theorem 2.7.1]). It follows that problem (3.3.8) as the dual of (3.3.14) has a finite value and attains its solution, which is to say that (ii) implies (iii).

Finally, if (iii) is true then there cannot be any arbitrage in $\mathcal{T}$ because adding an arbitrage to the optimal solution of (3.3.8) will improve it. Thus, (iii) implies (i) and we have completed a cyclic proof of the equivalence of (i), (ii) and (iii). ●

3.3.2 Relationship between Dual of Portfolio Utility Maximization, Lagrange Multiplier and Martingale Measure

Although no arbitrage is equivalent to the existence of an equivalent martingale measure is well known, as pointed out in [67] the proof of Theorem 3.3.1 using a class of utility functions says more. It tells us that the risk neutral measure is, in fact, a scaling of the solution to the dual of the portfolio utility maximization problem. Moreover, since the dual solution corresponding to the Lagrange multipliers of the primal portfolio utility maximization problem (see [9]), we see that the equivalent martingale measure can also be explained as the scaling of the Lagrange multiplier of the portfolio utility maximization problem.

To see this relationship explicitly, let us write the utility optimization problem (3.3.8) as

$$\inf\{\mathbf{E}[(-u)(x)] : x - G_T(\Theta) - w_0 = 0, \Theta \in \mathcal{T}\}. \tag{3.3.15}$$

The existence of the solution to the dual of (3.3.15) implies the existence of a Lagrange multiplier $\lambda \in RV(\Omega, \mathcal{F}, P)$ such that the Lagrangian

$$\begin{aligned} L((x,\Theta),\lambda) &= \mathbf{E}[(-u)(x)] + \langle \lambda, x - G_T(\Theta) - w_0 \rangle \\ &= \mathbf{E}[(-u)(x) + \lambda(x - w_0)] - \langle \lambda, G_T(\Theta) \rangle \end{aligned}$$

attains minimum at solution (x^*, Θ^*) to the problem (3.3.8). It follows that, for any $P(\omega) \neq 0$,

$$\lambda(\omega) \in -\partial(-u)(x^*(\omega)) \subset (0, +\infty) \tag{3.3.16}$$

and, since $\Theta \mapsto \langle \lambda, G_T(\Theta) \rangle$ is linear,

$$\langle \lambda, G_T(\Theta) \rangle = 0, \text{ for all } \Theta \in \mathcal{T}. \tag{3.3.17}$$

It is easy to deduce from (3.3.17) that $\mathbf{E}[\lambda(S_t - S_{t-1}) \mid \mathcal{F}_{t-1}] = 0$. Thus, $Q = (\lambda/\mathbf{E}[\lambda])P$ is a martingale probability measure for market S equivalent to P.

3.3.3 Pricing Contingent Claims

Suppose that a contingent claim can only be traded at $t = 0$ and $t = T$ and its payoff at time $t = T$ is $\phi(S_T)$. To find out a reasonable price ϕ_0 for this contingent claim at time $t = 0$, we can again consider the portfolio utility optimization problem

$$\begin{aligned} &\text{minimize} && \mathbf{E}[(-u)(x)] \\ &\text{subject to} && x - \beta(\phi(S_T) - \phi_0 + G_T(\Theta)) - w_0 = 0, \\ &&& \Theta \in \mathcal{T}. \end{aligned} \tag{3.3.18}$$

Using the same arguement as in the previous subsection, we can show that there exists a Lagrange multiplier $\lambda \in RV(\Omega, \mathcal{F}, P)$ such that, for any $P(\omega) \neq 0$,

$$\lambda(\omega) \in -\partial(-u)(x^*(\omega)) \subset (0, +\infty)$$

and $Q = (\lambda/\mathbf{E}[\lambda])P$ is a martingale probability measure for market S equivalent to P. Moreover,

$$\phi_0 = \mathbf{E}^Q[\phi(S_T)]. \tag{3.3.19}$$

Formula (3.3.19) indicates that the martingale measure used to pricing a contingent claim is, in general, relies on the risk aversion of an agent. Thus, in an incomplete market, agents with different risk aversions and, therefore, different utility functions may reasonably price the same contingent differently. This is certainly consistent with the reality of the markets.

3.3.4 Solution to the Utility Optimization Problem

The discussion in section 2.3.3 can be extended to multi-period model.

Theorem 3.3.2 *Suppose that equivalent martingale measure Q on market S is unique and S has no arbitrage. Then portfolio optimization problem (3.3.15) is equivalent to*

$$\begin{aligned} \text{minimize} \quad & \mathbf{E}[(-u)(x)] \\ \text{subject to} \quad & \mathbf{E}^Q(x) = w_0. \end{aligned} \tag{3.3.20}$$

As we have seen in the one period case this is merely calculating the optimal end wealth using the Lagrangian. Proof is similar to that of the one period case and is omitted.

3.4 Hedging and Super Hedging

If the market price of an asset violates those specified by the fundamental theorem of asset pricing then in theory an arbitrage opportunity arises. We turn to the problem of how to take advantage of such an arbitrage opportunity.

3.4.1 Super- and Sub-hedging Bounds

Consider an European style contingent claim whose payoff at T is $\psi(S_T)$. By the fundamental theorem of asset pricing, the price of ψ at $t = 0$ must belong to the set $\{\mathbf{E}^Q[\psi(S_T)] : Q \in \mathcal{M}\}$ to be arbitrage free. Here $\mathcal{M}$ is the set of all martingale measures equivalent to P. It follows that

$$\overline{\psi} = \sup\{\mathbf{E}^Q[\psi(S_T)] : Q \in \mathcal{M}\} \tag{3.4.1}$$

and

$$\underline{\psi} = \inf\{\mathbf{E}^Q[\psi(S_T)] : Q \in \mathcal{M}\} \tag{3.4.2}$$

give us upper and lower bounds for the price of ψ. If the price of ψ fells outside of these bounds, an arbitrage will become possible. We call them super- and sub-hedging bounds, respectively. We focus on the super-hedging bound. The discussion about the sub-hedging bound can be reduced to that of a super hedging bound for $-\psi$ because

$$-\underline{\psi} = \sup\{\mathbf{E}^Q[-\psi(S_T)] : Q \in \mathcal{M}\}. \tag{3.4.3}$$

If the market price of ψ is above this super hedging bound how can we find an arbitrage strategy? It turns out that the key is to view (3.4.1) as a linear programming problem and consider its dual. As discussed before that for a linear programming problem and its dual, the constraint qualification condition ensuring the strong duality is, in fact, the feasiblity condition. So the key is to correctly formulate the dual problem of (3.4.1). We will use the Lagrange formulation. Let's assume $\{\Theta_n\}_{n=1}^N$ is a bases for the finite dimensional Banach space $ts[S]$. Then we can rewrite (3.4.1) as

$$\overline{\psi} = \sup_{Q \in M^+} \{\mathbf{E}^Q[\psi(S_T)] : \mathbf{E}^Q[G_T(\Theta)] = 0, \mathbf{E}^Q[1] = 1, \Theta \in ts[S]\} \tag{3.4.4}$$
$$= \sup_{Q \in M^+} \{\mathbf{E}^Q[\psi(S_T)] : \mathbf{E}^Q[1] = 1, \mathbf{E}^Q[G_T(\Theta_n)] = 0, n = 1, \ldots, N\},$$

where M^+ signifies the set of all positive measures. We can see that (3.4.4) is a linear programming problem. Moreover, the Lagrangian of (3.4.4) is

$$L(Q, \lambda) = \mathbf{E}^Q[\psi(S_T)] + \sum_{n=1}^{N} \lambda_n \mathbf{E}^Q[G_T(\Theta_n)] + \lambda_0(\mathbf{E}^Q[1] - 1), \tag{3.4.5}$$

where $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_N) \in \mathbb{R}^{N+1}$ is the Lagrange multiplier. Observe that elements $\Theta \in ts[S]$ can be represented as

$$\Theta = \sum_{n=1}^{N} \lambda_n \Theta_n$$

we can equivalently view (Θ, λ_0) as a Lagrange multiplier of the linear programming problem (3.4.4) and write the Lagrangian as,

$$L(Q, (\Theta, \lambda_0)) = \mathbf{E}^Q[\psi(S_T)] + \mathbf{E}^Q[G_T(\Theta)] + \lambda_0(\mathbf{E}^Q[1] - 1), \tag{3.4.6}$$

where $(\Theta, \lambda_0) \in ts[S] \times \mathbb{R}$. It is easy to verify that

$$\inf_{(\Theta, \lambda_0)) \in ts[S] \times \mathbb{R}} L(Q, (\Theta, \lambda_0)) = \begin{cases} \mathbf{E}^Q[\psi(S_T)] & Q \in \mathcal{M} \\ -\infty & \textit{otherwise}. \end{cases}$$

Thus, we can write

$$\overline{\psi} = \sup_{Q \in M^+} \inf_{(\Theta, \lambda_0) \in ts[S] \times \mathbb{R}} L(Q, (\Theta, \lambda_0)) \tag{3.4.7}$$

and by strong duality we have

$$\overline{\psi} = \inf_{(\Theta, \lambda_0) \in ts[S] \times \mathbb{R}} \sup_{Q \in M^+} L(Q, (\Theta, \lambda_0)) \tag{3.4.8}$$
$$= \inf_{\Theta \in ts[S]} \sup_{Q \in M^+} \{\mathbf{E}^Q[\psi(S_T) + G_T(\Theta)], \mathbf{E}^Q[1] = 1\}$$
$$= \inf_{\Theta \in ts[S]} \sup_{\omega \in \Omega} \{\psi(S_T)(\omega) + G_T(\Theta)(\omega)\}$$

The financial interpretation of the last expression in (3.4.8) is that a solution to problem (3.4.8), if exists, is a trading strategy that results in a payoff that is always bounded by the super-hedging bound. Thus, if the market price exceeds the super-hedging bound one has an arbitrage strategy.

The arbitrage trading strategy alluded to above can be found by solving the linear programing problem

$$\begin{aligned} \min \quad & t \\ s.t. \quad & t - G_T(\Theta)(\omega) \geq \psi(S_T)(\omega), \omega \in \Omega \\ & \Theta \in ts[S], t \in \mathbb{R}. \end{aligned} \tag{3.4.9}$$

Let $\bar{\Theta}$ and $\bar{t} = \overline{\psi}$ be the solution of (3.4.9). If the market price of the contingent claim at $t = 0$ is

$$\psi_0 > \overline{\psi}.$$

Then we can short one share of the contingent claim and follow the trading strategy $-\Theta$ (or equivalently, short the trading strategy Θ). By time $t = T$, we have

$$\bar{t} - G_T(\bar{\Theta})(\omega) \geq \psi(S_T)(\omega), \text{ for all } \omega \in \Omega.$$

That is to say the gain from the trading and cash amount $\overline{\psi}$ safely covers the short position in any possible economic state and the difference $\psi_0 - \overline{\psi}$ becomes our arbitrage profit.

3.4.2 Towards a Complete Market

If we know the prices of some European contingent claims, say $\phi_1, \ldots, \phi_K$ at $t = 0$ to be $c_1, \ldots, c_K$, respectively. Then to avoid arbitrage the estimate of the upper bound for a contingent claim ψ is

$$\sup\{\mathbf{E}^Q[\psi] : Q \in \mathcal{M}, \mathbf{E}^Q[\phi_k] = c_k, k = 1, \ldots, K\}. \tag{3.4.10}$$

Denote $c = (c_1, \ldots, c_K)$ and $\phi = (\phi_1, \ldots, \phi_K)$ we can write the Lagrangian of the constrained optimization problem (3.4.10) as

$$L(Q, (\Theta, \lambda_0, b)) = \mathbf{E}^Q[\psi(S_T)] + \mathbf{E}^Q[G_T(\Theta)] + \lambda_0(\mathbf{E}^Q[1] - 1) + b \cdot (\mathbf{E}^Q[\phi(S_T)] - c),$$

where $(\Theta, \lambda_0, b) \in ts[S] \times \mathbb{R} \times \mathbb{R}^K$.

Similar to the previous section we can verify that, by the strong lagrange duality,

$$\begin{aligned} \overline{\psi}|_\phi &= \inf_{(\Theta,\lambda_0,b)\in ts[S]\times\mathbb{R}\times\mathbb{R}^K} \sup_{Q\in M^+} L(Q, (\Theta, \lambda_0, b)) \\ &= \inf_{(\Theta,b)\in ts[S]\times\mathbb{R}^K} \sup_{Q\in M^+} \{\mathbf{E}^Q[\psi(S_T) + G_T(\Theta) + b \cdot (\phi - c)], \mathbf{E}^Q[1] = 1\} \\ &= \inf_{(\Theta,b)\in ts[S]\times\mathbb{R}^K} \sup_{\omega\in\Omega} \{\psi(S_T)(\omega) + G_T(\Theta)(\omega) + b \cdot (\phi(S_T)(\omega) - c)\}. \end{aligned} \tag{3.4.11}$$

The financial interpretation of the last expression in (3.4.11) is that a solution to problem (3.4.11), if exists, is a trading strategy that results in a payoff that is always bounded by the super-hedging bound. Thus, if the market price exceeds the super-hedging bound one has an arbitrage stratgy, which can be calculated using a liner programming problem similar to that of in (3.4.9).

Here with the additional tradable contingent claims $\phi_1, \ldots, \phi_K$, the upper bound for the no arbitrage price is lowered and correspondingly the lower bound will be increased so that we get a more accurate estimate of the price. If we add enough additional contingent claims as the tradable, the market eventually becomes complete in the sense that the upper and lower bounds will coincide to give us a unique price. There are many ways to characterize a complete market. In the context here the most direct way is to require the subspace

$$W = \{G_T(\Theta) + b \cdot (\phi - c) \mid (\Theta, b) \in ts[S] \times \mathbb{R}^K\} \tag{3.4.12}$$

of $RV(\Omega, \mathcal{F}, P)$ has a codimension 1 (the dimension of W is exactly 1 less than that of $RV(\Omega, \mathcal{F}, P)$).

3.4.3 Incomplete Market Arise from Complete Markets

We turn to consider an incomplete market arises from complete markets. A motivating example is a call option on a currency spread. For simplicity let us consider a one period economy where transactions take place at $t = 0$ and $t = 1$.The payoff of a call option on the spread of two different currencies C^1, C^2 with a strike K in terms of a third currency at $t = 1$ is then

$$(C_1^1 - C_1^2 - K)^+. \tag{3.4.13}$$

Since C^1 and C^2 are different currencies, it is reasonable to model their value in terms of the common currency at time $t = 1$ as random variables in two different probability spaces $(\Omega_1, \mathcal{F}^1, P_1)$ and $(\Omega_2, \mathcal{F}^2, P_2)$, respectively. We assume that both markets for C^1 and C^2 are complete. Moreover, we assume that P_i is the unique martingale measure for $C^i, i = 1, 2$. If we consider (3.4.13) to be a special form of the more general contingent claim $\psi = \psi(C_1^1, C_1^2)$, then ψ is a random variable on the product measure space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$. Our problem now is to seek a martingale measure π on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$, which prices ψ so as to consistent with the martingale measures P_1 and P_2, respectively. Consider a contingent claim $\phi^1(C^1)$ that depends only on C^1. We can view this payoff both as a random variable on $(\Omega_1, \mathcal{F}^1, P_1)$ and as a random variable on $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \pi)$. Thus requiring π to be consistent with P_1 is to require

$$\int_{\Omega_1} \phi^1(C_1^1) dP_1 = \int_{\Omega_1 \times \Omega_2} \phi^1(C_1^1) d\pi. \tag{3.4.14}$$

Since $\phi^1(C_1^1)$ is arbitrary this is to say that P_1 is the marginal probability measure of π on Ω_1. Similarly, P_2 must be the marginal probability measure of π on Ω_2. Clearly, product measure π that satisfies such marginal requirements is not unique. We see that despite the completeness of the financial markets on Ω_1 and Ω_2, in pricing a contingent claim with payoff as a random variable on the product measure space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$, we face an incomplete market.

