Martingale Optimal Transport and Portfolio Theory

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Vrushali Kumbhar

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2019

Supervisor: Dr. Anup Biswas

Certificate

This is to certify that this dissertation entitled Martingale Optimal Transport and Portfolio Theorytowards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Vrushali Kumbharat Indian Institute of Science Education and Research under the supervision of Dr. Anup Biswas, Associate Professor, Department of Mathematics , during the academic year 2018-2019.

Dr. Anup Biswas

Committee:

Dr. Anup Biswas

Dr. Anindya Goswami

This thesis is dedicated to Dr.Anup Biswas and Dr.Anindya Goswami.

Declaration

I hereby declare that the matter embodied in the report entitled Martingale Optimal Transport and Portfolio Theory are the results of the work carried out by me at the Department of Mathematics, Indian Institute of Science Education and Research, Pune, under the supervision of Dr. Anup Biswas and the same has not been submitted elsewhere for any other degree.

Vrushali Kumbhar

Acknowledgments

I feel blessed to have a mentor like Dr. Anup Biswas who has a contagious enthusiasm for everything in mathematics! When I struggled with my studies he was very patient in guiding me, I found much needed encouragement in his patience during crucial times. My initial projects with Dr Anindya Goswami were the building blocks of my learning process. I owe my strong mathematical foundations to Dr. Goswami. I am forever grateful to the brilliant minds of Dr. Anup Biswas and Dr. Anindya Goswami for their continual coaching throughout last three years.

IISER student community is a mix of very ambitious students and this created a very envigorating, competitive environment which kept me going even during my moments of discouragement. Thank you IISER for creating such a unique place where aspiring scientist can learn from the best faculties in the world.

Abstract

This project involved the study of Monge-Kantorovich problem of optimally transporting one distribution of mass to another. A cost is incurred while doing the transportation and the optimality is measured against this cost function. The properties of solutions when the solution to optimal transport exist is studied. An application to portfolio theory will be discussed, which amounts to finding a portfolio strategy, a strategy, which depends only on the current state of the market, which will give the investor a possibility of unbounded profit with probability 1. We study the dual version of Monge - Kantorovich problem for martingale measures which has a natural financial interpretation in terms of hedging options.

Contents

Abstract **xi**

1 Preliminaries **3**

1.1 Polish Space 3

1.2 Couplings 3

1.3 Push forward maps 5

1.4 Transport Plans 5

1.5 Tools from convex analysis 6

1.6 Tools for portfolio theory 6

1.7 Probability Theory 8

2 Optimal transport problem and Kantorovich Duality **11**

2.1 Monge- Kantorovich formulation of Transport Problem 11

2.2 Kantorovich Duality 12

2.3 Optimal Transportation theorem for quadratic cost function 15

2.4 Kantorovich- Rubinstein Distance 21

3 Fundamental Theorem of Optimal Transport **23**

3.1 Optimal transport on real line 23

3.2 Cyclic Monotonicity 26

3.3 Fundamental Theorem of optimal transport 27

4 Portfolio Theory **31**

4.1 **Introduction** 31

4.2 Simplex Model 36

4.3 From portfolio to optimal transport 37

5 Martingale Optimal Transport **43**

5.1 Martingale measures 43

5.2 Probability measures with given marginals 43

5.3 Duality: Martingale Optimal Transport in two dimensions. 46

Introduction

Foundation of Optimal mass transportation can be traced back to Gaspard Monge's famous paper of 1781:Memoire sur la theorie des d'eblais et des remblais. The problem here is to minimize the cost of transporting a given distribution of mass from one location to another both having the same volume. Since then it has become a classical subject in the field of probability theory, economics and optimization. Following the seminal discoveries of Brenier in his 1987 paper, optimal transportation has received a much renewed attention in the last 20 years. It has become an increasingly powerful tool at the interface between partial differential equations, geometry, fluid dynamics, probability theory and functional analysis. Optimal Transportation theory has a wide range of important applications. In the context of machine learning/signal processing, one often has to deal with collections of samples that can be interpreted as probability distributions and Optimal transport is a perfect tool to compare empirical probability distributions. It appears that Optimal Transport seems to be an approach with a long mathematical tradition and with a rich enough set of mathematical tools that may have an outsized impact on future deep learning theoretical work.

The standard approach in the pricing of options, the price of a contingent claim is calculated using a postulated model as a suitably discounted, risk neutral expectation of the payoff under that model. In practice we can observe the traded option prices, but know little about the underlying model. If we know single price we can calculate the volatility using the Black-Scholes model. But if we know the prices of multiple call options together, then they will typically be inconsistent with the Black Scholes model. On the other hand, if we know the vanilla option prices, for all strikes and maturities, then we can find a unique martingale diffusion consistent with all those prices. A rather unconventional way of finance based pricing has come up in the recent years, stream of model-free finance. The research attempts to investigate prices that are consistent in the presence of arbitrage given the current prices of traded options on the same underlying. The option bounds are obtained using methods based on Skorokhod embedding [3] and the theory of optimal transport [9]. The motivation of this project was to understand model free finance using optimal transportation theory.

The project started with the study of probability theory and then shifted focus on learning Optimal transportation theory [9] and eventually moved on to study of geometric ideas in probability theory. The interplay between portfolio theory and optimal transport was explored in [4]. As an application of optimal transport, in [5] we studied how only finite number of traded option prices as data input can be used to obtain model independent option bounds using the duality theory for martingale measures.

Chapter 1

Preliminaries

1.1 Polish Space

Polish spaces are complete separable metric spaces. More precisely, $(\mathcal{X}, d)$ is said to be a Polish space if $(\mathcal{X}, d)$ is complete metric space and it has a countable dense subset.

1.2 Couplings

- **Couplings**

Definition 1.2.1. *Let* $(\mathcal{X}, \mu)$ *and* $\mathcal{Y}, \nu)$ *be two probability spaces. A coupling measure* π *is a probability measure on* $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$ *such that* π *admits marginals* μ *and* ν *on* $\mathcal{X}$ *and* $\mathcal{Y}$*, respectively. Following ways to state this equivalent condition:*

1. $(Proj_{\mathcal{X}})_{\#}\pi = \mu$, $(Proj_{\mathcal{Y}})_{\#}\pi = \nu$; *where* $Proj_{\mathcal{X}}$ *and* $Proj_{\mathcal{Y}}$ *are the projection maps* $(x, y) \to x$ *and* $(x, y) \to y$*, respectively*
2. *For all measurable* $A \subset \mathcal{X}$, $B \subset \mathcal{Y}$ *we have,* $\pi[A \times \mathcal{Y}] = \mu[A]$ *and* $\pi[\mathcal{X} \times B] = \nu[B]$
3. *For all integrable measurbale functions* ϕ, ψ *on* $\mathcal{X}, \mathcal{Y}$,

$$\int_{\mathcal{X}\times\mathcal{Y}} (\phi(x) + \psi(y))d\pi(x, y) = \int_{\mathcal{X}} \phi(x)d\mu + \int_{\mathcal{Y}} \psi(y)d\nu.$$

The space of coupling measures of (μ, ν) is denoted by $\Pi(\mu.\nu)$.