To find the upper bound for the price of ψ that is consistent with the no arbitrage principle we face the optimization problem

$$\bar{\psi} = \sup_{\pi \in \Pi(P_1, P_2)} \mathbf{E}^\pi[\psi], \tag{3.4.15}$$

where $\Pi(P_1, P_2)$ signifies the set of all probability measures on the product measure space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ whose marginals on Ω_1 and Ω_2 are P_1 and P_2, respectively.

Problem (3.4.15) turns out to be a Kantorovich mass transport problem. By Kantorovich'es duality theorem (a special case of the abstract linear programming duality) we have

$$\bar{\psi} = \sup_{\pi \in \Pi(P_1, P_2)} \mathbf{E}^\pi[\psi] = \min_{(\phi_1, \phi_2) \in G_\psi} \left(\mathbf{E}^{P_1}[\phi^1] + \mathbf{E}^{P_2}[\phi^2] \right), \tag{3.4.16}$$

where $G_\psi := \{(\phi^1, \phi^2) \in (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2) : \phi^1(\omega_1) + \phi^2(\omega_2) \geq \psi(C_1^1(\omega_1), C_1^2(\omega_2))\}$. The Kantorovich duality (3.4.16) shows that in principle one can implement the upper no arbitrage price bound $\bar{\psi}$ using the sum of two contingent claims ϕ^1 and ϕ^2 on sample spaces Ω_1 and Ω_2, respectively.

In this concise introduction we cannot afford a detailed discussion of the Kantorovich duality theorem. Instead we will exam the case when both sample spaces Ω_1 and Ω_2 are finite. In this case problem (3.4.15) reduces to a linear programming problem. We can achieve the decoupling alluded to in the Kantorovich duality theorem by directly using linear programming duality.

Example 3.4.1 (Estimate Upper No Arbitrage Bound in Finite Sample Spaces)
Suppose that both sample spaces Ω_1 and Ω_2 are finite. Denote $\Omega_1 = \{i : i = 1, \dots, L\}$ and $\Omega_2 = \{j : j = 1, \dots, M\}$, respectively. For briefty of the notation we denote

$$\psi_{ij} = \psi(C^1(i), C^2(j)).$$

Then the problem of finding an upper bound for the contingent claim $\psi(C^1, C^2)$ can be formulated as

$$\begin{aligned} \max \quad & \sum \psi_{ij}\pi_{ij} \\ s.t. \quad & \sum_j \pi_{ij} - \mu_i = 0, \ \sum_i \pi_{ij} - \nu_j = 0 \\ & \sum_i C_1^1(i)\mu_i = C_0^1, \ \sum_j C_1^2(j)\nu_j = C_0^2 \\ & \sum_i \mu_i = 1, \ \sum_j \nu_j = 1. \end{aligned} \tag{3.4.17}$$

The dual of the linear programming problem (3.4.17) is

$$\begin{aligned} \min \quad & \lambda_1 C_0^1 + \lambda_2 C_0^2 + \lambda_3 + \lambda_4 \\ s.t. \quad & u_i + v_j \geq \psi_{ij} \\ & \lambda_1 C_1^1(i) + \lambda_3 - u_i \geq 0 \\ & \lambda_2 C_1^2(j) + \lambda_4 - v_j \geq 0. \end{aligned} \tag{3.4.18}$$

Defining $\phi^1(C^1) = \lambda_1 C^1 + \lambda_3$ and $\phi^2(C^2) = \lambda_2 C^2 + \lambda_4$ we can rewrite (3.4.18) as

$$\begin{aligned} \min \quad & \phi^1(C_0^1) + \phi^2(C_0^2) \\ s.t. \quad & \phi^1(C_1^1(i)) + \phi^2(C_1^2(j)) \geq \psi_{ij} \end{aligned} \tag{3.4.19}$$

which is the Kantorovich dual form of (3.4.17). Note that ϕ^1 and ϕ^2 are linearly depend on C^1 and C^2, respectively. Thus, problem (3.4.19) is a linear programming problem.

3.5 Conic Finance

Real financial markets have frictions. Trading a financial asset one faces two different prices: ask and bid. Usually, the ask is strictly larger than the bid and one can only buy at the ask price and sell at the bid price. This violation of the one price principle complicates the modeling. The attainable gains from trading assets in such a more realistic market model is not a subspace but rather, in general, a cone. This leads to the name of conic finance.

3.5.1 Modeling Financial Markets with an Ask-Bid Spread

Let $\mathbb{F} = \{\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_T = \mathcal{F}\}$ be an information structure on the probaility space $(\Omega, \mathcal{F}, P)$ with a finite sample space that represents the economic states. Denote $\mathcal{X}$ the space of all $\mathbb{F}-$adapted cash streams $x = (x_t)_{t=0}^T$ endowed with the inner product

$$\langle x, y \rangle = \mathbf{E}\left[\sum_{t=0}^{T} x_t y_t\right].$$

Then $\mathcal{X}$ is a finite dimensional Hilbert space.

A financial market consists of M risky assets $S^m \in \mathcal{X}, m = 1, 2, \ldots, M$ and T riskless bonds $1^u, u = 1, 2, \ldots, T$ where $1^u_u = 1$ and $1^u_t = 0$ for $t \neq u$. At time t, to trade the rights to the *income stream* of S^i after t, there is a bid and ask price paire:

$$b^i_t \leq a^i_t \tag{3.5.1}$$

Thus, paying a^i_t one will get the income stream $(S^i_s)_{s=t+1}^T$. Similarly, receiving b^i_t one sells the income stream $(S^i_s)_{s=t+1}^T$ or in other words get the income stream $(-S^i_s)_{s=t+1}^T$. Considering the market friction makes the model more complicated comparing to the one price model we have used so far. A risky asset in a one price economy is described buy its prices (as random variables) at $T+1$ trading times. In a economy with bid and ask spread to describe a risky asset we need to use $2(T+1)$ income streams with corresponding prices. Similarly, the description of riskless asset also becomes more involved. We regard them as a series of riskless bonds maturing at time $u = 1, 2, \ldots, T$ whose payment streams are 1 when $t = u$ and 0 when $t \neq u$ and is denoted 1^u. The bid and ask prices of 1^u at $t < u$ are denoted g^u_t and h^u_t, respectively. They satisfy the following inequality

$$g^u_t \leq h^u_t. \tag{3.5.2}$$

A convenient way of thinking the trading of these income streams is to incorporate the buying cost or selling revenue into the income streams to view the resulting income streams as zero cost money streams. For example, the action of buying income stream $(S^i_s)_{s=t+1}^T$ at time t with ask price a^i_t is equivalent to acquiring the zero cost money stream S^{it} defined by

$$S^{it}_s = \begin{cases} 0 & s < t \\ -a^i_t & s = t \\ S^i_s & s > t. \end{cases} \tag{3.5.3}$$

Selling the above income stream at the bid price b_t yields the zero cost income stream $\tilde{S}^{it}$ defined by

$$\tilde{S}^{it}_s = \begin{cases} 0 & s < t \\ b^i_t & s = t \\ -S^i_s & s > t. \end{cases} \tag{3.5.4}$$

We observe that $\tilde{S}^{it}$ is different from $-S^{it}$ due to the spread between the ask and bid prices. Similarly the bond maturing at u generates zero cost income streams

$$1^{ut}_s = \begin{cases} 0 & s \neq u, t \\ -h^u_t & s = t \\ 1 & s = u, \end{cases} \quad \text{and } \tilde{1}^{ut}_s = \begin{cases} 0 & s \neq u, t \\ g^u_t & s = t \\ -1 & s = u. \end{cases} \tag{3.5.5}$$

Assuming that one can buy or sell any fraction of the cash stream alluded to above, suppose $\alpha^i_t, \tilde{\alpha}^i_t, \beta^u_t, \tilde{\beta}^u_t, i = 1, \ldots, M, u = 1, \ldots, T$ are nonnegative $\mathcal{F}_t$ measurable random variables, then

$$z = \sum_{t=0}^{T} \sum_{i=1}^{M} [\alpha^i_t S^{it} + \tilde{\alpha}^i_t \tilde{S}^{it}] + \sum_{t=0}^{T} \sum_{u=1}^{T} [\beta^u_t 1^{ut} + \tilde{\beta}^u_t \tilde{1}^{ut}], \tag{3.5.6}$$

is a cash stream that can be implemented by trading the available zero cost cash streams. Denote Z the collection of all cash streams of the form in (3.5.6). It is clear that Z is a cone. Define C to be the set of all cash streams $c \in \mathcal{X}$ such that there exists a $z \in Z$ with $z \geq c$, that is, C is all the cash streams that can be dominated by a cash stream in Z. Then it is easy to see that C is also a cone and $Z \subset C$. In general, C maybe larger than Z. The set C represents all the cash streams that can be dominated by a corresponding cash stream in Z, which can be implemented by using zero cost cash streams in the market by involving appropriate trading. For any $c \in C$, we can find $z \in Z$ defined by (3.5.6) such that $z \geq c$. We say $\alpha^i_t, \tilde{\alpha}^i_t, \beta^u_t$, and $\tilde{\beta}^u_t$ is a trading strategy that super implements c.

We note that when S^i is a cash stream that pays S^i_T with ask and bid prices at t both coincide with S^i_t and all $g^u_t = h^u_t = 1$, we recover the one price financial markets defined before as a special case.

3.5.2 Characterization of No Arbitrage by Utility Optimization

Using the model described in the previous section, we can extend the fundamental theorem of asset pricing to markets with a bid -ask spread. First we define arbitrage in such a market.

Definition 3.5.1 (Arbitrage Trading Strategy) *We say that a cash stream $c \in C\backslash\{0\}$ is an arbitrage if $c_t \geq 0$ for all $t = 0, 1, \ldots, T$.*

Denote $\mathcal{X}^+$ the cone in $\mathcal{X}$ with all the components are nonnegative, then there is no arbitrage trading strategy in the financial market described in the previous section if and only if

$$C \cap \mathcal{X}^+ = \{0\}. \tag{3.5.7}$$

Let u be a utility function satisfying the conditions (u1)–(u3). We consider the optimal trading problem

$$p = \max\left\{\sum_{t=0}^{T} \mathbf{E}[u(c_t)] : c \in w^0 + C\right\}, \tag{3.5.8}$$

where $w^0 \in \mathcal{X}^+$ is an initial endowment cash stream. We can characterize the no arbitrage in terms of the optimal trading problem (3.5.8):

Theorem 3.5.2 (No Arbitrage and Utility Maximization) *The financial market described in the previous section has no arbitrage trading strategy if and only if the optimal trading problem (3.5.8) has a finite optimal value $p < \infty$.*

Proof. Since one can always scale an arbitrage trading strategy with any arbitrarily large positive number, therefore $p < +\infty$ implies that there is no arbitrage trading strategy. On the other hand if $p = +\infty$, without loss of generality we assume that there is a sequence $z^n \in Z$ such that

$$\sum_{t=0}^{T} \mathbf{E}[u(w_t^0 + z_t^n)] \to +\infty. \tag{3.5.9}$$

Clearly $\|z^n\| \to +\infty$. Then taking a subsequence if necessary we can assume that $z^n/\|z^n\| \to z^* \in Z\backslash\{0\}$. By property (u3) $z_t^n \geq -w_t^0, t = 0, 1, \dots, T$. Thus, $z_t^* \geq 0$ implies that z^* is an arbitrage trading strategy. ●

3.5.3 Dual Characterization of No Arbitrage

We turn to the dual characterization of the no arbitrage and its implication for the price of financial assets. Define, for $x \in \mathcal{X}$,

$$f(x) = \sum_{t=0}^{T} \mathbf{E}[(-u)(x_t)], \tag{3.5.10}$$

we can rewrite the optimal trading problem (3.5.8) as

$$p = -\inf[f(x) + \iota_{w^0+C}(x)]. \tag{3.5.11}$$

Note that the (CQ) condition

$$0 \in \operatorname{int}[\operatorname{dom} \iota_{w^0+C} - \operatorname{dom} f] = \operatorname{int}[w^0 + C - \mathcal{X}^+] \tag{3.5.12}$$

holds. Thus, strong duality implies that

$$\begin{aligned} p &= -\max_{z \in \mathcal{X}}\{-\sigma_{w^0+C}(z) - f^*(-z)\} \qquad (3.5.13)\\ &= \min_{z \in \mathcal{X}} \left\{ \sum_{t=0}^{T} \mathbf{E}[(-u)^*(-z_t) + \langle w^0, z\rangle + \sigma_C(z)\right\}. \end{aligned}$$

Let x^*, z^* be solutions to the primal and dual problem (3.5.11) and (3.5.13), respectively. Condition (u2) implies that $\operatorname{dom}(-u)^* = (-\infty, 0)$ so that $z_t^* > 0$. Moreover,

$$z_t^* \in -\partial(-u)(x_t^*). \tag{3.5.14}$$

Finally, if the market has no arbitrage trading strategy then $p < +\infty$ in (3.5.13) which implies that $\sigma_C(z^*) < \infty$ or

$$z^* \in C^\circ := \{z^* \in \mathcal{X} : \langle z^*, c\rangle \le 0, \text{ for all } c \in C\}. \tag{3.5.15}$$

Relation (3.5.15) can be interpreted as scaling z^* one can derive a martingale measure. Let's look into the detials. We will use $\mathbf{E}_t$ to denote the conditional expectation with respect to $\mathcal{F}_t$ and *characteristic function* of a set $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise. Since for any $A \in \mathcal{F}_t$, $\chi_A 1^{ut}, \chi_A \tilde{1}^{ut} \in Z \subset C$ we have

$$\langle z^*, \chi_A 1^{ut}\rangle \le 0 \text{ and } \langle z^*, \chi_A \tilde{1}^{ut}\rangle \le 0 \tag{3.5.16}$$

which implies that

$$g_t^u z_t^* \le \mathbf{E}_t[z_u^*] \le h_t^u z_t^*. \tag{3.5.17}$$

In particular, $\pi_t = \mathbf{E}_t[z_{t+1}^*]/z_t^* \in [g_t^{t+1}, h_t^{t+1}]$ plays the role of a discounting factor in the interval between transaction times t and $t+1$. Defining a discounting process Γ_t recursively by

$$\Gamma_0 = 1, \Gamma_{t+1} = \Gamma_t \pi_t, \tag{3.5.18}$$

we see that Γ_t is $\mathbb{F}-$ adapted. Denoting $M_t = z_t^*/\Gamma_t$ we can verify

$$\mathbf{E}_t[M_{t+1}] = \mathbf{E}_t\left[\frac{z_{t+1}^*}{\Gamma_{t+1}}\right] = \mathbf{E}_t\left[\frac{z_{t+1}^*}{\Gamma_t \pi_t}\right] = \frac{1}{\Gamma_t}\mathbf{E}_t\left[\frac{z_{t+1}^*}{\pi_t}\right] = \frac{z_t^*}{\Gamma_t} = M_t. \tag{3.5.19}$$

In other words, M_t is a matingale and we have the decomposition

$$z_t^* = \Gamma_t M_t. \tag{3.5.20}$$

Defining $Q = M_T P$, we can rewrite (3.5.17) as

$$g_t^u \le \mathbf{E}_t^Q\left[\sum_{s=t+1}^{T} \frac{\Gamma_s}{\Gamma_t} 1_s^{ut}\right] \le h_t^u. \tag{3.5.21}$$

Similarly, for any $A \in \mathcal{F}_t$, $\chi_A S^{it}, \chi_A \tilde{S}^{it} \in Z \subset C$ implies

$$b_t^i z_t^* \le \mathbf{E}_t\left[\sum_{s=t+1}^{T} z_s^* S_s^i\right] \le a_t^i z_t^*. \tag{3.5.22}$$

Deviding by z_t^* and using the representation (3.5.20) we have

$$b_t^i \le \mathbf{E}_t\left[\sum_{s=t+1}^{T}\frac{\Gamma_s M_s}{\Gamma_t M_t}S_s^i\right] \le a_t^i, \tag{3.5.23}$$

or

$$b_t^i \le \mathbf{E}_t^Q\left[\sum_{s=t+1}^{T}\frac{\Gamma_s}{\Gamma_t}S_s^i\right] \le a_t^i. \tag{3.5.24}$$

in other words the discounted values of the cash flows related to both bonds and risky assets under the equivalent martingale measure Q fall between the bid and ask prices.

3.5.4 Pricing and Hedging

In a market model with bid-ask spread the pair of equivalent martingale measure and related discount factors (Q, Γ) plays the role of an equivalent martingale measure in a one price market model.

Definition 3.5.3 (Martingale Measure and Discount Factor) *Let Q be a measure equivalent to P and let $\Gamma = (\Gamma_t)_{t=0}^T$ be an $\mathbb{F}-$ adapted process with $\Gamma_0 = 1$. We say that (Q, Γ) is an equivalent martingale measure and discounting process pair corresponding to the T-period market in section 3.5.1 provided that they satisfy reloationships (3.5.21) and (3.5.24). We use $\mathcal{MD}$ to denote the collection of all such pairs.*

The set $\mathcal{MD}$ plays a role similar to the set of equivalent martingale measures in a one price economy and can be used to determine sub and sper-hedge bounds. For sake of brievity we only discuss the one period case to illustrate the idea. Readers can find a more technical discussion of the general multiperiod model in [60].

We will show that

$$u_0 = \max\left\{\mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}c_1\right] : (Q, \Gamma) \in \mathcal{MD}\right\}$$

defines a super hedging bound. A sub-hedging bound can be derived similarly. We represent u_0 as a linear programming problem

$$\begin{aligned} u_0 = \max \ & \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}c_1\right] \\ \textit{subject to} \ & \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}S_1^m\right] \le a_0^m, \ \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}(-S_1^m)\right] \le -b_0^m, m = 1, \ldots, M, \\ & \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}\right] \le h_0^1, \ \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}(-1)\right] \le -g_0^1. \end{aligned} \tag{3.5.25}$$

We formulate the dual problem using the Lagrange format. Let

$$(\Lambda, \gamma) = (\lambda_0^1, \dots, \lambda_0^M, \tilde{\lambda}_0^1, \dots, \tilde{\lambda}_0^M, \gamma_0^1, \tilde{\gamma}_0^1) \in \mathbb{R}_+^{2M+2} \tag{3.5.26}$$

be the Lagrange multipliers of linear programming problem (3.5.25). We conside the Lagrangian

$$\begin{aligned} L((Q,\Gamma),(\Lambda,\gamma)) = \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}c_1\right] + \gamma_0^1\left(h_0^1 - \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}\right]\right) + \tilde{\gamma}_0^1\left(\mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}\right] - g_0^1\right) \\ + \sum_{m=1}^M \lambda_0^m\left(a_0^m - \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}S_1^m\right]\right) + \sum_{m=1}^M \tilde{\lambda}_0^m\left(\mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}S_1^m\right] - b_0^m\right). \end{aligned} \tag{3.5.27}$$

We can see that

$$\inf_{(\Lambda,\gamma)\in\mathbb{R}_+^{2M+2}} L((Q,\Gamma),(\Lambda,\gamma)) = \begin{cases} \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}c_1\right] & (Q,\Gamma)\in\mathcal{MD} \\ -\infty & otherwise. \end{cases} \tag{3.5.28}$$

Thus, by the strong linear programming duality

$$\begin{aligned} u_0 &= \sup_{(Q,\Gamma)\in PM\times\mathbb{R}_+} \inf_{(\Lambda,\gamma)\in\mathbb{R}_+^{2M+2}} L((Q,\Gamma),(\Lambda,\gamma)) \\ &= \inf_{(\Lambda,\gamma)\in\mathbb{R}_+^{2M+2}} \sup_{(Q,\Gamma)\in PM\times\mathbb{R}_+} L((Q,\Gamma),(\Lambda,\gamma)), \end{aligned} \tag{3.5.29}$$

where PM signifies the set of all probability measures. For $(\Lambda, \gamma) \in \mathbb{R}_+^{2M+2}$ consider the zero cost portfolio of cash follow

$$(z_0(\Lambda,\gamma), z_1(\Lambda,\gamma)) = \gamma_0^1 1^{10} + \tilde{\gamma}_0^1 \tilde{1}^{10} + \sum_{m=1}^M (\lambda_0^m S^{m0} + \tilde{\lambda}_0^m \tilde{S}^{m0}). \tag{3.5.30}$$

We see that

$$L((Q,\Gamma),(\Lambda,\gamma)) = \mathbf{E}_0^Q\left[\frac{\Gamma_1}{\Gamma_0}(c_1 - z_1(\Lambda,\gamma))\right] - z_0(\Lambda,\gamma). \tag{3.5.31}$$

Note that $z_0(\Lambda,\gamma)$ and Γ_1/Γ_0 are constants. Moreover, for u_0 to be finite in (3.5.29) we need only to consider (Λ,γ) that makes $c_1 - z_1(\Lambda,\gamma) \le 0$. Let

$$\bar{\omega} \in \operatorname{argmax}\{c_1 - z_1(\Lambda,\gamma)\}.$$

In (3.5.29), taking Q to be the probability measure concentrated at $\bar{\omega}$ and letting $\Gamma_1/\Gamma_0 = g_0^1$ for all $\omega \in \Omega$, we derive

$$u_0 = \inf_{(\Lambda,\gamma)\in\mathbb{R}_+^{2M+2}} \left\{ g_0^1 \sup_{\omega\in\Omega}[c_1(\omega) - z_1(\Lambda,\gamma)(\omega)] - z_0(\Lambda,\gamma) \right\}. \tag{3.5.32}$$

We show that a solution $(\bar{\Lambda}, \bar{\gamma})$ to the minimization problem (3.5.32) provides a super-hedging strategy when the bid price $b_0(c)$ for payoff c_1 at $t = 0$ exceeds u_0. We observe that $(b_0(c), -c_1)$ is a zero cost cash flow. To get an arbitrage strategy we aquire the zero cost cash flows (3.5.30) with $(\Lambda,\gamma) = (\bar{\Lambda},\bar{\gamma})$, $(b_0(c), -c_1)$, and $-\sup_{\omega\in\Omega}[c_1(\omega) - z_1(\bar{\Lambda},\bar{\gamma})(\omega)]$ units of $\tilde{1}^{10}$. We can see that this portfolio's cash flow is, at $t = 0$,

$$z_0(\bar{\Lambda},\bar{\gamma}) + b_0(c) - g_0^1 \sup_{\omega\in\Omega}[c_1(\omega) - z_1(\bar{\Lambda},\bar{\gamma})(\omega)] \tag{3.5.33}$$
$$= z_0(\bar{\Lambda},\bar{\gamma}) + b_0(c) - (u_0 - z_0(\bar{\Lambda},\bar{\gamma})) = b_0(c) - u_0 > 0,$$

and at $t = 1$,

$$z_1(\bar{\Lambda},\bar{\gamma}) - c_1 + \sup_{\omega\in\Omega}[c_1(\omega) - z_1(\bar{\Lambda},\bar{\gamma})(\omega)] \geq 0. \tag{3.5.34}$$

Thus, our portfolio is indeed an arbitrage.

4

Continuous Financial Models

Summary. We turn to discuss continuous financial models. These models in general involve infinite dimensional spaces and are more complex. Our focus here is to use relatively simple models to illustrate the convex duality between the price of a contingent claim and the process of cash borrowed in delta hedging. This reveals the root of the convexity in contingent claims. Interestingly, when hedging with a contingent claim instead of the underlying, a similar duality in the sense of generalized Fenchel conjugate holds. Correspondingly, this generalized duality leads to the generalized convexity of the contingent claims with many interesting applications. Much of the material presented in this chapter appear here for the first time.

4.1 Continuous Stochastic Processes

4.1.1 Continuous Stochastic Processes

A continuous stochastic process is a generalization of the discrete stochastic process that we discussed before.

Definition 4.1.1 (Stochastic Process) *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and let* $[0, T]$ *be an interval. We call* $(X_t), t \in [0, T]$ *a stochastic process if for every* t, X_t *is a random variable on* $(\Omega, \mathcal{F}, P)$.

In financial applications the parameter t is usually time but not always. For example, it could be the so call local time when the calendar time is fixed at a point and the parameter t, in fact, reflects the change in the price space. Similar to the discrete case we also need to deal with gradually revealing information.

Definition 4.1.2 (Filtration) *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $[0, T]$ be an interval. We say $(\mathcal{F}_t), t \in [0, T]$ is a* filtration *if for every t, $\mathcal{F}_t \subset \mathcal{F}$ is a σ-algebra and, for any $s < t$,*

$$\mathcal{F}_s \subset \mathcal{F}_t.$$

As in the discrete case, $\mathcal{F}_t$ represents information available up to time t. The definition implicitly assumes that information once become available will never be forgotten.

Definition 4.1.3 (Adapted Stochastic Process) *Let $(\mathcal{F}_t), t \in [0, T]$ be a filtration on probability space $(\Omega, \mathcal{F}, P)$. We say a stochastic process (X_t) is $\mathcal{F}_t$-adapted provided that, for every t, X_t is $\mathcal{F}_t$ measurable.*

Intuitively, the value X_s of an adapted stochastic process becomes deterministic when the current time $t > s$.

4.1.2 Brownian Motion and Martingale

Brownian motion is a special continuous stochastic process that plays a crucial role in financial modeling. It is named after the Scottish botanist Robert Brown who in 1828 observed such a motion from pollen suspended in liquid. Louis Bachelier first used it to model the price of financial assets in his 1900 Ph. D. thesis and derived the famous Bachelier formula for option pricing. The mathematical property of Brownian motion was clearly elaborated by Robert Weiner who also provided a proof of the existence of a Brownian motion by construction. Paul Samuelson proposed the widely used geometric Brownian motion model for stock price movements in 1965, which is more realistic when modeling assets with nonnegative values. However, the geometric Brownian motion is continuous so that it does not allow any price jump which does happen to a stock price process from time to time. As the saying goes “All models are wrong. Some are wronger than others.” What we need to keep in mind is that models are approximations of the reality. They are *not* reality.

Definition 4.1.4 (One-dimensional Brownian Motion) *A stochastic process $\{B_t : t \in [0, T)\}$ is called a standard Brownian motion if*

1. $B_0 = 0$,

2. *for* $0 \le t_1 < t_2 < \ldots < t_k \le T$*, the random variables*

$$B_{t_2} - B_{t_1}, B_{t_3} - B_{t_2}, \ldots, B_{t_k} - B_{t_{k-1}}$$

are independent,

3. *for* $0 \le s \le t \le T$*,* $B_t - B_s$ *has a Gaussian distribution with mean* 0 *and variance* $t - s$*,*
4. *for* ω *in a set of probability one, the path* $B_t(\omega)$ *is continuous.*

Definition 4.1.5 (Multi-dimensional Brownian Motion) *A vector stochastic process* $\{B_t : t \in [0,T]\}$ *in* $\mathbb{R}^n$ *is called a standard Brownian motion if* $B_t = (B_t^1, B_t^2, \ldots, B_t^n)$ *where* $B_t^i, i = 1, 2, \ldots, n$ *are independent standard one-dimensional Brownian motions. If* B_t *is a standard Brownian motion. Then* $x + B_t$ *is called a Brownian motion starting from* x*.*

Remark 4.1.6 The existence of a stochastic process satisfying all the conditions laid out in Definition 4.1.4 is not automatically guaranteed. By and large, there are two ways to prove the existence:

- by construction pioneered by Wiener (see e.g. [55]), or
- by Kolmogorov's extension theorem (see e.g. [42]).

We are satisfied with known the existence of Brownian motions for our applications.

If in a given probability space there is a Brownian motion then one can also define a Brownian motion in a different yet similar probability space. Thus, Brownian motion is not uniquely defined. However, since every Brownian motion has the same properties laid out in Definition 4.1.4, their effects are equivalent. We usually pick a 'convenient' version for the purpose of a concrete application.

For each Brownian motion B_t, defining the $\sigma-$algebra represents the information contained in B_t up to time t by $\mathcal{F}_t$ we get a nature filtration associated with B_t. In fact, we can take $\mathcal{F}_t$ to be the $\sigma-$algebra generated by the collection of preimages of Borel sets under $B_s, s < t$. In the sequel whenever we discuss a Brownian motion we always assume that it is accompanied by this filtration.

Somewhat more general than a Brownian motion is the martingale process.

Definition 4.1.7 (Martingale) *Let* $\mathcal{F}_t$ *be a filtration for the probability space* $(\Omega, \mathcal{F}, P)$*. We say* M_t *is a* $(P, \mathcal{F}_t)-$martingale *if* M_t *is adapted to the filtrition* $\mathcal{F}_t$*, for all* $t > 0$*,* $\mathbf{E}|M_t| < \infty$ *and for all* $s < t$*,*

$$\mathbf{E}^P[M_t|\mathcal{F}_s] = M_s.$$

Similar to the discrete case a martingale can be think of representing the wealth process in playing a fair game. A Brownian motion B_t is clearly a martingale and it is also easy to check that $M_t = B_t^2 - t$ is also a martingale. So martingale is not necessarily a Brownian motion. However, martingales are only slightly more general than the Brownian motion as the following Levy's theorem shows (which we state without proof).

Theorem 4.1.8 (The Levy Characterization of Brownian Motion) *Let $X(t) = (X_1(t), \ldots, X_n(t))$ be a continuous stochastic process on $(\Omega, \mathcal{F}, Q)$. Then $X(t)$ is a Brownian motion with respect to Q if and only if*
(i) $X(t)$ is a martingale w.r.t. Q, and
(ii) $X_i(t)X_j(t) - \delta_{ij}t$ is a martingale w.r.t. Q for all $i, j = 1, \ldots, n$.
Here δ_{ij} is the Kronecker delta defined by $\delta_{ij} = 0$ when $i \neq j$ and $\delta_{ii} = 1$.

For $n = 1$ we have the characterization of one-dimensional Brownian motion.

Theorem 4.1.9 (The Levy Characterization of Brownian Motion) *Let $X(t)$ be a scalar continuous stochastic process on $(\Omega, \mathcal{F}, Q)$. Then $X(t)$ is a Brownian motion with respect to Q if and only if*
(i) $X(t)$ is a martingale w.r.t. Q, and
(ii) $X^2(t) - t$ is a martingale w.r.t. Q.

4.1.3 The Itô Formula

The Itô formula is an important tool in analyzing continuous stochastic processes.

Theorem 4.1.10 (Basic Form of the Itô formula)*Let $f(x,t) \in C^{2,1}$ and let B_t be a one dimensional Brownian motion. Then*

$$df(B_t,t) = f_t(B_t,t)dt + f_x(B_t,t)dB_t + \frac{1}{2}f_{xx}(B_t,t)dt. \tag{4.1.1}$$

The Itô formula presented in (4.1.1) is a shorthand for

$$f(B_t,t) = f(0,0) + \int_0^t f_t(B_s,s)ds \tag{4.1.2}$$
$$+\int_0^t f_x(B_s,s)dB_s + \frac{1}{2}\int_0^t f_{xx}(B_s,s)dt.$$

This formula (4.1.1) looks like an usual chain rule except for the last term. A rigorous proof is beyond the scope of this short book. Below are some heuristics that can help in understanding the Itô formula.

We know that $f(B_t,t) - f(0,0) = \int_0^t df(B_t,t)$. Expand $df(B_t,t)$ using the Taylor's expansion. Since terms of order $o(dt)$ will vanish in the integration process we need only do this to the second order. That gives us

$$df(B_t,t) = f_t(B_t,t)dt + f_x(B_t,t)dB_t + \frac{1}{2}f_{xx}(B_t,t)dB_t^2$$
$$+\frac{1}{2}f_{tt}(B_t,t)dt^2 + f_{tx}(B_t,t)dtdB_t.$$

Since $dt^2, dtdB_t$ are $o(dt)$ the last two terms can be omitted and we have

$$df(B_t,t) = f_t(B_t,t)dt + f_x(B_t,t)dB_t + \frac{1}{2}f_{xx}(B_t,t)dB_t^2.$$

By the properties of the Brownian motion, we can replace dB_t^2 by dt giving us the Itô formula (4.1.1).