Remark 1.2.1. *Couplings always exist, there is always at least a trivial coupling i.e; $\pi = \mu \times \nu$.*

- **Deterministic coupling**

Definition 1.2.2. *A coupling (X, Y) is said to be deterministic if there exists a measurable function $T : \mathcal{X} \to \mathcal{Y}$ such that $Y = T(X)$ and (X, Y) is a coipling of (μ, ν). To say that (X, Y) is a coupling of μ and ν it is equivalent to saying any of the following:*

1. *X has law μ and $Y = T(X)$, where $T_{\#}\mu = \nu$.*
2. *X has law μ and $Y = T(X)$, where T is a change of variables from μ to ν : for all $\nu-$ integrable function ϕ,*

$$\int_{\mathcal{Y}} \phi(y) d\nu(y) = \int_{\mathcal{X}} \phi(T(x)) d\mu(x);$$

3. *$\pi = (id, T)_{\#}\mu$*

- **Wasserstein Distance**

Definition 1.2.3. *Let $(\mathcal{X}, d)$ be a Polish metric space, and $1 \leq p < \infty$. For any two probability measures μ and ν on X, the Wasserstein distance between the probability measures of the order p is defined by*

$$W_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \Big\{ \int d(x, y)^p d\pi(x, y) \Big\}^{\frac{1}{p}}. \tag{1.1}$$

- **Wasserstein Space**

Definition 1.2.4. *The Wasserstein Space of order $p \geq 1$ is defined as*

$$P_p(\mathcal{X}) := \{\mu \in P(\mathcal{X}); \quad \int_{\mathcal{X}} d(x_0, x)^p \mu(dx) < +\infty\}, \tag{1.2}$$

where $x_0 \in \mathcal{X}$ is arbitrary. It is easy to see that the space is independent of the choice of x_0. W_p defines a distance on $P_p(\mathcal{X})$.

1.3 Push forward maps

For a Borel measurable map $T : \mathcal{X} \to \mathcal{Y}$, we define its corresponding push forward operator $T_\# : P(\mathcal{X}) \to P(\mathcal{Y})$. In particular, for $\mu \in P(\mathcal{X})$ we have

$$T_\# \mu(B) = \mu(T^{-1}(B)), \quad \text{for all Borel set } B \subset \mathcal{Y}.$$

For discrete measures, push forward operations consists of simply moving the positions of all the points in the support of the measures i.e. given $\alpha = \sum_i a_i \delta_{x_i}$ we get

$$T_\# \alpha := \sum_i a_i \delta_{T(x_i)}.$$

For more general measures, with density functions, the push forward measures is defined as follows. For $T : \mathcal{X} \to \mathcal{Y}$, the push-forward measure $\beta = T_\# \alpha \in P(\mathcal{Y})$ of some $\alpha \in P(\mathcal{X})$ satisfies

$$\forall h \in \mathcal{C}_b(\mathcal{Y}), \qquad \int_Y h(y) d\beta(y) = \int_X h(T(x)) d\alpha(x),$$

where $\mathcal{C}_b$ denotes the collection of all bounded continuous functions. Equivalently, for any measurable set $B \subset Y$, one has

$$\beta(B) = \alpha(\{x \in X : T(x) \in B\}) = \alpha(T^{-1}(B)).$$

Intuitively, the measurable map $T : \mathcal{X} \to \mathcal{Y}$ can be interpreted as a function moving a single point from a measurable space to another. $T_\#$ is an extension of T that can move the entire probability measure on $\mathcal{X}$ toward a new probability measure on $\mathcal{Y}$.

1.4 Transport Plans

Transport plan is a probability measure $\pi(\mu, \nu)$ which is in $\Pi(\mu, \nu)$, Where μ, ν are the probability measures we are interested in transporting.

1.5 Tools from convex analysis

1. (**Convex functions.**) Consider a function $f : I \to \mathbb{R}$, where I is an interval in $\mathbb{R}$. We say that the function f is a convex function if, for any points x and y in I and any $t \in [0, 1]$ we have,
$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

2. (**Jensen's inequality.**) If $f(x)$ is a convex function and X is a random variable then, we have,
$$f(E[X]) \leqslant E[f(X)].$$

3. (**c-concave functions.**) Given a function $\phi : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ and a lower semi-continuous cost function $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$, ψ is called a c-concave on $\mathcal{Y}$, if
$$\psi(y) = \inf_{x \in \mathcal{X}} (c(x, y) - \phi(x))$$
for $y \in \mathcal{Y}$. Let $\phi : \mathcal{X} \to \mathbb{R} \cup \{-\infty\}$ be a function not identical to $-\infty$, then
$$\phi^c(y) = \inf_{x \in \mathcal{X}} (c(x, y) - \phi(x))$$
is called a c-transform of ϕ, which is a c-concave on $\mathcal{Y}$. We have that ϕ is a c-concave function on $\mathcal{X}$, if and only if $\phi = \phi^{cc}$. The c-superdifferential of $\partial_c\phi \subset \mathcal{X} \times \mathcal{Y}$ of a c-concave function ϕ is defined as the set of all pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ such that for all $z \in \mathcal{X}$
$$\phi(z) \leqslant \phi(x) + (c(z, y) - c(x, y)).$$

4. (**C-monotone set.**) A set $\Gamma \subset \mathcal{X} \times \mathcal{Y}$ is a c-monotone if and only if there is a c-concave function ϕ such that $\Gamma \subset \partial_c\phi$.

1.6 Tools for portfolio theory

Definition 1.6.1. *(Bregman Divergence) Given a convex function $\Psi(x)$ defined on a convex set Ω, we have*
$$\Psi(u) + \nabla\Psi(u) \cdot (u - v) \geqslant \Psi(v), \qquad (u, v) \in \Omega. \tag{1.3}$$

The difference assigns a non negative number (positive number if Ψ is strictly convex) to the pair (u, v). Bregman divergence is like a metric but it satisfies neither the symmetry nor the triangle inequality to be as a certified metric. The Bregman diverence associated with the function $\Psi(x)$ defined on the convex set Ω is given by:

$$\mathrm{BD}(u, v) = \Psi(u) \; - \; \Psi(v) \; - \; \langle \nabla\Psi(v), u - v \rangle; \qquad u, v \in \Omega. \tag{1.4}$$

Kulback Leibler Divergence

When log function is used in Bregman divergence we get the kullback Leibler divergence. It is also used as a notion for "distance" between two probability distributions. The Kullback Leibler divergence between two probability density f and g is given by:

$$\mathrm{KL}(f||g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx \tag{1.5}$$

The relative entropy is also known as Kullback Leibler divergence

Definition 1.6.2. *(Exponentially concave functions). A function ϕ is called an exponentially concave function if e^{ϕ} is a concave function.*

Definition 1.6.3. *(l- divergence). A $l-$ divergence of an exponentially concave function ψ defined on a convex set ω is given as follows:*

$$T(u|v) = \log(1 + \psi(v) \cdot (u - v)) - (\psi(u) - \psi(v)) \qquad \ldots u, v \in \omega \tag{1.6}$$

Definition 1.6.4. *(Unit Simplex(Δ_n)). For $n \geq 2$ we define the open unit simplex Δ_n in $\mathbb{R}^n$ by*

$$\Delta_n := \{\mathbf{p} = (p_1, \ldots, p_n) \in (0, 1)^n : \; p_1 + \cdots + p_n = 1\} \tag{1.7}$$

The closure of Δ_n is denoted by $\overline{\Delta_n}$

Definition 1.6.5. *(Relative Entropy).The relative entropy is the function $H(\cdot \mid \cdot)$ on $\overline{\Delta_n} \times \overline{\Delta_n}$ defined by*

$$H(p \mid q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}, \qquad p, q \in \overline{\Delta_n} \tag{1.8}$$

Definition 1.6.6. *(Excess growth rate). Let $\pi \in \overline{\Delta_n}$ be fixed. We define the functional*

$T_\pi(\cdot|\cdot)$ *on* $\Delta \times \Delta$ *by*

$$T_\pi(q|p) = \log\left(\sum_{i=1}^{n} \pi_i \frac{q_i}{p_i}\right) - \sum_{i=1}^{n} \pi_i \log \frac{q_i}{p_i}, \qquad p, q \in \Delta_n \tag{1.9}$$

1.7 Probability Theory

1.7.1 Conditional Expectation with respect to a sub σ-algebra

On a general probability space $(\Omega, \mathcal{F}, \mathcal{P})$, the conditional probability of an event $B \in \mathcal{F}$ occurring given that an event $A \in \mathcal{F}$ has already occurred is defined by $P(B \mid A) = \frac{P(A \cap B)}{P(A)}$ as long as $P(A) \neq 0$. For a fixed A, $P(\cdot \mid A)$ is the probability measure on $(\Omega, \mathcal{F})$ and that $P(A \mid A) = 1$. Following this, if X is an random variable, the conditional expectation of X given A is given by:

$$\int_\Omega X dP(\cdot \mid A) = E[X \mid A].$$

Generalizing, to allow conditioning on an arbitrary $\sigma-$ algebra $\mathcal{G} \subset \mathcal{F}$. Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space. Suppose X is a real valued integrable random variable defined on the probability space, then the conditional expectation of X given $\mathcal{G}$ is a $\mathcal{G}$- measurable, integrable random variable Y such that, for all $A \in \mathcal{G}$

$$E[I_A X] = \int_A X dP = \int_A Y dP = E[I_A Y].$$

Y is denoted by $E[X|\mathcal{G}]$.