Graphically we can illustrate by drawing the graph of f_x around point B_t, then $df(B_t,t)$ is the area under the graph of f_x (see Fig 4.1). We can see that $f_x(B_t,t)dB_t$ represnts the approximation of the area using Euler's mathod while $\frac{1}{2}f_{xx}(B_t,t)dB_t^2 \sim \frac{1}{2}f_{xx}dt$ corrects the "triangle" part to get to an approximation using the trapezoid rule.

The heuristic argument leads us to the following simple rule in handling the differential term arising in the Taylor expansion of a function of the Itô process usually called *box algebra*.

	dt	dB_t
dt	0	0
dB_t	0	dt

Example 4.1.11 Below is a nice application illustrating the power of the Itô formula. Define $\beta_k(t) = \mathbf{E}B_t^k$. Itô formula gives us

$$\beta_k(t) = \frac{1}{2}k(k-1)\int_0^t \beta_{k-2}(s)ds.$$

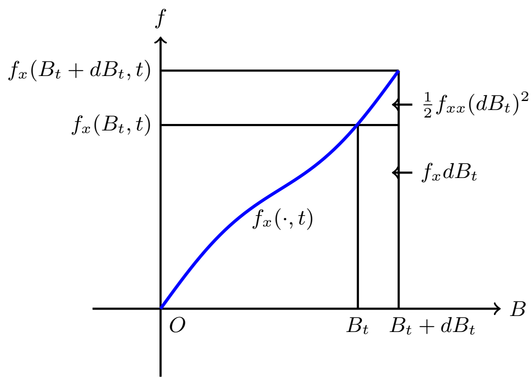


Fig. 4.1. Graphic illustration of the Itô formula

We can use this to easily get $\mathbf{E}B_t^3 = 0$ and $\mathbf{E}B_t^4 = 3t^2$. Those are mostly used in financial applications. By induction, in general $\mathbf{E}B_t^{2k+1} = 0$ and

$$\mathbf{E}B_t^{2k} = \frac{(2k)!t^k}{2^k k!}.$$

Itô Processes

Let B_t be a one-dimensional Brownian motion with respect to filtration $\mathcal{F}_t$ on $(\Omega, \mathcal{F}, P)$. Then

$$X_t = X_0 + \int_0^t \mu(s, \omega) ds + \int_0^t \sigma(s, \omega) dB_s$$

is called a (1-dim) Itô processes if μ, σ are $\mathcal{F}_t$ adapted,

$$P\left[\int_0^t \sigma(s, \omega)^2 ds < \infty \text{ for all } t \geq 0\right] = 1$$

and

$$P\left[\int_0^t |\mu(s, \omega)| ds < \infty \text{ for all } t \geq 0\right] = 1.$$

In shorthand we write

$$dX_t = \mu dt + \sigma dB_t.$$

Here μ is a drift and σ indicates magnitude of the variation of the random part. It is often useful to write stochastic process in this form if we can. A Brownian motion is an example of an Itô process where $\mu = 0$ and $\sigma = 1$. The Itô formula can be generalized to Itô process with dX_t replacing dB_t.

Theorem 4.1.12 (The General Itô formula) *Let $f(t,x) \in C^2$ and let X_t be an Itô process. Then*

$$df(X_t,t) = f_t(X_t,t)dt + f_x(X_t,t)dX_t + \frac{1}{2}f_{xx}(X_t,t)(dX_t)^2.$$

Example 4.1.13 Applying the Itô formula to $f(x) = x^2$ we have

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t).$$

Example 4.1.14 (Integration by Parts) The pattern in handling $f(x) = x^2$ holds in more general setting. Let $g(s)$ be a continuous function with bounded variation with respect to $s \in [0,t]$. Applying the Itô formula to $f(t,x) = g(t)x$ we have

$$\int_0^t g(s)dB_s = g(t)B_t - \int_0^t g'(s)B_s ds.$$

Example 4.1.15 Here is an example of using the general Itô formula. Let $X_t = \mu t + \sigma B_t$. Then $dX_t = \mu dt + \sigma dB_t$. Using the box algebra we have

$$\begin{aligned} df(X_t,t) &= f_t dt + f_x dX_t + \frac{1}{2}f_{xx}dX_t^2 \\ &= f_t dt + \mu f_x dt + \sigma f_x dB_t + \frac{1}{2}\sigma^2 f_{xx} dt \end{aligned}$$

Example 4.1.16 Letting $f(t,x) = tx$ we have

$$tB_t = \int_0^t B_s ds + \int_0^t s dB_s$$

or

$$\int_0^t s dB_s = tB_t - \int_0^t B_s ds.$$

The Multidimensional Itô Formula

Let $X_t = (X_t^1, \ldots, X_t^n)$ be an n-dimensional Itô process satisfying

$$dX_t = \mu dt + \sigma dB_t,$$

where μ is an n-dimensional vector, σ an $n \times m$ matrix and B_t an n-dimensional Brownian motion. We require the components of μ and σ satisfy similar conditions in the definition of the one-dimensional Itô process. Let $g(t, x) : [0, \infty) \times R^n \to R^p$ has continuous second order partial derivatives. Then, for $Y_t = g(t, X_t)$,

$$dY_t^k = \frac{\partial g_k}{dt} dt + \sum_{i=1}^{n} \frac{\partial g_k}{\partial x_i} dX_t^i + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g_k}{\partial x_i \partial x_j} dX_t^i dX_t^j. \tag{4.1.3}$$

The following multi-dimensional box algebra is a convenient tool in simplifying the multi-dimensional Itô formula

	dt	dB_t^1	dB_t^2	$\ldots$	dB_t^n
dt	0	0	0	$\ldots$	0
dB_t^1	0	dt	0	$\ldots$	0
dB_t^2	0	0	dt	$\ldots$	0
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
dB_t^n	0	0	0	$\ldots$	dt

Example 4.1.17 (Integration by Parts) Let X_t, Y_t be Itô processes in R. Applying the Itô formula to $f(X_t, Y_t) = X_t Y_t$ we have

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + dX_t dY_t.$$

The integral form in the following is the general integration by parts formula

$$\int_0^t X_s dY_s = X_t Y_t - X_0 Y_0 - \int_0^t Y_s dX_s - \int_0^t dX_s dY_s.$$

Remark 4.1.18 The term $dX_t dY_t$ is called the quadratic covariation of X_t an Y_t and is often denoted $d\langle X, Y \rangle_t$.

Martingale Representation

The Itô formula is a crucial tool in proving the following important martingale representation theorem. This representation theorem further highlights the close relationship between martingales and Brownian motions. As an application oriented class we will omit the proof and directly present the result.

Theorem 4.1.19 (Martingale Representation) *Let B_t be an n-dimensional Brownian motion generating filtration $\mathcal{F}_t^n$. Suppose that M_t is an $(P, \mathcal{F}_t^n)-$ martingale and that $M_t \in L^2(P)$ for all $t \geq 0$. Then there exists a unique stochastic process $v \in \mathcal{V}^n$ such that*

$$M_t = \mathbf{E}M_0 + \int_0^t vdB_s.$$

Dual Itô Formula

Let $f(x,t) \in C^{3,1}$ and let X_t be an Itô process. Then using the quadratic covariation in Remark 4.1.18 we can write the general Itô formula in Theorem 4.1.12 as

$$df(X_t, t) = f_t(X_t, t)dt + f_x(X_t, t)dX_t + \frac{1}{2}d\langle f_x(X,t), X\rangle_t. \tag{4.1.4}$$

Now assume that f is convex in x for all t. We use $f^*(y,t)$ to signify the conjugate of f with respect to variable x. Define $Y_t = f_x(X_t, t)$. We see that X_t, Y_t satisfies the Fenchel equality

$$f(X_t, t) + f^*(Y_t, t) = X_tY_t. \tag{4.1.5}$$

It follows that

$$f_t(X_t, t) + f_t^*(Y_t, t) = 0, \tag{4.1.6}$$

$$Y_t = f_x(X_t, t) \text{ and } X_t = f_y^*(Y_t, t), \tag{4.1.7}$$

and using Example 4.1.17

$$df(X_t, t) + df^*(Y_t, t) = X_tdY_t + Y_tdX_t + d\langle X, Y\rangle_t. \tag{4.1.8}$$

Combining (4.1.4), (4.1.6) and (4.1.8) we derive the following Dual Itô formula

$$\begin{aligned} df(X_t, t) &= f_t(X_t, t)dt + Y_tdX_t + \frac{1}{2}d\langle Y, X\rangle_t \qquad (4.1.9)\\ df^*(Y_t, t) &= f_t^*(Y_t, t)dt + X_tdY_t + \frac{1}{2}d\langle X, Y\rangle_t. \end{aligned}$$

4.1.4 Girsanov Theorem

In financial applications, prices of stocks and other assets are often described by a Itô process of the form

$$dS_t = \mu dt + \sigma dB_t$$

where μ models a drift reflecting the large trend of the asset price and σ describes the volatility of the random fluctuation of the price process. In analyzing the price process, the important part is the impact of σ. The Girsanov theorem allows us to 'absorb' the drift μ by using a change of the probability measure. This is very similar to the equivalent martingale measure that absorbs the excess gains for the risky assets in the discrete model.

Theorem 4.1.20 (Girsanov Theorem) *Let S_t be an Itô process of the form*

$$dS_t = \mu(t,\omega)dt + \sigma(t,\omega)dB_t, t \in [0,T], S_0 = 0,$$

where B_t is a $(P, \mathcal{F}_t)-$Brownian motion and μ, σ are bounded and $\sigma > c > 0$ for some constant c. Then

1. *for $u = \mu/\sigma$, $M_t = \exp\left(-\int_0^t u(s,\omega)dB_s - \frac{1}{2}\int_0^t u^2(s,\omega)ds\right), t \in [0,T]$, is a $(P, \mathcal{F}_t)-$martingale.*
2. $$dQ(\omega) = M_T(\omega)dP(\omega).$$
   is a probability measure on $\mathcal{F}_T$ and
3. *$\hat{B}(t) = \int_0^t u(s,\omega)ds + B(t)$ is a Brownian motion w.r.t. Q and*
4. $$dS_t = \sigma(t,\omega)d\hat{B}(t).$$

Proof. (Sketch) Let $X_t = \int_0^t u(s,\omega)dB_s + \frac{1}{2}\int_0^t u^2(s,\omega)ds$ we have

$$dX_t = udB_t + \frac{1}{2}u^2 dt.$$

By direct calculation we have

$$dM_t = -u\exp(-X_t)dB_t$$

and, therefore, M_t is a martingale by the martingale representation theorem.

Since M_t is a martingale and $M_0 = 1$,

$$Q(\Omega) = \mathbf{E}^Q[1] = \mathbf{E}^P[M_T] = 1.$$

Thus, Q is a probability measure on $\mathcal{F}_T$. We note that $dQ = M_t dP$ on $\mathcal{F}_t$. In fact, for any bounded $\mathcal{F}_t-$measurable function f,

$$\int_\Omega f dQ = \int_\Omega f M_T dP = \mathbf{E}[fM_T] = \mathbf{E}[\mathbf{E}[fM_T|\mathcal{F}_t]]$$
$$= \mathbf{E}[f\mathbf{E}M_T|\mathcal{F}_t] = \mathbf{E}[fM_t] = \int_\Omega f M_t dP.$$

To show $\hat{B}_t$ is a Brownian motion, we turn to check the conditions in the Levy characterization of Theorem 4.1.8. We do only (i) and (ii) is similar. Using the product rule we can verify that $M_t\hat{B}_t$ is a martingale with respect to P. Now for $s < t$, and $A \in \mathcal{F}_s$ we have

$$\int_A \mathbf{E}^Q[\hat{B}_t|\mathcal{F}_s]dQ$$
$$= \int_A \hat{B}_t dQ = \int_A \hat{B}_t M_t dP = \mathbf{E}^P[1_A M_t \hat{B}_t]$$
$$= \mathbf{E}^P[\mathbf{E}^P[1_A M_t \hat{B}_t|\mathcal{F}_s]] = \mathbf{E}^P[1_A M_s \hat{B}_s]$$
$$= \int_A \hat{B}_s M_s dP = \int_A \hat{B}_s dQ.$$

Since $A \in \mathcal{F}_s$ is arbitrary, $E^Q[\hat{B}_t|\mathcal{F}_s] = \hat{B}_s$. ●

Measure Q is called the martingale measure for process S_t.

4.2 Bachelier and Black-Scholes Formulae

4.2.1 Pricing Contingent Claims

Let S_t be an Itô process

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dB_t$$

that represents the price process of a certain financial asset. Here B_t is a Brownian motion in a probability measure space $(\Omega, \mathcal{F}, P)$ with filtration $\mathcal{F}_t$. Assume for simplicity that the risk free interest rate is 0 and that μ, σ are bounded and $\sigma \geq c > 0$ for some constant c. Suppose that we want to price a European style contingent claim on S_t with the payoff $f(S_T)$ at the maturity T. We can proceed as follows. First using the Girsanov theorem we can write

$$dS_t = \sigma(S_t, t)dW_t$$

where W_t is a Brownian motion in $(\Omega, \mathcal{F}, Q)$ with filtration $\mathcal{F}_t$ where Q is a martingale measure for S_t equivalent to P. Similar to the discrete version of the fundamental theorem of asset pricing, we can write down the no arbitrage price function for the contingent claim at any time $t \in [0, T]$ and price x as

$$v(x, t) = \mathbf{E}^Q[f(S_T)|S_t = x]. \tag{4.2.1}$$

Next we explicitly calculate the price function for call options under the Bachelier and Black-Scholes models.

Bachelier Formula

Bachelier modeled the price of a stock in his 1900 pioneering paper [3] by

$$dS_t = \mu dt + \sigma dB_t$$

where μ and σ are constant. This model was thought unrealistic because stock price cannot become negative. However, now we can see it as a good approximation for pair trading or forward for currency swap contracts. Consider the price of a call option with a strike K maturing at T. Then formula (4.2.1) reduces to

$$B(x, t) = \mathbf{E}^Q[(S_T - K)^+|S_t = x], \tag{4.2.2}$$

where Q is an equivalent martingale measure with respect to the price process S_t. Since under Q the dynamics of the price process is

$$dS_t = \sigma dW_t$$

where W_t is a Q Brownian motion, we have

$$S_T = x + \sqrt{T - t}\sigma W_1,$$

where $W_1 \sim N(0,1)$. Thus,

$$\begin{aligned} B(x,t) &= \mathbf{E}^Q[(x+\sqrt{T-t}\sigma W_1 - K)^+] \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(x-K+\sqrt{T-t}\sigma y)^+ e^{-\frac{y^2}{2}}\,dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{\frac{K-x}{\sigma\sqrt{T-t}}}^{\infty}(x-K+\sqrt{T-t}\sigma y)e^{-\frac{y^2}{2}}\,dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-K}{\sigma\sqrt{T-t}}}(x-K-\sqrt{T-t}\sigma z)e^{-\frac{z^2}{2}}\,dz \ (z=-y) \end{aligned} \tag{4.2.3}$$

We can write (4.2.3) concisely as

$$B(x,t) = (x-K)N\left(\frac{x-K}{\sigma\sqrt{T-t}}\right) + \sigma\sqrt{T-t}N'\left(\frac{x-K}{\sigma\sqrt{T-t}}\right), \tag{4.2.4}$$

where

$$N(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{z^2}{2}}\,dz.$$

Black-Scholes Formula

Black and Scholes modeled the price of a stock as a geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

where μ and σ are constant. Consider the price of a call option with a strike K maturing at T. Again formula (4.2.1) reduces to

$$C(x,t) = \mathbf{E}^Q[(S_T - K)^+ | S_t = x], \tag{4.2.5}$$

where Q is an equivalent martingale measure with respect to the price process S_t. Now under Q the dynamics of the price process is

$$dS_t = \sigma S_t dW_t$$

where W_t is a Q Brownian motion. We have

$$S_T = x\exp\left(-\frac{-\sigma^2(T-t)}{2} + \sqrt{T-t}\sigma W_1\right), \tag{4.2.6}$$

where $W_1 \sim N(0,1)$. Thus,

$$\begin{aligned} C(x,t) &= \mathbf{E}^Q\left[\left(x\exp\left(-\frac{-\sigma^2(T-t)}{2} + \sqrt{T-t}\sigma W_1\right) - K\right)^+\right] \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(x\exp\left(-\frac{-\sigma^2(T-t)}{2} + \sqrt{T-t}\sigma y\right) - K\right)^+ e^{-\frac{y^2}{2}}\,dy \\ &= \frac{1}{\sqrt{2\pi}}\int_{\frac{\ln\left(\frac{K}{x}\right)+\frac{\sigma^2(T-t)}{2}}{\sigma\sqrt{T-t}}}^{\infty}\left(x\exp\left(-\frac{-\sigma^2(T-t)}{2} + \sqrt{T-t}\sigma y\right) - K\right) e^{-\frac{y^2}{2}}\,dy \end{aligned} \tag{4.2.7}$$

which can be represented as

$$C(x,t) = xN(d_+) - KN(d_-), \tag{4.2.8}$$

where

$$d_\pm = \frac{\ln\left(\frac{x}{K}\right) \pm \frac{\sigma^2(T-t)}{2}}{\sigma\sqrt{T-t}}.$$

4.2.2 Convexity

Convexity and generalized convexity play important roles in dealing with option pricing and hedging. Both Bachelier and Black-Scholes formulae involve interesting convexity with respect to their various parameters.