1.7.2 Properties of conditional expectation

1. For $\mathcal{G} \subseteq \mathcal{F}$, the conditional expectation $E[X \mid \mathcal{G}]$ exists for all $X \in \mathcal{L}^1(\Omega)$, and is unique up to a set of probability zero.

2. (**Linearity.**) Suppose X and Y are integrable random variables, and $\alpha, \beta, \gamma \in \mathbb{R}$ are constants. Then

$$E[\alpha A + \beta Y + \gamma|\mathcal{G}] = \alpha E[X|\mathcal{G}] + \beta E[Y|\mathcal{G}] + \gamma, \qquad a.s.$$

3. Suppose X and Y are integrable random variables and $X \leqslant Y$ a.s. then,

$$E[X|\mathcal{G}] \leqslant E[Y|\mathcal{G}] \,.$$

4. (**Dominated/Monotone Convergence.**) Suppose $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of integrable random variables which converge almost surely to an integrable random variable X. If either there exists a non-negative random variable Y with $|X_n| \leqslant Y$, or X_n is a monotone increasing, or X_n is a monotone decreasing, then

$$E[X|\mathcal{G}] = \lim_n E[X_n|\mathcal{G}] \qquad a.s.$$

5. (**Tower Property.**) Suppose $\mathcal{G}_1$ and $\mathcal{G}_2$ are sub $\sigma-$ algebras of $\mathcal{F}$ such that $\mathcal{G}_2 \subseteq \mathcal{G}_1 \subseteq \mathcal{F}$. Then, for any integrable random variable X,

$$E\left[E[X|\mathcal{G}_1|\mathcal{G}_2]\right] = E[X|\mathcal{G}_2] \qquad a.s.$$

6. (**Taking out the known.**) Suppose X is an integrable random variable and Y an $\mathcal{G}-$measurable random variable, such that the product XY is integrable. Then $E[XY|\mathcal{G}] = YE[X|\mathcal{G}]$ a.s.

1.7.3 Martingales

Definition 1.7.1. *A real valued stochastic process $\{X_n\}$ is called a super-martingale with respect to the $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$ where $\mathbb{T} = \{0, 1, 2, 3, ...\}$, if,*

1. *each X_n is $\mathcal{F}_n$ measurable, i.e. $\{X_n\}_{n\in\mathbb{T}}$ is adapted to $\{\mathcal{F}_n\}_{n\in\mathbb{T}}$,*
2. *$E[|X_n|] < \infty$, for all $n \in \mathbb{T}$, and*
3. *$X_n \geqslant E[X_m|\mathcal{F}_n]$ almost surely, for all $m \geqslant n$,*

When we replace "$\geqslant$" in property (c) with "$\leqslant$", then X is called a submartingale. If the sequence X is both a submartingale and a supermartingale, then it is called a martingale.

Chapter 2

Optimal transport problem and Kantorovich Duality

Like many other areas in mathematics the field of optimal transport was born several times. In the eighteenth century, the geometer Gaspard Monge was the progenitor to formalize the mass transport problem; and long later it was rediscovered by Leonoid Kantorovich. The formalization of the transport problems are as follows:

2.1 Monge- Kantorovich formulation of Transport Problem

- **Monge Formulation of optimal transport**
  Consider you have to extract a certain amount of sand from the ground and this sand is to be transported to different places wherein it is to be incorporated in a construction. The places where the sand is to be extracted from and the places where it is to be transported to are known a priori. The assignment of certain extracted mass of soil from the place of extraction to the place of construction is to be determined. The assignment problem matters because the transportation of mass of sand from the place of extraction to the place of construction is costly. Monge assumed that the cost of transportation of a unit mass was given by the product of mass by a certain distance. Also, the sand extracted from one site can be supplied to only one construction site.

There is no splitting of sand dug from one site. This makes the assignment a bijection map between the extraction and construction sites. Let the sites of extraction be modelled by μ and the sites of construction be modelled by ν. Let $c(x, y)$ be the cost function such that the cost of transportation for a particle of sand extracted from x to the site of construction y. The Monge's problem can be written as follows:

$$Minimize \qquad I[T] = \int_{\mathcal{X}} c(x, T(x)) d\mu(x) \tag{2.1}$$

over the set of all measurable maps T such that $T_{\#}\,\mu = \nu$

- **Kantorovich formulation of optimal transport**
  Long later Monge's problem was rediscovered by a Russian mathematicain Leonoid Kantorovich. Knatorovich made some changes to the restrictions of Monge's formulation. He stated that sand dug from one site can be supplied to more than one construcion sites and similarly, one construction site can import sand from multiple sites of extraction. Instead of optimizing on transport maps, Kantorovich proposed to optimize over couplings of measures. Mathematically, it can written as:

$$Minimize \qquad I[\pi] = \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\, d\pi(x, y) \qquad for\ \pi\ \in \Pi(\mu, \nu) \tag{2.2}$$

  Kantorovich stated and proved, by means of functional analytical tools, a duality theorem which plays a crucial role in the problem of optimal coupling. He also devised a convenient notion of distance between probability measures should be the optimal transport cost from one measure to the other if the cost is chosen as distance function. This distance is now-a-days called the Kantorovich - Rubinstein distance. Its was after many years that Kantorovich made connections to Monge's work and now the problem of optimal coupling is jointly called Monge-Kantorovich problem

2.2 Kantorovich Duality

Theorem 2.2.1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Polish spaces, let $\mu \in P(\mathcal{X})$ and $\nu \in P(\mathcal{Y})$ and let $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ be a lower semi-continuous cost function. Whenever $\mu \in P(\mathcal{X})$ and*

$\nu \in P(\mathcal{Y})$ and $(\phi, \psi) \in \mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$, define

$$I[\pi] = \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)\, d\pi(x,y), \qquad J(\phi,\psi) = \int_{\mathcal{X}} \phi\, d\mu + \int_{\mathcal{Y}} \psi\, d\nu. \tag{2.3}$$

Define $\Pi(\mu, \nu)$ to be the set of all Borel probability measures π on $\mathcal{X} \times \mathcal{Y}$ such that for all measurable sets $A \subset X$ and $B \subset Y$,

$$\pi[A \times \mathcal{Y}] = \mu[A], \qquad \pi[\mathcal{X} \times B] = \nu[B],$$

and define Φ_c to be the set of all measurable functions $(\phi.\psi) \in \mathcal{L}^1 d(\mu) \times \mathcal{L}^1(d\nu)$ satisfying

$$\phi(x) \,+\, \psi(y) \leqslant c(x,y)$$

for $d\mu$-almost all $x \in X$, $d\nu$- almost all $y \in Y$.
Then,

$$\inf_{\Pi(\mu,\nu)} I[\pi] = \sup_{\Phi_c} J(\phi,\psi) \tag{2.4}$$

Sketch of proof: The idea is to rewrite the constrained infimum problem as an infsup problem and then apply minimax principle to exchange the two i.e, replacing a "infsup" by a "supinf".
We begin by writing the indicator function of Π as follows:

$$\inf_{\pi\in\Pi(\mu,\nu)} I[\pi] = \inf_{M_+(X\times\mathcal{Y})} \left(I[\pi] + \left\{ \begin{array}{c} 0 \text{ if } \pi \in \Pi(\mu,\nu) \\ +\infty \text{ else} \end{array} \right. \right), \tag{2.5}$$