We start with the Bachelier formula and use $I = \sqrt{T-t}\sigma$ and forward price $X = x - K$ to simplify notation. We will also use their ratio moneyness $m = X/I$. Using these new variables the we can write the Bachelier formula (4.2.3) as

$$B(X, I) = \mathbf{E}^Q[(X + IW_1)^+] = XN\left(\frac{X}{I}\right) + IN'\left(\frac{X}{I}\right). \tag{4.2.9}$$

Since for any fixed w, $(X + Iw)^+$ is a sublinear function of (X, I), so is $B(X, I)$. Thus, we have representation

$$B(X, I) = XB_X + IB_I. \tag{4.2.10}$$

Comparing with (4.2.9) we see that

$$B_X = N\left(\frac{X}{I}\right) \text{ and } B_I = N'\left(\frac{X}{I}\right). \tag{4.2.11}$$

We see that the sublinear property of the Bachelier formula brings us much convenience in calculating B_X and B_I.

The sublinearity of B also mens that its conjugate is an indicator function of some convex set M and we have the representation

$$B = \sigma_M \text{ and } B^* = \iota_M.$$

By the definition of conjugate function we can calculate that

$$\begin{aligned} M &= \{(X^*, I^*) : I^* + mX^* \le mN(m) + N'(m)\} \\ &= \{(N(m), I^*) : I^* \le N'(m), m \in \mathbb{R}\}. \end{aligned} \tag{4.2.12}$$

We now turn to the Black-Scholes formula. First direct calculation verifies

$$\frac{\partial C(x,t)}{\partial x} = N(d_+). \tag{4.2.13}$$

We observe that the variable x appears in the expressions of $C(x, t)$ in three separate places. Yet curiously the calculation result of the partial derivative with respect to x contains only the partial derivative with respect to the linear term of x. This is rather similar to the simple formula for B_X in (4.2.11). In the next section we will show the reason is related to the convexity of C in x and Fenchel-Legendra transform of C in x is related to the delta hedging. It is nature to ask whether C is also convex with respect to σ. It turns out the answer is negative. Yet if we compensate C by a multiple of an at money call it becomes convex.

We start by calculating the partial derivative of C with respect to σ:

$$C_\sigma = xN'(d_+)\frac{\partial d_+}{\partial \sigma} - KN'(d_-)\frac{\partial d_-}{\partial \sigma}. \tag{4.2.14}$$

Observing that

$$xN'(d_+) = KN'(d_-) = \sqrt{\frac{xK}{2\pi}} \exp\left(-\frac{(\ln(x/K))^2}{2\tau\sigma^2} - \frac{\tau\sigma^2}{8}\right) \tag{4.2.15}$$

and

$$d_+ - d_- = \sigma\sqrt{\tau} \tag{4.2.16}$$

we can simplify the expression of C_σ to

$$C_\sigma = \sqrt{\frac{xK\tau}{2\pi}} \exp\left(-\frac{(\ln(x/K))^2}{2\tau\sigma^2} - \frac{\tau\sigma^2}{8}\right). \tag{4.2.17}$$

It follows that

$$C_{\sigma\sigma} = \sqrt{\frac{xK\tau}{2\pi}} \exp\left(-\frac{(\ln(x/K))^2}{2\tau\sigma^2} - \frac{\tau\sigma^2}{8}\right)\left(\frac{(\ln(x/K))^2}{\tau\sigma^3} - \frac{\tau\sigma}{4}\right). \tag{4.2.18}$$

Defining

$$f(\sigma) := C - \sqrt{xK}\left[N\left(\frac{\sqrt{\tau}\sigma}{2}\right) - N\left(-\frac{\sqrt{\tau}\sigma}{2}\right)\right]$$

(note inside the hard bracket is the percentage premium of an at the money call option) we have

$$\begin{aligned} f''(\sigma) &= C_{\sigma\sigma} + \sqrt{xK\tau}\frac{\tau\sigma}{4}N'\left(\frac{\sqrt{\tau}\sigma}{2}\right) \\ &= \sqrt{xK\tau}N'\left(\frac{\sqrt{\tau}\sigma}{2}\right)\exp\left(-\frac{(\ln(x/K))^2}{2\tau\sigma^2}\right)\frac{(\ln(x/K))^2}{\tau\sigma^3} \\ &\quad +\sqrt{xK\tau}\frac{\tau\sigma}{4}N'\left(\frac{\sqrt{\tau}\sigma}{2}\right)\left(1 - \exp\left(-\frac{(\ln(x/K))^2}{2\tau\sigma^2}\right)\right) \geq 0. \end{aligned} \tag{4.2.19}$$

We note that

$$N\left(\frac{\sqrt{\tau}\sigma}{2}\right) - N\left(-\frac{\sqrt{\tau}\sigma}{2}\right)$$

is the price of an at the money call. Thus, the Black-Scholes call price C compensated by a multiple $(-\sqrt{x/K})$ of an at the money call as a function of σ is convex. We can also phrase this in terms of generalized convexity. Note that f is convex and, therefore, can be supported from below by an affine function. Thus, the Black-Scholes call price C as a function of σ can be supported from below by a function of the form

$$\sqrt{xK}\left[N\left(\frac{\sqrt{\tau}\sigma}{2}\right) - N\left(-\frac{\sqrt{\tau}\sigma}{2}\right)\right] + y\sigma - b.$$

Define

$$c(\sigma, y) = \sqrt{xK}\left[N\left(\frac{\sqrt{\tau}\sigma}{2}\right) - N\left(-\frac{\sqrt{\tau}\sigma}{2}\right)\right] + y\sigma$$

Then the Black-Scholes call price C as a function of σ is $\Phi_{c(1)}-$convex using the notation in Section 1.5.

4.2.3 Duality

We turn to explore the reason why the derivative of the Black-Scholes call formula C has a simple derivative with respect to x. To understand this phenomenon we need to go back to the original derivation of the Black-Scholes formula in [5]. Black and Scholes derive formula (4.2.8) by considering a portfolio of N_t shares of the underlying to hedge a short position of one share of the European call option:

$$S_t N_t - C(S_t, t). \tag{4.2.20}$$

They want to choose N_t in such a way that the resulting portfolio (4.2.20) has riskless gains, that is

$$N_t dS_t - dC(S_t, t) = 0. \tag{4.2.21}$$

Using the Itô formula we have

$$N_t dS_t = \frac{\partial C}{\partial x} dS_t + \left(\frac{\partial C}{\partial t} + \frac{1}{2} \frac{\partial^2 C}{\partial x^2} \right) dt. \tag{4.2.22}$$

It follows that

$$N_t = \frac{\partial C}{\partial x} \tag{4.2.23}$$

and C must satisfies the Black-Scholes partial differential equation

$$\frac{\partial C}{\partial t} + \frac{1}{2} \frac{\partial^2 C}{\partial x^2} = 0, \tag{4.2.24}$$

with terminal condition

$$C(x, T) = (x - K)^+. \tag{4.2.25}$$

The Black-Scholes partial differential equation (4.2.24) with the terminal condition (4.2.25) provide an alternative derivation of the Black-Scholes formula (4.2.8) via the Feynmann-Kac formula.

Relationships (4.2.20) and (4.2.23) reveals that when portfolio (4.2.20) has riskless gains its value equals to the Fenchel - Legendra transform of the no arbitrage option price. Since Merton has shown that the Black-Scholes option price $C(S_t, t)$ is convex in S_t, we have the following duality:

$$C^*(N_t, t) = \sup_{S_t} [N_t S_t - C(S_t, t)], \tag{4.2.26}$$

and

$$C(S_t, t) = \sup_{S_t} [N_t S_t - C^*(S_t, t)], \tag{4.2.27}$$

where the conjugate operation is with respect to the first variable. These relationships reveal that for each fixed t the option value is a convex function of the stock price and the cash borrowed $C^*(N_t, t)$ is a convex function of the share of the of the stock in the hedging portfolio. The same relationship also holds for the Bachelier formula. Thus, the simple form of the partial derivative of C in (4.2.13) is a consequence of the Fenchel-Young equaility in Proposition 1.3.1. This duality argument also explains the simplicity of B_X but as mentioned before B_X can be derived more directly using the sublinear property of the Bachelier formula B.

4.3 Duality and Delta Hedging

The duality relationship in delta hedging observed in the previous section for the Bachelier and Black-Scholes formulae also holds in more general setting.

4.3.1 Delta Hedging

We consider a diffusion process S_t satisfying

$$dS_t = \sigma S_t dW_t, \tag{4.3.1}$$

where W_t is a standard Brownian motion under measure Q (so that Q is a martingale measure for S_t). We assume that the risk free rate is 0. Consider a contingent claim on S_t of European style with maturity at $T > 0$ and a terminal payoff $f(S_T)$ at $t = T$. Denoting the price of the European contingent claim at time t by $v(S_t, t)$. We use a portfolio of N_t shares of the underlying S_t to hedge a short position of one share of the European call option:

$$S_t N_t - v(S_t, t). \tag{4.3.2}$$

The gain of this portfolio is

$$N_t dS_t - dv(S_t, t). \tag{4.3.3}$$

Applying the Itô formula we get we can rewrite (4.3.3) as

$$N_t dS_t - \left(v_t + \frac{\sigma^2 x^2}{2} v_{xx}\right) dt + v_x \sigma dW_t$$

To ensure a riskless gain we need

$$N_t = v_x(S_t, t). \tag{4.3.4}$$

Then the gain in portfolio reduces to

$$\left(v_t + \frac{\sigma^2 x^2}{2} v_{xx}\right) dt.$$

Now no arbitrage requires this quantity to be 0. Thus, v must satisfy the Black-Scholes PDE

$$v_t + \frac{\sigma^2 x^2}{2} v_{xx} = 0. \tag{4.3.5}$$

with terminal condition

$$v(x, T) = f(x), \tag{4.3.6}$$

where f is the payoff of the target at T.

4.3.2 Duality

Using (4.2.1) we know that

$$v(x,t) = \mathbf{E}^Q[f(S_T)|S_t = x] \tag{4.3.7}$$
$$= \mathbf{E}^Q\left[f(x\exp\left(-\frac{\sigma^2}{2}(T-t) + \sqrt{T-t}\sigma W_1\right)\right],$$

where $W_1 \sim N(0,1)$ under measure Q. Thus we see that v is convex in x provided that f is convex.

Fixing t, $v_x(\cdot,t)$ is a monotone increasing function. Thus, we can represent the pricing portfolio $S_tN_t - v(S_t,t)$ graphically in Fig. 4.2 and Fig. 4.3.

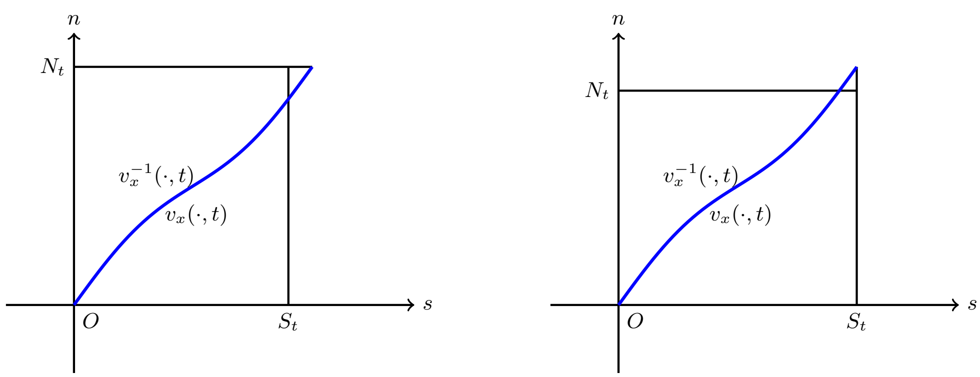


Fig. 4.2. Hedging portfolio

We see from those graphs the similarity with Fenchel duality. Indeed whenever the terminal payoff f of the European contingent claim is convex we have the following duality relationship:

$$v^*(N_t,t) = \sup_{S_t}[S_tN_t - v(S_t,t)] \tag{4.3.8}$$

and

$$v(S_t,t) = \sup_{N_t}[S_tN_t - v^*(N_t,t)]. \tag{4.3.9}$$

Here the conjugate v^* is the cash borrowed process when we maintaining a self-financing hedging portfolio. Relationship (4.3.8) corresponds to that the hedging portfolio has riskless gain and relationship (4.3.9) shows that the hedging portfolio $S_tN_t - v^*(N_t,t)$ is self-financing.

To implement this hedging, N_t must satisfy the Fenchel equality

$$v(S_t,t) + v^*(N_t,t) = S_tN_t. \tag{4.3.10}$$

Then $N_t = v_x(S_t,t)$ is a function of S_t and $S_t = v_n^*(N_t,t)$ is a function of N_t. Moreover,

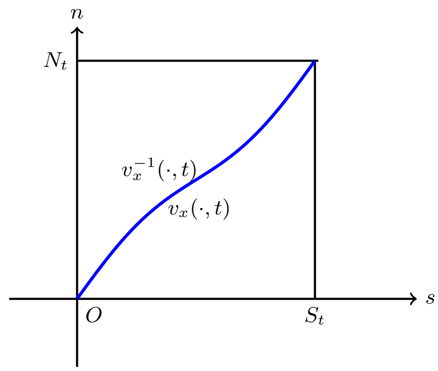


Fig. 4.3. Equality holds when $N_t = v_x(S_t,t), S_t = v_x^{-1}(N_t,t)$

$$\frac{\partial v}{\partial t} = -\frac{\partial v^*}{\partial t}$$

and

$$v_{xx}v^*_{nn} = 1.$$

Substituting the above into (4.3.5) we derive

$$-\frac{\partial v^*}{\partial t} + \frac{\sigma^2 x^2 \tilde{v}_{xx}^2}{2} v^*_{nn} = 0. \tag{4.3.11}$$

4.3.3 Time Reversal

In particular, if we reverse the time by setting $\tau = T - t$ then equation (4.3.11) becomes

$$\frac{\partial v^*}{\partial \tau} + \frac{\sigma^2 x^2 \tilde{v}_{xx}^2}{2} v^*_{nn} = 0. \tag{4.3.12}$$

Since equations (4.3.12) and (4.3.5) have the same form this suggests that in reverse time the cash borrowed process v^* should be a martingale just like v is a martingale in time t.

Let us fix the notation first. We use τ to denote the reversed time. For a stochastic process $P_t, t \in [0,T]$ we define its time reversal by $\hat{P}_\tau = P_t$ provided that $t + \tau = T$. Let us denote Δ an infinitesimal increment of time. Setting $\tau + t + \Delta = T$, we have

$$dP_t = P_{t+\Delta} - P_t = \hat{P}_\tau - \hat{P}_{\tau+\Delta} = -d\hat{P}_\tau.$$

We note that if W_t is a Brownian motion under measure Q then so is $\hat{W}_\tau$ under the same measure. The time reversal of a function of a stochastic process is defined below using $N_t = v_x(S_t,t)$ as an example

$$\hat{N}_\tau = v_x(\hat{S}_\tau, \tau).$$

The time reversal for the differential of a product stochastic processes needs to be dealt with caution. For example, we can write (4.3.1) as

$$S_{t+\Delta} - S_t = \sigma S_t (W_{t+\Delta} - W_t).$$

Letting $t + \tau + \Delta = T$ we have

$$\begin{aligned} d\hat{S}_\tau &= \hat{S}_{\tau+\Delta} - \hat{S}_\tau = -(S_{t+\Delta} - S_t) = -dS_t \\ &= -\sigma S_t (W_{t+\Delta} - W_t) = -\sigma \hat{S}_{\tau+\Delta}(\hat{W}_\tau - \hat{W}_{\tau+\Delta}) \\ &= \sigma(\hat{S}_\tau + d\hat{S}_\tau) d\hat{W}_\tau. \end{aligned} \tag{4.3.13}$$

Iterating (4.3.13) and eliminating zero terms we have

$$d\hat{S}_\tau = \sigma^2 \hat{S}_\tau d\tau + \sigma \hat{S}_\tau d\hat{W}_\tau. \tag{4.3.14}$$

We see that although S_t is a martingale its time reversal $\hat{S}_\tau$ is not.

Now we turn to $\hat{N}_\tau$. Using Itô's formula we have

$$\begin{aligned} d\hat{N}_\tau &= \frac{\partial v_x}{\partial t} d\tau + \frac{1}{2} \frac{\partial^2 v_x}{\partial x^2} (d\hat{S}_\tau)^2 + \frac{\partial v_x}{\partial x} d\hat{S}_\tau \\ &= \left[\frac{\partial v_x}{\partial t} + \frac{\partial v_x}{\partial x} \sigma^2 \hat{S}_\tau + \frac{1}{2} \frac{\partial^2 v_x}{\partial x^2} \sigma^2 \hat{S}_\tau^2 \right] d\tau + \frac{\partial v_x}{\partial x} \sigma S_\tau d\hat{W}_\tau. \end{aligned} \tag{4.3.15}$$

Differentiating (4.3.5) with respect to x we have

$$\frac{\partial v_x}{\partial t} + \frac{\partial v_x}{\partial x} \sigma^2 x + \frac{1}{2} \frac{\partial^2 v_x}{\partial x^2} \sigma^2 x^2 = 0.$$

It follows that

$$d\hat{N}_\tau = \frac{\partial v_x}{\partial x} \sigma S_\tau d\hat{W}_\tau \tag{4.3.16}$$

is a martingale.

Finally we consider the time reversal of the hedging portfolio (cash borrowed) process $H_t = v^*(N_t, t)$. Using the dual Itô formula (4.1.9) we have

$$\begin{aligned} dv &= v_t dt + N_t dS_t + \frac{1}{2} d\langle S, N\rangle_t \\ dH_t &= dv^* = v_t^* dt + S_t dN_t + \frac{1}{2} d\langle S, N\rangle_t. \end{aligned} \tag{4.3.17}$$

Combining (4.3.17) with the riskless gain condition $dv = N_t dS_t$ and $v_t + v_t^* = 0$ from (4.1.6) we have

$$\begin{aligned} dH_t &= H_{t+\Delta} - H_t = S_t dN_t + d\langle S, N\rangle_t \\ &= (S_t + dS_t) dN_t = S_{t+\Delta}(N_{t+\Delta} - N_t). \end{aligned} \tag{4.3.18}$$

Letting $t + \tau + \Delta = T$ we have

$$\hat{H}_\tau - \hat{H}_{\tau+\Delta} = \hat{S}_\tau (\hat{N}_\tau - \hat{N}_{\tau+\Delta})$$

or

$$d\hat{H}_\tau = \hat{S}_\tau d\hat{N}_\tau = \frac{\partial v_x}{\partial x} \sigma S_\tau^2 d\hat{W}_\tau. \tag{4.3.19}$$

Thus, $\hat{H}_\tau$ is also a martingale.

4.4 Generalized Duality and Hedging with Contingent Claims

Financial innovations in the past several decades have led to the creation of many new types of financial derivatives. They become increasingly liquid and, thus, can also be used as hedging devices. What happens when we use a contingent claim instead the underlying to construct a hedging portfolio for the purpose of pricing and hedging a target contingent claim? It turns out that a duality also emerges between the value of the target contingent claim and the cash borrowed process in terms of generalized duality which naturally corresponds to a generalized convexity concept (see e.g. [39]). Moreover, similar to the classical option pricing theory, the no arbitrage value of the contingent claim derived this way preserves the generalized convexity of the terminal payoff.

4.4.1 Preservation of Generalized Convexity in the Value Function of a Contingent Claim

Consistency of Generalized Convexity

Let S_t be a diffusion process

$$dS_t = \mu(S_t, t)dt + \sigma(S_t, t)dW_t, \tag{4.4.1}$$

where W_t is a standard Brownian motion. We assume again that the risk free rate is 0. Consider a target contingent claim on S_t of European style with maturity at $T > 0$ and a terminal payoff $f(S_T)$ at $t = T$. Suppose that a different contingent claim, we call it hedging claim, on S_t is traded on the market with price $p(S_t, t)$ at all time $t \in [0, T]$. For uniqueness in what follows we always assume that p and v are smooth functions bounded by $\alpha e^{\beta x^2}$ for some $\alpha, \beta > 0$. Our main result is:

Theorem 4.4.1 (Consistency of Generalized Convexity) *Define* $c_t(x, y) = p(x, t)y$ *and assume that* f *is* $\Phi_{c_T(1)}-$*convex. Then*

(i) Partial differential equation $v_t + \frac{\sigma^2}{2}v_{xx} = 0$, $v(x, T) = f(x)$, *uniquely determines an arbitrage free price for the target claim;*
(ii) for any $t \in [0, T]$, $v(\cdot, t)$ *is* $\Phi_{c_t(1)}-$*convex; and*
(iii) N_t *determined by*

$$v(N_t, t)^{c_t(1)} + v(S_t, t) = p(S_t, t)N_t,$$

makes the portfolio of the hedging instrument and the riskless asset $p(S_t, t)N_t - v^{c_t(1)}(N_t, t)$ *riskless.*

Proof. We price v by forming a potentially self-financing portfolio of statically shorting one share of the target contingent claim with N_t units of the hedging claim. Then

$$p(S_t, t)N_t - v(S_t, t). \tag{4.4.2}$$

is the cash borrowed resulting from this portfolio. Self-financing implies that

$$N_t dp(S_t, t) = dv(S_t, t). \tag{4.4.3}$$

Applying the Itô formula we get

$$\begin{aligned} N_t &\left[\left(p_t + \mu p_x + \frac{\sigma^2}{2} p_{xx}\right) dt + p_x \sigma dW_t\right] \\ - &\left(v_t + \mu v_x + \frac{\sigma^2}{2} v_{xx}\right) dt + v_x \sigma dW_t \end{aligned} \tag{4.4.4}$$

To ensure riskless gains we need N_t to satisfy the equation

$$v_x(S_t, t) = N_t p_x(S_t, t). \tag{4.4.5}$$

Then the gain in portfolio reduces to

$$N_t \left(p_t + \frac{\sigma^2}{2} p_{xx}\right) dt - \left(v_t + \frac{\sigma^2}{2} v_{xx}\right) dt.$$

Now no arbitrage requires this quantity to be 0. Thus

$$N_t \left(p_t + \frac{\sigma^2}{2} p_{xx}\right) dt = \left(v_t + \frac{\sigma^2}{2} v_{xx}\right) dt.$$

Since p is arbitrage free,

$$p_t + \frac{\sigma^2}{2} p_{xx} = 0.$$

Thus, v must also satisfy the Black-Scholes PDE

$$v_t + \frac{\sigma^2}{2} v_{xx} = 0. \tag{4.4.6}$$

with terminal condition

$$v(x, T) = f(x), \tag{4.4.7}$$

where f is the payoff of the target at T.

We show that $v^{c_t(1)c_t(2)}$ satisfies the same Black-Scholes PDE as v does. Observe that $x \to p(x, T)$ is strictly monotone, which implies that $x \to p(x, t)$ is invertible, i.e., $x = x(p, t)$. We can define

$$\tilde{v}(p, t) = v(x(p, t), t) + \iota_{\mathrm{range}(p(\cdot, t))}(p).$$

Then we have

$$\begin{aligned} \tilde{v}^*(N_t, t) &= \sup_{P_t} [P_t N_t - \tilde{v}(P_t, t)] \\ &= \sup_{S_t} [p(S_t, t)N_t - v(S_t, t)] = v^{c_t(1)}(N_t, t). \end{aligned}$$

Similarly, for any $P_t = p(S_t, t)$,

$$\begin{aligned}\tilde{v}^{**}(P_t, t) &= \sup_{N_t}[P_t N_t - \tilde{v}^*(N_t, t)] \\ &= \sup_{N_t}[p(S_t, t)N_t - v^{c_t(1)}(N_t, t)] = v^{c_t(1)c_t(2)}(S_t, t).\end{aligned}$$

Thus, we need only to show that $\tilde{v}$ and $\tilde{v}^{**}$ satisfy the same Black-Scholes PDE. We do so through the PDE for the cash borrowed $\tilde{v}^*$. Changing variables we have

$$\begin{aligned}\frac{\partial v}{\partial t} &= \frac{\partial \tilde{v}}{\partial t} + \tilde{v}_p \frac{\partial p}{\partial t} \\ v_x &= \tilde{v}_p p_x \\ v_{xx} &= \tilde{v}_p p_{xx} + \tilde{v}_{pp} p_x^2.\end{aligned}$$

Substituting them into

$$\frac{\partial v}{\partial t} + \frac{\sigma^2}{2} v_{xx} = 0$$

and using

$$\frac{\partial p}{\partial t} + \frac{\sigma^2}{2} p_{xx} = 0$$

we have

$$\frac{\partial \tilde{v}}{\partial t} + \frac{\sigma^2 p_x^2}{2} \tilde{v}_{pp} = 0. \tag{4.4.8}$$

Thus, using Fenchel equality

$$\tilde{v}(P_t, t) + \tilde{v}^*(N_t, t) = P_t N_t$$

we have

$$n = \tilde{v}_p, p = \tilde{v}^*_n, \frac{\partial \tilde{v}}{\partial t} = -\frac{\partial \tilde{v}^*}{\partial t}$$

and

$$\tilde{v}_{pp} \tilde{v}^*_{nn} = 1.$$

Substituting the above into (4.4.8) we derive

$$-\frac{\partial \tilde{v}^*}{\partial t} + \frac{\sigma^2 p_x^2 \tilde{v}_{pp}^2}{2} \tilde{v}^*_{nn} = 0. \tag{4.4.9}$$

To derive the PDE for $\tilde{v}^{**}$ we start from P_t and N_t satisfying the Fenchel equality

$$\tilde{v}^{**}(P_t, t) + \tilde{v}^*(N_t, t) = P_t N_t.$$

Then we have

$$n = \tilde{v}^{**}_p, p = \tilde{v}^*_n, \frac{\partial \tilde{v}^{**}}{\partial t} = -\frac{\partial \tilde{v}^*}{\partial t}$$

and

$$\tilde{v}^{**}_{pp} \tilde{v}^*_{nn} = 1.$$

Since $\tilde{v}_{pp} \tilde{v}^*_{nn} = 1$ substituting the above relationship into (4.4.9) yields

$$\frac{\partial \tilde{v}^{**}}{\partial t} + \frac{\sigma^2 p_x^2}{2} \tilde{v}^{**}_{pp} = 0.$$

We see that $\tilde{v}$ and $\tilde{v}^{**}$ satisfy the same Black-Scholes differential equaiton. Since $v(x,t) = \tilde{v}(p,t)$ and $\tilde{v}^{**}(p,t) = v^{c_t(1)c_t(2)}(x,t)$ for $x = x(p,t)$ we conclude that $v(x,t)$ and $v^{c_t(1)c_t(2)}(x,t)$ also satisfy the same Black-Scholes differential equation.

Finally, since $v(\cdot,T)$ is $\Phi_{c_T(1)}-$convex we have $v(x,T) = v^{c_T(1)c_T(2)}(x,T)$. That is v and $v^{c_T(1)c_T(2)}$ satisfy the same terminal condition. Thus, they must be the same for all t, i.e. $v(x,t) = v^{c_t(1)c_t(2)}(x,t)$ so that $v(\cdot,t)$ is $\Phi_{c_t(1)}-$convex. ●

Remark 4.4.2 Function $c_t(x,y) = p(x,t)y$ is known when we know the price of claim p that we use to hedge.

Fixing t and defining $\tilde{v}(p,t) = v(x(p,t),t)$, we can represent the portfolio $p(S_t,t)N_t - v(S_t,t)$ graphically in Fig 4.4-4.5

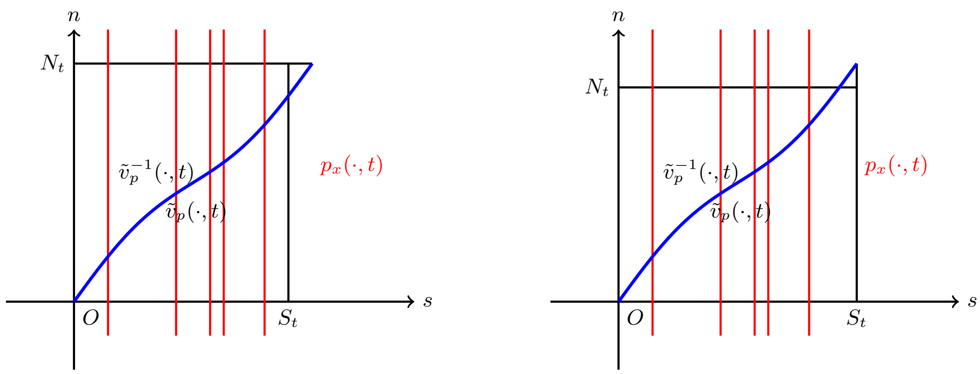


Fig. 4.4. Hedging portfolio

We see that these graphs are almost exact replications of the graphic representation of the hedging portfolio $S_tN_t - v(S_t,t)$. The only difference is that the sn-plane is weighted by $p_x(\cdot,t)$. This implies the following generalized Fenchel duality relationship .

$$v^{c_t(1)}(N_t,t) = \sup_{S_t}[p(S_t,t)N_t - v(S_t,t)] \tag{4.4.10}$$

and

$$v(S_t,t) = \sup_{N_t}[p(S_t,t)N_t - v^{c_t(1)}(N_t,t)]. \tag{4.4.11}$$

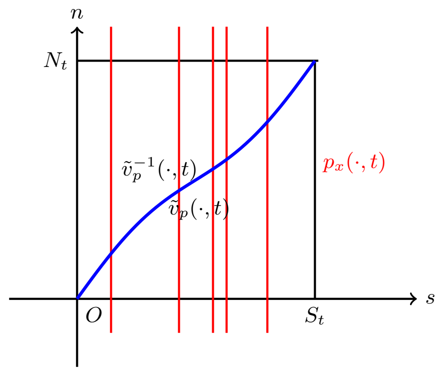


Fig. 4.5. Equality holds when $p_x(S_t,t)N_t = v_x(S_t,t)$

Relationship (4.4.10) can be interpreted as a cash borrowed process having the property of riskless gains and equation (4.4.11) shows that the hedging portfolio $p(S_t,t)N_t - v^{c_t(1)}(N_t,t)$ of the hedging claim and cash is self-financing. The key of the formal proof of Theorem 4.4.1 is to verify that $v(\cdot,t)$ is $\Phi_{c_t(1)}$−convex.