Where $M_+(\mathcal{X} \times \mathcal{Y})$ denotes the space of non-negative Borel measures on $\mathcal{X} \times \mathcal{Y}$. Now,

$$\left\{ \begin{array}{c} 0 \text{ if } \pi \in \Pi(\mu,\nu) \\ +\infty \text{ else} \end{array} \right\} = \sup_{(\phi,\psi)} \left[\int_{\mathcal{X}} \phi(x) d\mu + \int_{\mathcal{Y}} \psi(y) d\nu - \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x) + \psi(y)] d\pi(x,y) \right] \tag{2.6}$$

where the supremum runs over all $(\phi, \psi) \in \mathcal{C}_b \times \mathcal{C}_b$. Using this, we write the left hand side of the equation (2.4) as

$$= \inf_{\pi\in M_+(\mathcal{X}\times\mathcal{Y})} \sup_{(\phi,\psi)} \left\{ \int_{\mathcal{X}\times Y} c(x,y) d\pi + \int_{\mathcal{X}} \phi(x) d\mu + \int_{\mathcal{Y}} \psi(y) d\nu - \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x) + \psi(y)] d\pi(x,y) \right\} \tag{2.7}$$

Taking for granted that we can apply the minimax principle.

$$= \sup_{\phi,\psi} \inf_{\pi \in M_+(\mathcal{X}\times\mathcal{Y})} \left\{ \int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi + \int_{\mathcal{X}} \phi(x)d\mu + \int_{\mathcal{Y}} \psi(y)d\nu - \int_{\mathcal{X}\times\mathcal{Y}} [\phi(x)+\psi(y)]d\pi(x,y) \right\} \tag{2.8}$$

Taking the infimum inside the brackets and rearranging the terms,

$$= \sup_{(\phi,\psi)} \left\{ \int_{\mathcal{X}} \phi(x)d\mu + \int_{\mathcal{Y}} \psi(y)d\nu - \sup_{\pi \in M_+(\mathcal{X}\times\mathcal{Y})} \int_{\mathcal{X}\times Y} [\phi(x)+\psi(y)-c(x,y)]d\pi(x,y) \right\} \tag{2.9}$$

Let us compute the supremum inside the curly brackets. The function $\zeta(x,y) = \phi(x) + \psi(y) - c(x,y)$ takes a positive value at some point (x_0, y_0). Then by choosing, $\pi = \lambda\delta_{(x_0,y_0)}$ and letting $\lambda \to \infty$ we see the supremum is infinite. On the other hand, if ζ is non-positive, ($d\mu \otimes d\nu$ -everywhere), the supremum is clearly obtained for $\pi = 0$. Thus,

$$\sup_{\pi \in M_+(\mathcal{X}\times\mathcal{Y})} \int_{\mathcal{X}\times Y} [\phi(x)+\psi(y)-c(x,y)]d\pi(x,y) = \left\{ \begin{array}{c} 0 \text{ if } (\phi,\psi) \in \Phi_c \\ +\infty \text{ else} \end{array} \right\} \tag{2.10}$$

Plugging this into equation (2.9), we obtain,

$$(2.9) = \sup_{(\phi,\psi)\in\Phi_c} J(\phi,\psi)$$

as desired. $\square$

Remark 2.2.1. *It follows that when c is bounded, one can restrict the supremum of the right hand side of 2.4 to those pairs (ϕ^{cc}, ϕ^c) where ϕ is bounded and,*

$$\phi^c(y) = inf_{x\in\mathcal{X}}[c(x,y) - \phi(x)] \quad \phi^{cc}(x) = \inf_{y\in\mathcal{Y}}[c(x,y) - \phi^c(y)] \tag{2.11}$$

It is easy to show that $(\phi^{cc})^c = \phi^c$. The pair (ϕ^{cc}, ϕ^c) is called the pair of conjugate c-concave functions. ϕ^c is measurable as it can be written as $\lim_{l\to\infty} \psi_l$ where ψ_l

$$\phi_l(y) = \inf_{x\in\mathcal{X}}[c(x,y) - \phi(x)]$$

where c_l is an increasing family of uniformly bounded functions converging pointwise to c

2.3 Optimal Transportation theorem for quadratic cost function

Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$ and let the cost function be the square of the Euclidean norm. We take the cost function to be $c(x, y) = \frac{|x-y|^2}{2}$. So the cost function will be:

$$I[\pi] = \int_{\mathbb{R}^n \times \mathbb{R}^n} \frac{|x-y|^2}{2} d\pi(x, y) \tag{2.12}$$

Let μ and ν be probability measures on $\mathbb{R}^n$, both having finite second order moments, in the following sense

$$\mathcal{M} = \int_{\mathbb{R}^\ltimes} \frac{|x|^2}{2} \, d\mu(x) + \int_{\mathbb{R}^\ltimes} \frac{|y|^2}{2} \, d\nu(y) < \infty. \tag{2.13}$$

This condition ensures that $I[\pi]$ is always finite on $\Pi(\mu, \nu)$. Next, we formulate the dual problem with quadratic cost. The condition for (ϕ, ψ) to belong to Φ_c is:

$$\phi(x) + \psi(y) \leq \frac{|x-y|^2}{2} \tag{2.14}$$

holding for $\underset{\cdot}{\mu}-$ almost all x and $\underset{\cdot}{\nu}-$ almost all y in $\mathbb{R}^n$. We expand the R.H.S of this equation and rearranging the terms we have,

$$x \cdot y \leq \left(\frac{|x|^2}{2} - \phi(x) \right) + \left(\frac{|y|^2}{2} - \psi(y) \right) \tag{2.15}$$

We consider the new unknown functions

$$\tilde{\phi}(x) = \frac{|x|^2}{2} - \phi(x), \qquad \tilde{\psi}(y) = \frac{|y|^2}{2} - \psi(y) \tag{2.16}$$

We can then express the dual problem in the following manner.

$$\inf_{\Pi(\mu,\nu)} I[\pi] = M_2 - \sup\{ \int (x \cdot y) d\pi(x, y); \qquad \pi \in \Pi(\mu, \nu) \} \tag{2.17}$$

and

$$\sup_{\Phi_c} J = M_2 - \inf\{ J(\phi, \psi); \quad (\tilde{\phi}, \tilde{\psi}) \in \tilde{\Phi} \} \tag{2.18}$$

where $\tilde{\Phi}$ is the set of all pairs $\tilde{\phi}, \tilde{\psi}$ in $\mathcal{L}^1(d\mu) \times \mathcal{L}^1(d\nu)$ such that for almost all x, y

$$x \cdot y \leq \tilde{\phi}(x) + \tilde{\psi}(y) \tag{2.19}$$

Then equation 2.4 becomes,

$$\sup\left\{\int (x \cdot y) d\pi(x,y);\ \ \pi \in \Pi(\mu,\nu)\right\} = \inf\left\{J(\phi,\psi);\ \ (\phi,\psi) \in \tilde{\Phi}\right\} \tag{2.20}$$

Here we state without proving an important theorem which uses convex conjugate function in minimizing for the quadratic cost function.

Theorem 2.3.1. *Let (μ,ν) be probability measures on $\mathbb{R}^n$, with finite second order moments. Let $\tilde{\phi}_c$ be defined by equation (2.18). Then there exists a pair of lower semi-continuous functions (ϕ,ϕ^c) on $\mathbb{R}^n$, such that,*

$$\inf_{\tilde{\Phi}_c} J = J(\phi,\phi^c)$$

Lemma 1. *The minimization problem* $\inf\{I[\pi];\, \pi \in (\mu,\nu)\}$ *admits a minimizer.*

Proof. First of all, $\Pi(\mu,\nu)$ is a non-empty set. $\Pi(\mu,\nu)$ is compact for weak topology of probability measures (topology induced by $\mathcal{C}_b(\mathbb{R}^n \times \mathbb{R}^n)$. To show the tightness of the set $\Pi(\mu,\nu)$.(The tightness will imply relative compactness with respect to weak topology). Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$, let $\delta > 0$, given a set $K \subset \mathcal{X}$, $L \subset \mathcal{Y}$ such that,

$$\mu[\mathcal{X} \setminus K] \leq \delta \qquad \nu[\mathcal{Y} \setminus L] \leq \delta.$$

Whenever $\pi \in \Pi(\mu,\nu)$,

$$\pi[(X \times Y) \setminus (K \times L)] \leq \pi[X \times (Y \setminus L)] + \pi[(X \setminus K) \times L] \leq 2\delta$$

Since conditions which define $\Pi(\mu,\nu)$ are continuous with respect to weak topology, $\Pi(\mu,\nu)$ is weakly closed and hence compact. This implies that there exists a minimizer for I. Let $(\pi_k)_{k\in\mathbb{N}}$ be a minimizing sequence; then it admits a cluster point $\pi_* \in \Pi(\mu,\nu)$. Given a cost function $c(x,y)$, it can be expressed as supremum of a sequence of bounded, non-decreasing functions $(c_l)_{l\in\mathbb{N}}$. By successively invoking monotone convergence theorem, the fact that π_* is a cluster point, the inequality $c_l \leq c$ and minimizing property of π_* we have;

$$
\begin{aligned}
\int c(x,y)d\pi_*(x,y) &= \lim_{l\to\infty}\int c_l(x,y)d\pi_*(x,y) \\
&\leq \lim_{l\to\infty}\limsup_{k\to\infty}\int c_l(x,y)d\pi_k(x,y) \\
&\leq \limsup_{k\to\infty}\int c(x,y)d\pi_k(x,y) = \inf I.
\end{aligned} \tag{2.21}
$$

□

- **Knott Smith Criterion.** It states that $\pi \in \Pi(\mu,\nu)$ is the optimal coupling if and only if there exists a convex lower semi-contimnuous function ϕ such that

$$
Supp(\pi) \subset Graph(\partial\phi) \tag{2.22}
$$

or equivalently,

$$
for\, d\pi -\ almost\, all\, (x,y), \qquad y \in \partial\phi(x) \tag{2.23}
$$

Moreover, in that case, the pair (ϕ,ϕ^*) has to be the minimizer of the problem

$$
\inf\left\{\int_{\mathbb{R}^n}\phi d\mu + \int_{\mathbb{R}^n}\psi d\nu; \qquad \forall(x,y), \quad x\cdot y \leqslant \phi(x) + \psi(y)\right\}
$$

Where

$$
\partial\phi
$$

is the sub-differential of the convex function ϕ defined by,

$$
\partial\phi(x) := \{y\, :\, \phi(z) \geq \phi(x) + y\cdot(z-x), \forall z \in \mathbb{R}^n\} \tag{2.24}
$$

A sub-differential is in general a set. However if ϕ is differentiable at x then $\partial\phi(x) = \nabla\phi(x)$.

2.3.1 Breiner's theorem

It turns out that Monge's cost $c(x,y) = |x-y|$ is much harder to deal with than the quadratic cost. The situation for quadratic cost is much simpler as it mirrors the Hilbert space geometry of $\mathbb{L}^2$ among the Banach spaces $\mathbb{L}^p$ with $p \geq 1$. Brenier proved that there is a unique solution to the quadratic cost and characterized it as a convex gradient.

Theorem 2.3.2. *If μ does not give mass to small sets then there is unique optimal π, which is*

$$d\pi(x, y) \;=\; d\mu(x)\, \delta[y \;=\; \nabla\phi(x)] \tag{2.25}$$

or equivalently

$$\pi = (Id \,\times\, \nabla\phi)\#\mu \tag{2.26}$$

where $d\phi$ is the gradient of a convex function which pushes μ forward to ν: $\nabla\phi\#\mu \;=\; \nu$. Moreover,

$$Supp(\nu) = \overline{\nabla\phi(supp(\mu)}.$$

As a corollary, $\nabla\phi$ is the unique solution to Monge transportation problem:

$$\int_{\mathbb{R}^n} |x - \nabla\phi(x)|^2 d\mu(x) = \inf_{T\#\mu=\nu} \int_{\mathbb{R}^n} |x - T(x)|^2 d\mu(x),$$

or equivalently,

$$\int_{\mathbb{R}^n} x \cdot \nabla\phi(X) d\mu(x) = \sup_{T\#\mu=\nu} \int_{\mathbb{R}^n} x \cdot T(x) d\mu(x).$$

Proof. We only provide the key steps involved in the proof. For more detail we cite [9].

1. By Lemma 1, there exists a minimizer. By Theorem 2.3.1, there exists a pair of convex conjugate functions (ϕ, ϕ^c) optimal for the dual problem. Then writing the Kantorovich duality for quadratic cost function in the form of equation 2.20 we have

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} (x \cdot y) d\pi(x, y) \;=\; \int_{\mathbb{R}^n} \phi(x) d\mu + \int_{\mathbb{R}^n} \phi^c(y) d\nu$$

$$= \int_{\mathbb{R}^n\times\mathbb{R}^n} [\phi(x) + \phi^c(y)] d\pi(x, y)$$

Equivalently we have,

$$\int_{\mathbb{R}^n\times\mathbb{R}^n} [\phi(x) + \phi^c(y) - x \cdot y] d\pi(x, y) = 0$$

The functions inside the square bracket is non-negative. So, it has to vanish for $d\mu-$

almost all (x, y). Given a proper lower semicontinuous convex function ϕ on $\mathbb{R}^n$, for all $(x, y) \in \mathbb{R}^n$ we have the following

$$x \cdot y = \phi(x) + \phi^c(y) \iff y \in \partial\phi(x) \iff x \in \partial\phi^c(y) \tag{2.27}$$

Using this we can say that

$$\text{for } d\pi - \text{almost all } (x, y),\ y \in \partial\phi(x) \tag{2.28}$$

2. Conversely let $\pi \in \Pi(\mu, \nu)$. Then by the same argument,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y) d\pi(x, y) = \int_{\mathbb{R}^n} \phi(x) d\mu + \int_{\mathbb{R}^n} \phi^c d\nu \tag{2.29}$$

So, π and (ϕ, ϕ^c) are optimal on both sides. This concludes the proof of Knott-Smith criterion.

3. Assume that μ does not give mass to small sets, and let ϕ be as above. $\mu \in \mathcal{L}^1(d\mu)$, it is $d\mu-$ almost everywhere finite: μ[Dom(ϕ)]=1. The border of the convex set ∂Dom(ϕ) of a convex set Dom(ϕ) is a small set. Therefore, μ[Int(Dom(ϕ)]=1. Now on Int(Dom(ϕ)), the set of non-differentiability of ϕ is a small set. On the whole, $d\mu-$ almost every point of X is a differentiability point for ϕ. So, for $d\mu-$ almost all x, the subdifferential of ϕ at a point x is $\{\nabla\phi(x)\}$. And, the statement true for $d\mu-$ almost all x is true for $d\pi-$ almost all (x, y), we obtain that $y = \nabla\phi(x)$ for $d\pi$ almost all(x, y).

   We have so far proved that any optimal transference plan takes from $(Id \times \nabla\phi)\#\mu$, for some convex ϕ such that $\nabla\phi\#\mu = \nu$, and that there is atleast one such transference plan.