4.4.2 Determining the Hedging Process

While in principle the PDE with terminal condition (4.4.6) and (4.4.7) determines an arbitrage free and $\Phi_{c_t(1)}$-convexity preserving contingent claim pricing function v, to determine the hedging process one must know the dynamics of N_t and $H_t = v(\cdot,t)^{c_t(1)}(N_t)$.

Defining $n(x,t) := v_x(x,t)/p_x(x,t)$, equation (4.4.5) implies that the hedging process is

$$N_t = n(S_t,t). \tag{4.4.12}$$

Differentiating (4.4.6) with respect to x we derive the PDE governing n:

$$n_t + \frac{\sigma^2}{2} n_{xx} = -\frac{n_x \sigma}{p_x}(p_x\sigma)_x. \tag{4.4.13}$$

We turn to the hedging process N_t. Using Itô's formula we have

$$dN_t = \left(n_t + \mu n_x + \frac{\sigma^2}{2} n_{xx}\right) dt + n_x \sigma dW_t \tag{4.4.14}$$

Using (4.4.13) we can simplify (4.4.14) to

$$dN_t = n_x \left(\left[\mu - \sigma\frac{(p_x\sigma)_x}{p_x}\right] dt + \sigma dW_t\right) \tag{4.4.15}$$

We see that N_t is in general not a martingale unless $\mu - \sigma\frac{(p_x\sigma)_x}{p_x} = 0$.

Next we discuss the dynamic of the cash borrowed process H_t. We have seen that no arbitrage forces $v(\cdot,t) = v(\cdot,t)^{c_t(1)c_t(2)}$. Thus, by (4.4.10) and (4.4.11) we have

$$H_t(N_t) + v(S_t,t) = p(S_t,t)N_t. \tag{4.4.16}$$

Due to the self-financing condition (3.2.4) we have

$$\begin{aligned} dH_t &= p dN_t + d\langle p, N\rangle_t \\ &= n_x p\sigma dW_t + \sigma^2\left[p_x^2 n_x + \frac{1}{2}p_{xx}p_x^2 n - pn_x\frac{(p_x\sigma)_x}{p_x\sigma}\right]dt \end{aligned} \tag{4.4.17}$$

In general H_t is not a martingale. However, in some special case it could be. For example, if $p(x,t) = x$, i.e. the hedging is done with the price process S_t itself then $p_x = 1, p_{xx} = 0$ and equation (4.4.17) is simplified to

$$dH_t = \sigma n_x\left(S_t dW_t + [\sigma - S_t\sigma_x]\,dt\right). \tag{4.4.18}$$

Now when S_t follows a geometric Brownian motion where $\sigma(x,t) = \sigma(t)x$, we have $\sigma = x\sigma_x$ and H_t is a martingale.

4.4.3 Hedging with $p-$multiple ETF

Double long and short ETFs and triple long and short ETFs are now available for many major indices. They provides convenient tools for hedging. We discuss in this section the general $p-$multiple long ETF as a hedging tool. We will need the following special case of Theorem 2.2.3.

Proposition 4.4.3 *The function $x^q, x \ge 0$ is $\Phi_{[x^p y](1)}-$convex if either $q > 0$ and $p < q$ or $q < 0$ and $q < p$. Similarly, the function $-x^q, x \ge 0$ is $\Phi_{[x^p y](1)}-$convex if either $p > q > 0$ or $p < q < 0$.*

Proof. We prove only for the case x^q. The discussion for $-x^q$ is similar. Let $u(x) = x^q, x \ge 0$. It is easy to calculate that

$$R(x) = -\frac{xu''(x)}{u'(x)} = 1 - q. \tag{4.4.19}$$

When $q > 0$ and $p < q$, u is an increasing function and $R(x) = 1 - q < 1 - p$ and when $q < 0$ and $p > q$, u is a decreasing function and $R(x) = 1 - q > 1 - p$. Now the conclusion of the proposition directly follows that of Theorem 2.2.3. ●

Suppose S_t satisfies the diffusion process

$$dS_t = \sigma S_t dB_t. \tag{4.4.20}$$

Consider an Enropean styled option with payoff S_T^q at $t = T$. Denote the value of this option at time t by $v(S_t, t)$. Solving (4.4.6) with terminal condition $v(S_T, T) = S_T^q$, we can determine that

$$v(S_t, t) = S_t^q e^{\frac{q(q-1)}{2}\sigma^2(T-t)}.$$

It is easy to verify that

$$\frac{dv(S_t, t)}{v(S_t, t)} = q\frac{dS_t}{S_t}.$$

Thus, v is a q-multiple of S_t. Similarly, a p-multiple of S_t has a no arbitrary price

$$P_t = S_t^p e^{\frac{p(p-1)}{2}\sigma^2(T-t)}.$$

Theorem 4.4.4 (Hedging with Multiple of ETF) *Let S_t be the price of an asset satisfying the diffusion equation (4.4.20). Suppose that either $q > 0$ and $p < q$ or $q < 0$ and $q < p$. Then a q-multiple long ETF of $S_t, t \in [0, T]$ can always be dynamically hedged with an arbitrage free self-financing portfolio involving a p-multiple ETF of S_t. Moreover, for any $t \in [0, T]$, the arbitrage free price of the q-multiple ETF is $\Phi_{[x^p y](1)}$-convex.*

Proof. By Theorem 4.4.1 we need only to check that $v(x, T) = x^q$ is $\Phi_{[x^p y](1)}$-convex. This follows directly from Proposition 4.4.3. ●

In this case we can explicitly calculate that the hedging process is

$$N_t = \frac{q}{p} S_t^{q-p} e^{[\frac{q(q-1)}{2} - \frac{p(p-1)}{2}]\sigma^2(T-t)}$$

and the cash borrowed process is

$$H_t = \frac{q-p}{p} v(S_t, t).$$

Note that the cash borrowed process is always a martingale. In particular, for $q = 4$ and $p = 2$, we see that the no arbitrage price of the quadruple long ETF at any given time $t \in [0, T]$ is $\Phi_{[x^2 y](1)}-$convex and such a process can be hedged by a double ETF.

Remark 4.4.5 It is worthy to observe that when $q \in (0, 1)$ and $p < q$ the $\Phi_{[x^p y](1)}$-convex functions are, in fact, concave. We can see that $\Phi_{[x^p y](1)}$-convex functions represent a wide spectrum of convex and concave functions with different strengths. A few graphic illustrations are included in Fig 4.6–4.9.

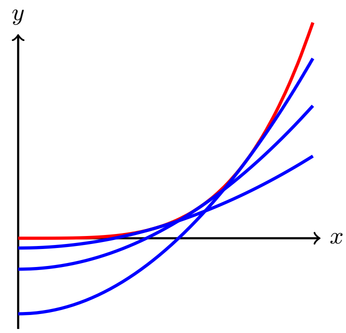


Fig. 4.6. Graphic illustration of $q = 4$ and $p = 2$

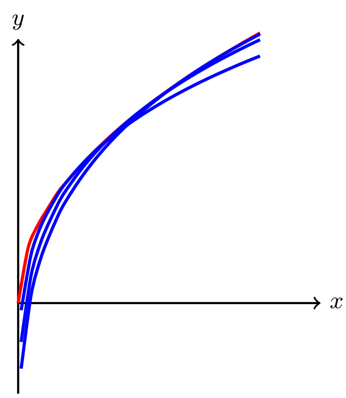


Fig. 4.7. Graphic illustration of $q = 1/2$ and $p = 1/4$

The above discussion can be applied to q-multiple short ETF of S_t. We summarize the result in the following Theorem.

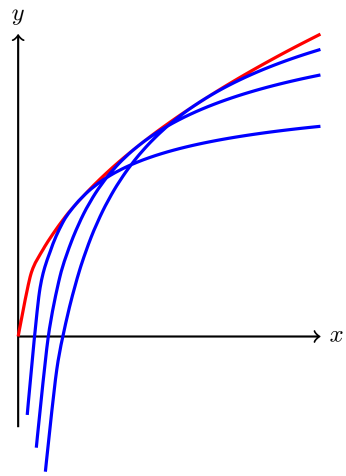


Fig. 4.8. Graphic illustration of $q = 1/2$ and $p = -1/2$

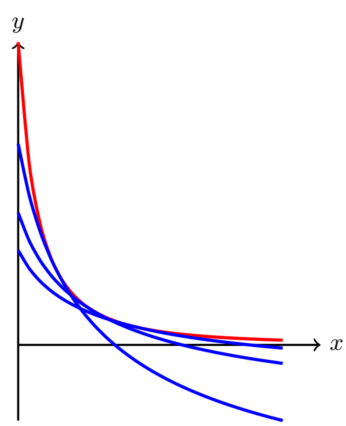


Fig. 4.9. Graphic illustration of $q = -2$ and $p = -1/2$

Theorem 4.4.6 *Let S_t be the price of an asset satisfying the diffusion equation (4.4.20). Suppose that either $p > q > 0$ or $p < q < 0$ and $q < p$. Then a q-multiple short ETF of $S_t, t \in [0, T]$ can always be dynamically hedged with an arbitrage free self-financing portfolio involving a p-multiple long ETF of S_t. Moreover, for any $t \in [0, T]$, the arbitrage free price of the q-multiple short ETF is $\Phi_{[x^p y](1)}$-convex.*

Proof. The proof is the same as that of the proof of Theorem 4.4.4 except we need to use the second part of Proposition 4.4.3. ●

Generalized convexity is also shows up in other financial related functions. The following are two simple examples.

Example 4.4.7 (Stock Price as a Contingent Claim of Company's Asset)

Leland proposed the following perspective of stock price in [31]. Consider a company's activity has value a_t at $t \in [0, +\infty)$ with dynamics

$$da_t = \sigma a_t dW_t.$$

where σ is a constant. Assume that the risk free rate is r and that there is no dividend. Let's first view the stock price $S(a_t)$ as a perpetual claim on a_t. Then $S(a_t)$ satisfies the ordinary differential equation

$$\frac{\sigma^2 x^2}{2} S_{xx} + rxS_x - rS = 0.$$

So that

$$S(a_t) = ba_t - ca_t^q,$$

where $q = -r/\sigma^2 < 0, b, c > 0$.

Now suppose that the company has outstanding bond maturing at T with a total amount K. Then the stock price u becomes a contingent claim on a_t with terminal payoff

$$u(a_T, T) = (ba_T - ca_T^q - K)^+.$$

It is easy to check that for x sufficiently large u is an increasing function and

$$-\frac{xu''(x)}{u'(x)} \leq 1 - q.$$

Thus, for K sufficiently large $u(x, T)$ is a $\Phi_{[x^q y](1)}-$convex function. It follows from Theorem 4.4.1 that, $u(\cdot, t)$ is also $\Phi_{[x^q y](1)}-$convex.

Example 4.4.8 (Normal Kernel) Consider the scaled normal kernel

$$n(x) = e^{-kx^2/2}, \; x \geq 0, k > 0.$$

We can verify that $-xn''(x)/n'(x) = kx^2 - 1 \geq -1$ but there is no upper bound. Thus, the decreasing function $e^{-kx^2/2}, x \geq 0$ is $\Phi_{[x^p y](1)}$-convex for any $p \geq 2$. Due to the symmetry of both $e^{-kx^2/2}$ and $|x|^p y - b$ with respect to the vertical axis we conclude that this property also holds when $x < 0$. So that $e^{-x^2/2}$ is $\Phi_{[|x|^p y](1)}$-convex for any $p \geq 2$..

We note that in both Example 4.4.7 and Example 4.4.8 the functions involved are neither convex nor concave.

4.4.4 Reducing the Volatility of the Hedging Process

When there are multiple hedging claims available in the market, it is usually the case that for a given target contingent claim there are many different ways to hedge. Choosing an appropriate hedging device that fits better in generalized convexity often can help reducing the volatility of the hedging process.

Example 4.4.9 (Hedging q-multiple Long ETF Using p-multiple) *Suppose that S_t is a diffusion process*

$$dS_t = \sigma S_t dW_t, t \in [0, T].$$

Let v be the value of the q-multiple long ETF of S_t. Suppose either $q > 0, p < q$ or $q < 0$,$p > q$ Then the process for the hedging shares has been explicitly calculated as

$$N_t = \frac{q}{p} S_t^{q-p} e^{[\frac{q(q-1)}{2} - \frac{p(p-1)}{2}]\sigma^2 (T-t)}$$

and the cash borrowed process is

$$H_t = \frac{q-p}{p} v(S_t, t).$$

Note that the closer the p to q, the smoother the cash borrowed process H_t which is an proxy for the value of the hedging portfolio.

Example 4.4.10 (Normal Kernel) Now consider S_t following a Bachelier model $S_t = \sigma W_t$ and let $v(S_t, t), t \in [0, T]$ be the no arbitrage price of a contingent claim with payoff $f(x) = e^{-x^2/2}$ at T.

It is easy to directly calculate that

$$v(S_t, t) = \frac{1}{\sqrt{\sigma^2(T-t)+1}} \exp\left(-\frac{S_t^2}{2(\sigma^2(T-t)+1)}\right).$$

In this case, we can dynamically replicate v using either S_t (v is not convex in S_t) or its double long ETF $P_t = S_t^2 + \sigma^2(T-t)$ with respect to which v is convex.

When hedging with S_t we can calculate that share of hedging $N_t^S = v_x = -S_t v/(\sigma^2(T-t)+1)$. The cash borrowed process is

$$H_t^S = S_t N_t^S - v = -\frac{S_t^2 + \sigma^2(T-t)+1}{\sigma^2(T-t)+1} v.$$

When hedging with P_t we can similarly calculate that the share of hedging $N_t^P = v_x/p_x = -v/2(\sigma^2(T-t)+1)$. The cash borrowed process becomes

$$H_t^P = S_t N_t^P - v = -\frac{S_t^2/2 + 3\sigma^2(T-t)/2+1}{\sigma^2(T-t)+1} v.$$

We can see that hedging with P_t results in a smoother cash borrowed process because the random change related to the uncertain stock price is only half that of hedging with S_t.

4.4.5 The Volatility Trade

Now consider S_t following a diffusion process

$$dS_t = \sigma_t S_t dW_t, t \in [0, T].$$

Let us assume that the volatility σ_t^2 is unknown. We further assume that the market implies a constant volatility σ_h^2 which is, say, known to be too high by a certain trader. Can he take advantage of the situation? Carr and Madan has shown in [11] that the answer is yes if there is a contingent claim whose no arbitrage price $v(S_t, t)$ is convex in S_t.

In this example we show that generalize convexity can help us to derive a similar volatility trade when $v(S_t, t)$ has a certain generalized convexity properties. Let $p(S_t, t)$ be the no arbitrage price of a hedging claim with $p(\cdot, t)$ strictly monotone. Let $c_t(x, y) = p(x, t)y$. We assume that $v(\cdot, T)$ is $\Phi_{c_T(1)}$−convex but not necessarily convex in S_t such as in Examples 4.4.7 and 4.4.8.

Denote again

$$\tilde{v}(p, t, \sigma_h) = v(x(p, t), t, \sigma_h).$$

We have already seen that $\tilde{v}(p, t, \sigma_h)$ is convex in p. Here σ_h is added to emphasize that the trader views that $\tilde{v}(p, t, \sigma_h)$ follows the constant volatility σ_h implied by the market in trading.