4. **Uniqueness.** Let $\overline{\phi}$ be another convex function such that $\nabla\overline{\phi}_{\#}\mu = \nu$, and accordingly, $(\overline{\phi}, \overline{\phi^c})$ is an optimal pair of dual problem. Therefore, we have,

$$\int_{\mathbb{R}^n} \overline{\phi} d\mu + \int_{\mathbb{R}^n} \overline{\phi^c} d\nu = \int_{\mathbb{R}^n} \phi d\mu + \int_{\mathbb{R}^n} \phi^c d\nu. \tag{2.30}$$

   Let π be the optimal transference plan associated with ϕ. We have,

$$\int_{\mathbb{R}^n \times \mathbb{R}^n} [\overline{\phi}(x) + \overline{\phi^c}(y)] d\pi(x, y) = \int_{\mathbb{R}^n \times \mathbb{R}^n} [\phi(x) + \phi^c(y)] d\pi(x, y)$$

$$= \int_{\mathbb{R}^n \times \mathbb{R}^n} (x \cdot y) d\pi(x, y).$$

Since, $\pi = (Id \times \nabla\phi)_{\#}\mu$, the above can be written as,

$$\int_{\mathbb{R}^n} [\overline{\phi}(x) + \overline{\phi^c}(\nabla\phi(x))] d\mu(x) = \int_{\mathbb{R}^n} x \cdot \nabla\phi(x) d\mu(x).$$

Thus,

$$\int_{\mathbb{R}^n} [\overline{\phi}(x) + \overline{\phi^c}(\nabla\phi(x)) - x \cdot \nabla\phi(x)] d\mu(x) = 0.$$

Since the integrand is non-negative, it has to vanish $d\mu-$ almost everywhere. Using 2.27 we can conclude that,

$$\nabla\phi(x) \in \partial\overline{\phi}(x)$$

for $d\mu-$ almost every x. Since $\overline{\phi}$ is differentiable $\mu-$ almost everywhere, we have,

$$\nabla\phi(x) = \nabla\overline{\phi}(x)$$

We have shown the uniqueness of the solution of Monge-Kantorovich problem and also the uniqueness of of gradient of a convex function $\nabla\phi$ such that $\nabla\phi_{\#}\mu = \nu$

5. To prove that Supp(ν)=$\overline{\nabla\phi}$(Supp(μ)). Let $x \in$ Supp(μ) be a differential point of ϕ, let $y = \nabla\phi(x)$. We know that for any $\epsilon > 0$ with $\nabla\phi(B_\delta(x)) \subset B_\delta(y)$, and in particular,

$$\nu[B_\epsilon(y)] \geq \mu[\nabla\phi^{-1}(\nabla\phi(B_\delta(x)))] \geq \mu[B_\delta(x)]$$

But $\mu[B_\delta(x)] > 0$, for all x lies in support of μ; therefore, $\nu[B_\epsilon(y)] > 0$ too. Since, $\epsilon$is arbitrary we deduce that $y \in$ Supp(ν) We conclude that,

$$\nabla(Supp(\mu)) \subset Supp(\nu) \tag{2.31}$$

6. On the other hand, $\nu[\nabla\phi Supp(\mu))] \geq \mu[Supp(\mu)] = 1$. So, ν is concentrated on $\nabla\phi$(Supp μ). Therefore, by definition of support,

$$Supp(\nu) \subset \overline{\nabla\phi(Supp(\mu))} \tag{2.32}$$

Combining equations 2.31 and 2.32 we get

$$Supp(\nu) = \overline{\nabla\phi(Supp(\mu))}$$

2.4 Kantorovich- Rubinstein Distance

Kantorovich and Rubinstein made the duality more explicit for the case when the cost function is the distance function, i.e; $c(x, y) = d(x, y)$. The Wasserstein distance $\mathcal{W}_1$, distance is also called the Kantorovich Rubinstein distance.

Theorem 2.4.1. *Let $X = Y$ be Polish space and let d be a lower sem-continuous metric on X. Let $\mathcal{V}_d$ be the cost of optimal transportation of cost $c(x, y) = d(x, y)$, i.e;*

$$\mathcal{V}_d = \inf_{\pi\in\Pi(\mu,\nu)} \int_{X\times X} d(x, y)d\pi(x, y).$$

Let Lip(X) denote the space of Lipschitz functions on X and,

$$|| \phi || \equiv \sup_{x\neq y} \frac{| \phi(x) - \phi(y) |}{d(x, y)}$$

Then,

$$\mathcal{V}_d = \sup\left\{ \int_X \phi(\mu - \nu); \qquad \phi \in \cap\mathcal{L}^1(d \mid \mu - \nu \mid); \quad || \phi ||_{Lip} \leqslant 1\right\}$$

Proof. Define $d_n = \frac{d}{1+n^{-1}d}$, this sequence satisfies $d_n \leq d$ and for all x, y the quantity $d_n(x, y)$ converges monotonically to $d(x, y)$ as $n \to \infty$. (The set of all 1-Lipschitz functions for d_n is included in the set for 1-Lipschitz functions for d. We are going to assume that all Lipschitz functions are bounded and hence integrable with respect to μ, ν. So, in the view of 2.3.1 we only need to check that

$$\sup_{(\phi,\psi)\in\Phi_d} J(\phi, \psi) = \sup\left\{ \int_X \phi d(\mu - \nu); \qquad ||\phi||_{Lip} \leq 1\right\} \tag{2.33}$$

From Remark 2.2.1 we know that,

$$\sup_{(\phi,\psi)\in\Phi_d} J(\phi,\psi) = \sup_{\phi\in\mathcal{L}^1(d\mu)} J(\phi^{dd},\phi^d),$$

where

$$\phi^d(y) = \inf_{x\in X}[d(x,y)-\phi(x)], \qquad \phi^{dd} = \inf_{y\in Y}[d(x,y)-\phi^d(y)]$$

ϕ^d, being an infimum of 1-Lipschitz functions, bounded from below at some point x_0, is Lipschitz. So,

$$-\phi^d(x) \le \inf_y [d(x,y)-\phi^d(y)] \le -\phi^d$$

where the right inequality follows from the the choice $x=y$ in the infimum and the left inequality by the 1-Lipschitz property. This means that $\phi^{dd} = -\phi^d$ and

$$\begin{aligned}\sup_{\Phi_c} J(\phi,\psi) \le \sup_{\phi\in\mathcal{L}^1(d\mu)} J(\phi^{dd},\phi^d) &= \sup_{\phi\in\mathcal{L}^1 d\mu} J(-\phi^d,\phi^d) \\ &\le \sup_{||\phi||_{Lip}\le 1} J(\phi,-\phi) \le \sup_{\Phi_c} J(\phi,\psi)\end{aligned} \tag{2.34}$$

So there is equality everywhere and result follows. □

Chapter 3

Fundamental Theorem of Optimal Transport

3.1 Optimal transport on real line

Let us consider two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with cumulative distribution functions F and G respectively. The cumulative function for μ has the following form:

$$F(x) = \int_{-\infty}^{x} d\mu = \mu((-\infty, x]).$$

Here, F is right continuous, non-decreasing, $F(+\infty) = 1$ and $F(-\infty) = 0$. Also, the inverse of F on [0,1] is,

$$F^{-1}(t) = \inf \big\{ x \in \mathbb{R} : F(x) > t \big\}.$$

In general, $F^{-1}(F(x)) \geq x$ and $F(F^{-1}(t)) \geq t$. If F is invertible then $F^{-1}(F(x)) = x$ and $F(F^{-1}(t)) = t$.

For probability measures on product space $\mathbb{R} \times \mathbb{R}$, the cumulative distribution can be represented by joint two dimensional cumulative distribution functions:

$$H(x_0, y_0) = \int_{R(x_0, y_0)} d\pi = \pi[R(x_0, y_0)],$$

where $R(x_0, y_0)$ is the rectangle made of all points $(x, y) \in \mathbb{R}^2$ with $x \leqslant x_0, y \leqslant y_0$. The function H on $\mathbb{R}^2$ which is non-decreasing, right continuous in both the variables x and y, and has limits 0 and 1 at $(-\infty, -\infty)$ and $(+\infty, +\infty)$ gives rise to a unique probability measure on $\mathbb{R}^2$. To see this, note that H determines the mass of every rectangles with sides parallel to the co-ordinate axes and that such rectangles generate all Borel Sets in $\mathbb{R}^2$.

We now state the solution of optimal transport on real line in terms of the cumulative distribution functions.