Itô's formula tells us that

$$v(S_T, T) - v(S_t, t) - \int_t^T \tilde{v}_p dP_t = \int_t^T \left[\frac{\partial \tilde{v}}{\partial s} + \frac{\tilde{v}_{pp} p_x^2 S_s^2}{2} \sigma_s^2\right] ds.$$

The left hand is the trading portfolio and the right hand is the P&L. Since the trader follows the constant volatility σ_h implied by the market in trading

$$\frac{\partial \tilde{v}}{\partial t} = -\frac{\tilde{v}_{pp} p_x^2 S_t^2}{2} \sigma_h^2.$$

Thus,

$$P\&L = \int_t^T \frac{\tilde{v}_{pp} p_x^2 S_s^2}{2} (\sigma_s^2 - \sigma_h^2) ds$$

where $v_{pp} > 0$. We see that the trader can take advantage of the over estimation on volatility by the market by dynamically trading the portfolio

$$v(S_T, T) - v(S_t, t) - \int_t^T \tilde{v}_p dP_t.$$

Comments

Chapter 1 Sections 1.1-1.4 give a concise summary of standard convex analysis duality theory, which is pioneered by Fenchel [18], Moreau [41] and Rockafellar [43]. Our exposition follows [9, 21] emphasizing the variational approach by focusing on convex programming. We also highlight the role of subdifferential of the optimal value function as the set of Lagrange multipliers and the set of dual solutions.

Generalized convexity, conjugacy and related duality discussed in Section 1.5 can be traced back to Moreau. It gained more attention recently due to diverse applications and also due to its role in mass transport theory [61]. Our main references here are [30, 39]. Their applications in hedging with contingent claims are discussed in Section 4.4.

Chapter 2 Section 2.1 provides a unified treatment of the classical Markowitz portfolio theory [36], CAPM model [51] and Sharpe ratio [52]. Following [67] we emphasize that the underlying mathematical tools for all these applications are minimizing a quadratic function with linear constraint, a simplest form of convex programming. Convex duality is essential in revealing the structure of the solutions with a practical financial meaning.

Section 2.2 deals with the portfolio problem from the perspective of utility optimization. Utility function has a long history goes back to the work of Bernoulli [4] in 1738 related to the St. Petersburg paradox. The relevance to financial problem comes in as optimizing the utility of a portfolio simultaneously accounts for investors pursuing capital growth and risk aversion. The concavity of utility functions means convex analysis is essential. Different agents have different degree of risk aversion. They can be measured by using either absolute risk aversion coefficients or relative risk aversion coefficients [1, 49]. Interestingly, utility functions with those risk aversion coefficients bounded at a given level can be characterized by generalized convexity discussed in Section 1.5. These new characterizations are included in subsection 2.2.2.

Growth optimal portfolio theory [32] and Kelly'es criterion [27, 34, 35, 56, 57, 58] as a money management tool in investment are discussed as an illustration of such utility optimization problems. In particular, following [66] we highlights that optimizing the expected log utility for a portfolio of cash and a given investment strategy on historical performance data amounts to measure the useful information implied by the investment strategy and can be used as a measure to compare different investment strategies.

Fundamental theorem of asset pricing (FTAP) relates no arbitrage to the existence of a martingale measure that can be used to price assets in a financial market. Cox, Ross and Rubinstein observed such a principle in their classical work related to option pricing in complete markets [12, 13]. General FTAPs were discussed in [23, 22, 29, 15] with progressing generality, usually with a proof based on separation arguments. Dybvig and Ross [17] observed that in an incomplete market the martingale measures are related to the risk aversion of market agent. In Section 2.3 we approach the FTAP from the perspective of convex duality. We show that in an incomplete market, a martingale measure is, in fact, a scaling of the dual solution to a portfolio utility maximization problem. We also illustrate with example that this relationship helps us to understand that in an incomplete market, a martingale measure provides a reference price for a certain agent to improve their utility rather than arbitrage.

Section 2.4 deals with risk measures, a concept that plays important roles for both financial institutions and regulatory agencies. Diversification reduces risk implies the convexity of risk measures. We focus on coherent risk measures proposed by Artzner, Delbaen, Eber and Heath in [2]. Coherent risk measures are sublinear, a particular type of convex function. Duality is involved in providing a dual characterization of a coherent risk measure as the conjugate of an indicator function of a cone, called acceptance cone. Interestingly, the generating set for the acceptance cone is closely related to the practice of stress tests. Convex duality also provides several equivalent description of the coherent risk measures in terms of linear preference and value bonds. Moreover, the same argument is at the core of the discussion of good deal in financial markets as explained in Jaschke and Küchler [24]. Beside providing a framework to understand risk measures and their relationship with other important financial concepts, convex duality methods also help to amend widely used nonconvex risk measure value at risk [25] to the convex conditional value at risk proposed by Rockafellar and Uryasev in [44, 45].

Chapter 3 Sections 3.1–3.3 demonstrate that many of the results in the previous chapter also persist in the more general setting of a multiperiod economy. We use the general model laid out in S. Roman's textbook [47].

Section 3.4 discusses super hedging (and symmetrically subhedging) bounds in incomplete markets. Following Kahalé [26] we show that the super hedging bound of a given contingent claim has a dual representation. On one hand it is the supremum of all the prices derived through martingale measure and on the other hand it can be represented as the cost of the smallest super hedging portfolio. When the sample space is finite, the super hedging portfolio in the second representation can be derived by solving a linear programming problem. The linear programming duality can also be used to analyze narrowing the gap between the super and sub- hedging bounds by adding contingent claims with known prices. When discussing contingent claims related to currency spread, incomplete markets may arise from complete markets. Considering supper hedging bounds in this kind of problems usually leads to a Kantorovich mass transportation problem [61]. We illustrate the solution process with an example on a finite sample space using linear programming duality.

Section 3.5 discusses a model for financial markets with bid and ask spread. The main difference with a simplified one price financial market is that the attainable payoff set due to trading is, in general, a convex cone rather than a subspace. This leads to the title conic finance as coined by Madan in [37, 38]. Besides a concise representation of the basic conic finance model, we also discuss new refined fundamental theorem of asset pricing as well as super and sub-hedging price bounds. These results are taken from [60] emphasizing the role of convex duality.

Chapter 4 Section 4.1 summarizes facts on continuous models that we need later. To be concise we are satisfied with a heuristic description of most of the material. Readers interested in further details may consult [42, 55]. The dual Itô formula is a first taste of the role of duality in continuous model. It develops the generalized Itô formula using quadratic covariance in [19].

Section 4.2 discusses convexity and generalized convexity emerged in Bachelier and Black-Scholes formulae. The importance of these convexity properties is highlighted in applying them in the computation of Greeks and in illustrating the delta hedging is, in fact, the Fenchel-Legendra transform of the pricing formula. This is the observation in Carr [10] for more general settings and discussed in greater detail in Section 4.3.

It turns out that if one hedges using a contingent claim rather than the underlying itself, similar duality still persists in the sense of generalized duality that we discuss in Section 4.4. The general principles are summarized in Section 4.4.1 and 4.4.2. A number of examples are included to illustrate their applications in financial practice. How to hedge with the popular multiple ETFs of indices are discussed in detail in Section 4.4.3. What are also discussed in this section are examples of generalized convexity of Leland's model of stock price as contingent claims of company's assets [31] and the general convexity of the normal kernel. The common theme here is that they all follow from characterizations of the generalized convexity using the relative risk aversion coefficient and the absolute risk aversion coefficient. Hedging with derivatives can help to reduce the risk and to expand the range of volatility trading which is proposed in [11]. These are discussed in Sections 4.4.4 and 4.4.5, respectively. Much of the materials regarding to these duality and generalized duality relationships appear here for the first time. We believe that this is an area that worthy further attention.

Inaddition, survey papers [14, 48, 64, 67] have also been valuable references.

References

1. K. J. Arrow, Aspects of the Theory of Risk Bearing. The Theory of Risk Aversion. Helsinki: Yrjo Jahnssonin Saatio. Reprinted in: Essays in the Theory of Risk Bearing, Markham Publ. Co., Chicago, 1971, 90109, 1965.
2. P. Artzner, F. Delbaen, J.-M. Eber and D. Heath, Coherent measures of risk, Mathematical Finance, 9 (1999), 203228.
3. L. Bachelier, "Théorie de la spéculation", Annales Scientifiques de l'École Normale Supérieure, 3 (17), (1900) pp. 2186.
4. D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22:23–36, 1954/1738.
5. F. Black and M. Scholes, The pricing of options and corporate liabilities, Journal of Political Economy, 81:637–645, 1973
6. T. Bjork, *Arbitrage Theory in Continuous Time*, Oxford University Press, New York, 2009.
7. J.M. Borwein and A.S. Lewis, *Convex analysis and nonlinear optimization*, Springer-Verlag, 2000. Second edition, 2005.
8. J.M. Borwein and Q.J. Zhu, *Techniques of Variational Analysis*, Springer-Verlag, 2005.
9. J.M. Borwein and Q.J. Zhu, A variational approach to Lagrange multipliers, J Optim Theory Appl (2016) 171: 727-756. doi:10.1007/s10957-015-0756-2
10. P. Carr, Option as Optimization: A Dual Approach to Derivatives Pricing, Quant USA, New York, July, 2014.
11. P. Carr and D. Madan, Toward a theory of volatility trading, 2002.
12. J. Cox and S. Ross. The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, 3:144–166, 1976.
13. J. Cox, S. Ross, and M. Rubinstein. Option pricing: a simplified approach. *Journal of Financial Economics*, 7:229–263, 1979.
14. K. R. Dahl, Convex Duality and Mathematical Finance, Thesis for M. Sci. at University of Oslo, 2012.
15. F. Delbaen and W. Schachermayer. A general version of the fundamental theorem of asset pricing. *Math. Ann.*, 300:463–520, 1994.
16. S. Doleski and S. Kurcyusz, On $\Phi-$ convexity in extremal problems, SIAM J. Control and Optimization, 16, (1978) 277-300.
17. P. Dybvig and S. A. Ross. Arbitrage, state prices and portfolio theory. *Handbook of the Economics of Finance*, 2003.

18. W. Fenchel, Convex cones, sets and functions, Lecture Notes, Princeton University, Princeton, NJ, 1951.
19. Föllmer H., Protter P., and Shiryaev A.N., "Quadratic covariation and an extension of Itô's formula", *Bernoulli*, **1**: 149–169, 1995.
20. A. H. Hamel and F. Heyde, Duality for set-valued measures of risk, SIAM J. Financial Math. 1 (2010) 66-95.
21. D. Gale, A geometric duality theorem with economic applications, *The Review of Economic Studies*, **34** (1967), 19–24.
22. J.M. Harrison and S. Pliska. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes Appl.*, 11:215–260, 1981.
23. J.M. Harrison and D. M. Kreps, Martingales and arbitrage in multiperiod securities markets, *J. Econom. Theory*, 20 (1979) 381–408.
24. S. Jaschke and U. Küchler, Coherent risk measures and good-deal bounds, Finance Stochast. 5 (2001) 181-200.
25. P. Jorion. *Value at Risk*. McGraw-Hill, New York, 1997.
26. N. Kahalé, Sparse calibrations of contingent claims, Mathematical Finance, 20 (2010) 105-115.
27. J. L. Kelly. A new interpretation of information rate. *Bell System Technical Journal*, 35:917–926, 1956.
28. A.J. King Duality and martingale: a stochastic programming perspective on contingent claims. *Math. Progam., Ser. B*, 91:543–562, 2002.
29. D. Kramkov and W. Schachermayer. The asymptotic elasticity of utility functions and optimal investment in incomplete markets. *Ann. Appl. Probability*, 9:904–950, 1999.
30. S.S. Kutateladze and A.M. Rubinov, Minkowski duality and its applications, Russian Math. Surveys, 27 (1972) 137-192.
31. H. Leland, Corporate debt value, bond covenants, and optimal capital structure. Journal of Finance, 49(4):1213–1252, September 1994.
32. J Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47:13–37, 1965.
33. Lopez de Prado, M. Vince, R. and Zhu, Q.J. Optimal Risk Budgeting under a Finite Investment Horizon, (2013) SSRN 2364092.
34. L. C. Maclean, E. O. Thorp, and W. T. Ziemba. Good and bad properties of the Kelly criterion. *in "The Kelly Capital Growth Investment Criterion, Theory and Practice", Maclean, L. C. and Thorp, E. O. and Ziemba, W. T. Editors*, pages 563–574, 2010.
35. L. C. Maclean, E. O. Thorp, and W. T. Ziemba. *Editors: The Kelly Capital Growth Investment Criterion, Theory and Practice*. World Scientific Handbook in Financial Economics Series: Volume 3. World Scientific, Singapore, 2011.
36. H. Markowitz. *Portfolio Selection*. Cowles Monograph, 16. Wiley, New York, 1959.
37. D. Madan, Asset Pricing Theory for Two Price Economies, 2014.
38. D. Madan and W. Schoutens, Applied Conic Finance, Cambridge University Press, 2016.
39. J. E. Martinez-Legaz, Generalized convex duality and its economic applications.
40. R. Merton, Theory of rational option pricing, Bell Journal of Economics and Management Sciences, 4:141–183, 1973.
41. J. J. Moreau, Fonctionelles convexes, College de France, Lecture notes, 1967.

42. B. Oksendal, Stochastic Differential Equations, 6th edition, Springer, 2003.
43. R. T. Rockafellar, Convex Analysis, 1970.
44. R.T. Rockafellar and S. Uryasev, Optimization of conditional value at risk, J. Risk, 2 (2000) 21–41.
45. R. T. Rockafellar and S. Uryasev, Conditional value-at-risk for general loss distributions, Journal of Banking and Finance **26** (2002) 1443–1471.
46. R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, 1998.
47. S. Roman. *Introduction to the Mathematics of Finance.* Springer, New York, 2004.
48. Convex duality in stochastic optimization and mathematical nance, Math. Oper. Res. 36 (2011) 340–362.
49. J. W. Pratt, Risk Aversion in the Small and in the Large, Econometrica. 32 (12): 122136, 1964.
50. C. Shannon and W. Weaver, The Mathematical Theory of Communication. Urbana, Illinois: University of Illinois Press, 1949.
51. W. F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk.
52. W. F. Sharpe. Mutual fund performance. *Journal of Business*, 1:119–138, 1966. *Journal of Finance*, 19:425–442, 1964.
53. S. E. Shreve, *Stochastical Calculus for Finance I*, Springer, 2004.
54. S. E. Shreve, *Stochastical Calculus for Finance II*, Springer, 2004.
55. J. M. Steele, *Stochastic Calculus and Financial Applications*, Springer, 2001.
56. Edward O. Thorp. *Beat the Dealer.* Random House, New York, 1962.
57. Edward O. Thorp. Portfolio choice and the Kelly criterion. *Business and Econom. Statist. Proc. of American Statistical Association*, pages 215–224, 1971.
58. Edward O. Thorp and Sheen T. Kassouf. *Beat the Market.* Random House, New York, 1967.
59. Y. Tsuzuki, On the optimal super and sub-hedging strategies, CARF, preprints, The university of Tokyo, 2013.
60. M. Vazifedan and Q.J. Zhu, No arbitrage principle in conic finance, working paper.
61. C. Villani, Topics in Optimal Transportation, Graduate Studies in Mathematics (58), America Math. Soc. 2003.
62. R Vince. *The New money Management: A Framework for Asset Allocation.* John Wiley and Sons, New York, 1995.
63. Vince, R. and Zhu, Q.J., Optimal Betting Sizes for the Game of Blackjack, SSRN 2324852, to appear Journal of Portfolio Management, 2016.
64. J. Xia and J. A. Yan, Convex duality theory for optimal investment, preprint, 2006.
65. C. Zălinescu, On duality gaps in linear conic problems, Preprint, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 2010. (www.optimization-online.org/DB_HTML/2010/09/2737.html).
66. Q. J. Zhu, Mathematical analysis of investment systems, Journal of Mathematical Analysis and Applications, 326 (2007) 708-720.
67. Q. J. Zhu, Convex Analysis in Mathematical Finance, Nonlinear Analysis: Theory Method and Applications 75, (2012) 1719-1736.

Index