Theorem 3.1.1. *Let μ, ν be two probability measures on $\mathbb{R}$ with respective cumulative functions F and G. Let π be the probability measure on $\mathbb{R}^2$ with joint cumulative distribution function*

$$H(x, y) = min(F(x), G(y)). \tag{3.1}$$

Then, π belongs to $\Pi(\mu, \nu)$, and is optimal in Kantorovich transportation problem between μ and ν for the quadratic cost function $c(x, y) = |x - y|^2$. Moreover the value of optimal transportation cost is

$$\mathcal{T}(\mu, \nu) = \int_0^1 |F^{-1}(t) - G^{-1}(t)|^2 dt \tag{3.2}$$

Proof. For a cumulative function F, we shall denote by $F(x^-)$, the left limit $\lim_{z\uparrow x} F(z)$, which always exists because of monotonicity. There is no need to introduce the right limits since it is always right continuous.

(a) *We claim:*

$$Supp(\pi) \subset \{(x, y) \in \mathbb{R}^2; \quad F(x^-) \leqslant G(y) \text{ and } G(y^-) \leqslant F(x)\} \tag{3.3}$$

To prove this, assume that for instance $F(x^-) > G(y)$. From the fact that both F and G are right continuous functions and the right continuity of G, we can say that if x' belongs to a small neighbourhood of x and y' belongs to the small neighbourhood of y, then $F(x') > G(y')$. So,

$$H(x', y') = min[F(x'), G(y')] = G(y')$$

Thus, on a small rectangle centered at (x, y), the function H does not depend on the first variable x'. This easily entails that $d\pi = dH$ assigns zero mass to this rectangle, so $(x, y) \notin Supp(\pi)$ It does not follow that $G^{-1} \circ F(x^-) \leqslant G^{-1} \circ F(x)$ But one can check that monotonicity condition holds for the support of π.

(*Monotonicity:*) Here by monotonicity we mean that, if Γ a subset of $\mathbb{R}^2$ is said to be monotonous if

$$(x_1, y_1), (x_2, y_2) \in \Gamma \implies [x_1 \leqslant x_2 \text{ and } y_1 \leqslant y_2] or [x_1 \geqslant x_2 \text{ and } y_1 \geqslant y_2] \quad (3.4)$$

On $\mathbb{R}$ gradients of a convex function coincide with non-decreasing functions. And subgradients are complete nondecreasing graphs, or maximal monotone subsets of $\mathbb{R}^2$.

Indeed, let $(x_1, y_1), (x_2, y_2)$ be two points in the support of π. Assume that $x_1 > x_2$; we will have to check if $y_1 \geqslant y_2$ We apply equation 3.3 and the fact that F is a nondecreasing, we get that;

$$G(y_1) \geqslant F(x_1^-) \geqslant F(x_2) \geqslant G(y_2^-).$$

If $G(y_1) > G(y_2^-)$, then we are done. Otherwise, $G(y_1) = F(x_1^-) = F(x_2) = G(y_2^-)$. If $y_2 > y_1$, this means that F is constant on $[x_2, x_1)$, and G on $[y_1, y_2)$. But this is impossible, in the sense that (x_1, y_1) and (x_2, y_2) cannot belong to the support of π. Consider for instance (x_2, y_2). Introduce a small number $\epsilon > 0$, then the rectangle R, whose endpoints have respective first and second co-ordinates equal to $(x + \epsilon, y_2 + \epsilon)$ and $(x - \epsilon, y_2 - \epsilon)$ has zero measure for π. The expression for the measure in terms of H is:

$$\pi[R] = H(x - \epsilon, y_2 - \epsilon) + H(x + \epsilon, y_2 + \epsilon) + H(x - \epsilon, y_2 + \epsilon) + H(x + \epsilon, y_2 - \epsilon)$$

Using the definition of H, the inequalities $x_2 < x_1$ and $y_2 > y_1$, and the nondecreasing properties of F and G, it is easy to show that the preceding expression cancels if ϵ is chosen small enough; then $\pi[R] = 0$. This shows that the assumption $y_2 > y_1$ is impossible and equation 3.3 holds true.

(b) Thus, π has its support included in the monotone subset of $\mathbb{R}^2$, hence it is included in the subdifferential of a convex function. By the Knott-Smith optimality criterion, π is the optimal transference plan. Next, we claim that,

$$\pi = (F^{-1} \times G^{-1})_{\#}\mathcal{L}, \quad (3.5)$$

where, $\mathcal{L}$ stands for Lebesgue measure in [0,1]. It is enough to check this identity on arbitrary rectangle of the form $R(x, y)$ and the equation 3.5 becomes,

$$\pi[R(x,y)] = \mid \{(F^{-1}(t), G^{-1}(t)) \in R(x,y)\} \mid$$

(the notation used $\mid A \mid = \mathcal{L}[A]$) The last quantity is

$$\mid \{t \in \mathbb{R}; F^{-1}(t) \leqslant x\} \cap \{t \in \mathbb{R}; G^{-1}(t) \leqslant y \ \ \mid \tag{3.6}$$

Depending on the cases, $\{F^{-1}(t) \leqslant x\}$ is $[0, F(x))$ or $[0, F(x)]$. Anyway the set whose Lebegue measure is taken is an interval with endpoints 0 and $\min[F(x), G(y)]$ and its measure is equal to $min[F(x), G(y)] = H(x, y)$. This proves the claim.

(c) As a consequence, for any nonnegative measurable function u on $\mathbb{R}^2$

$$\int_{\mathbb{R}^2} u(x,y)d\pi(x,y) = \int_0^1 u(F^{-1}(t), G^{-1}(t))dt$$

□

3.2 Cyclic Monotonicity

Definition 3.2.1. *A subset $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ is said to be cyclically monotone if it fulfills the following condition: for all $m \geqslant 1$, and for all $(x_1, y_1), (x_2, y_2), ...(x_m, y_m)$ in Γ*

$$\sum_{i=1}^{m} |x_i - y_i|^2 \leqslant \sum_{i=1}^{m} |x_i - y_{i-1}|^2 \tag{3.7}$$

with the convention $y_0 = y_m$, or equivalently,

$$\sum_{i=1}^{m} y_i \cdot (x_{i+1} - x_i) \leqslant 0 \tag{3.8}$$

with the convention $x_{m+1} = x_1$

Informally c-cyclic monotone plan is a plan that cannot be improved: it is not possible to perturb it to get a more economical plan.

3.2.1 Rockafellar's Theorem:

Theorem 3.2.1. *A nonempty subset* $\Gamma \subset \mathbb{R}^n \times \mathbb{R}^n$ *is cyclically monotone if and only if it is included in the subdifferential of a proper lower semi-continuous convex function* ϕ *on* $\mathbb{R}^n$*. Moreover, maximal(with respect to inclusion) cyclically monotone subsets are exactly the sub-differentials of proper lower semi-continuous convex functions.*

3.3 Fundamental Theorem of optimal transport

Theorem 3.3.1. *Consider a cost function* $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ *that is continuous. Let* $\mu \in \mathcal{P}(\mathcal{X})$ *and* $\nu \in \mathcal{P}(\mathcal{Y})$ *be such that*

$$c(x, y) \leqslant a(x) + b(y)$$

for some $a \in \mathcal{L}^1(\mu)$ *and* $b \in \mathcal{L}^1(\nu)$*. Consider the Monge- Kantorovich problem*

$$\inf_{\gamma \in \Pi(\mu,\nu)} \int c d\gamma.$$

For a coupling $\gamma \in \Pi(\mu, \nu)$*, the following statements are equivalent:*

(a) γ *is an optimal transport plan.*

(b) The support of γ *is c-cyclically monotone.*

(c) There exists a c-concave function ϕ *on* $\mathcal{X}$ *such that max*$\{\phi, 0\} \in \mathcal{L}^1(\mu)$ *and supp*$(\gamma) \subset \partial^c \phi$

We start the proof later. First we need the Theorem mentioned below in order to prove the implication $(a) \implies (b)$.

Theorem 3.3.2. *Let* $\mathcal{X}$ *and* $\mathcal{Y}$ *be two Polish Spaces, and let* $c : \mathcal{X} \times \mathcal{Y} \to [0, \infty)$ *be a continuous cost function. Consider Kantorovich problem for the pair* $\mu \in \mathcal{P}(\mathcal{X})$ *and* $\nu \in \mathcal{P}(\mathcal{Y})$*, and suppose that the optimal value* $\inf \int c d\gamma$ *is finite. If* γ *is an optimal coupling, then the support of* μ *is c-cyclically monotone*

Proof. (of Theorem 3.3.2) Proof by contradiction. Suppose that γ is the optimal coupling but the support is not c-cyclically monotone. Then we have, for a non-trivial permutation σ on $\{1, 2, ..., N\}$ where $N \geqslant 2$ such that,

$$\sum_{i=1}^{N} c(x_i, y_i) > \sum_{i=1}^{N} c(x_i, y_{\sigma(i)})$$

This allows us to construct a perturbation such that the coupling $\tilde{\gamma} = \gamma + \eta$ that is strictly better than γ, that is $\int cd\tilde{\gamma} < \int cd\gamma$.

Since the cost function is continuous, there exists open neighbourhood of x_i, U_i and y_i, V_i, $i = 1, 2...., N$, such that

$$\sum_{i=1}^{N} c(u_i, v_i) > \sum_{i=1}^{N} c(u_i, v_{\sigma(i)}$$

for all $u_i \in U_i$ and $v_j \in V_j$.

Now we want to move some mass in γ to a better permutation. The part to be removed is a multiple of

$$\sum_{i=1}^{N} \frac{1}{m_i} \gamma|_{U_i \times V_i}$$

where $m_i = \gamma(U_i \times V_i)$. We want to add these masses to $U_i \times V_{\sigma(i)}$ while keeping the marginals unchanged. Write $\gamma_i = \frac{1}{m_i}\gamma|_{U_i \times V_i}$. We can do so by adding

$$\sum_{i=1}^{m} ((\pi_1)_{\#}\gamma_i) \otimes ((\pi_2)_{\#\gamma_{\sigma(i)}})$$

Each term is the product of the first marginal of γ_i and the second marginal of $\gamma_{\sigma(i)}$. Here π_i are the projection maps. The perturbation is given by

$$\eta = \frac{min_{1 \leqslant i \leqslant m} m_i}{N} \sum_{i=1}^{N} [\gamma_i - ((\pi_1)_{\#}\gamma_i) \otimes ((\pi_2)_{\#}\gamma_{\sigma(i)})]$$

The multiple is to ensure that the perturbation is still a probability measure. It is now a routine exercise to check that the perturbed coupling gives a strictly smaller cost. $\square$

Proof. (of Theorem 3.3.1) The $(a) \implies (b)$ is a direct consequence of the previous proof. Now let us make an observation. If ϕ is a c-concave function on $\mathcal{X}$ and $y \in$

$\partial\phi^c(x)$, then for any $x' \in \mathcal{X}$ we have the following:

$$\phi(x) + \phi^c(y) = c(x, y)$$

$$\phi(x') + \phi^c(y) \leqslant c(x', y).$$

Taking the difference, we have

$$\phi(x') \leqslant \phi(x) + c(x', y) - c(x, y).$$

The cost of any transport plan is finite for any $\gamma \in \Pi(\mu, \nu)$ as:

$$\int_{\mathcal{X}\times\mathcal{Y}} c d\gamma \leqslant \int_{\mathcal{X}\times\mathcal{Y}} (a + b) d\gamma = \int_{\mathcal{X}} a d\mu + \int_{\mathcal{Y}} b d\nu < \infty$$

$(b) \implies (c)$ (We refer Theorem 3.2.1 for this part).
Let Γ denote the support of γ. We want to show that the support of γ is contained in the c-superdifferential of a c-concave function on $\mathcal{X}$. Let us fixed an arbitrary pair $(x_0, y_0) \in \Gamma$ and consider function defined as follows:

$$\phi(x) = \inf \left\{ \sum_{i=0}^{N-1} (c(x_{i+1}, y_i) - c(x_i, y_i)) \right\}$$

Where the infimum is taken over $N \geqslant 1$ and $(x_i, y_i) \in \Gamma$. To check that ϕ is a c-concave and its superdifferential is contains the support of γ i.e; Γ, we see that $max\{\phi, 0\} \in \mathcal{L}^1(\mu)$, and let $(x_1, y_1) = (x_0, y_0)$ and $N = 1$. Then,

$$\phi(x) \leqslant c(x, y_0) - c(x_0, y_0),$$

As because of our assumption $c(x, y) \leqslant a(x) + b(y)$ on the right hand side of the equation we have,

$$a(x) + (b(y_0) - c(x_0, y_0)),$$

which confirms that Γ is in the superdifferential of a c-concave function.
$(c) \implies (a)$
Let $\gamma \in \Pi(\mu, \nu)$ be such that Γ is contained in the support of a c-superdifferential function ϕ. Then for any $(x, y) \in \partial\phi$ we have,

$$\phi(x) + \phi^c(y) = c(x, y),$$

and for any general $(x, y) \in \mathcal{X} \times \mathcal{Y}$ we have,

$$\phi(x) + \phi^c(y) \leqslant c(x, y).$$

And, to prove that γ is indeed optimal we need to show that for any $\tilde{\gamma} \in \Pi(\mu, \nu)$, we need to prove that $\int c d\gamma \leqslant \int c d\tilde{\gamma}$, which can be shown as follows:

$$\int c d\gamma = \int \phi(x) + \phi^c(y) d\gamma = \int \phi(x) + \phi^c(y) d\tilde{\gamma} \leqslant \int c d\gamma.$$

This completes the proof. $\square$

Chapter 4

Portfolio Theory

Portfolio theory is a concrete application of the previous mathematical developments. Although we begin this chapter with some financial motivations, the main focus here is to study the geometry of probability distributions rather than the financial issues. Nevertheless, the financial problem are interesting and very important indeed.

4.1 Introduction

Consider investing in stock market. The collection of stocks that a investor holds is called the portfolio. Market portfolio is the collection of all these portfolios, it includes all the stocks in the market. In the forthcoming section we are going to study and compare two kinds of portfolio strategies *viz*. Capitalization weighted portfolio and rebalancing portfolio strategy.

4.1.1 Capitalization weighted portfolio

It is customary to summarize the performance of a stock using a market index; its value is average of collection of representative stocks in the market. Many of these indexes are capitalization weighted, that is the influence of a given stock is proportional to its market capitalization. Let's consider a specific example where the index has only two stocks (stock A and stock B). Suppose the prices at the moment are as follows:

Stock	Prices	# of shares	Market Cap	Market Weight
A	$40	20000	$ 8.0×10^5	0.3076
B	$60	30000	$ 1.8×10^6	0.6923

Here the market capitalization is the product of the number of outstanding shares and the rice of the stock. The market weight of stock A is, for example,

$$\frac{8 \cdot 10^5}{8 \cdot 10^5 + 1.8 \cdot 10^6} = 0.3076$$

Let's assume for simplicity that the value of index today is 100.
Suppose the stock prices tomorrow are $42 for stock A and $55 for stock B The new value of the index will be given by:

$$100 \cdot (0.3076 \cdot \frac{42}{40} + 0.6923 \cdot \frac{55}{60}) = 95.75 \qquad (-4.25\%)$$

Note: The market weight of stocks also change. The new weight of stock A is:

$$\frac{42 \cdot 20000}{42 \cdot 20000 + 55 \cdot 30000} = 0.3373$$

We redo the above calculation to get current value of market index. Add real life complications like dividends, split stocks and you will get how the real index works. We also observe that the value of index can be replicated, in principle, by investing according to the market weights. Because of its popularity in both practice and theory, the market portfolio (approximated by market index) is frequently regarded as a benchmark for portfolio managers ***'beating the market'***

4.1.2 Rebalancing:

Volitility harvesting attempts to capture profit by process called rebalancing. Let us use the example above to expalin the idea of rebalancing. Suppose we start with a initial capital of $2000. We spend the money equally to buy stocks of A and stocks of B.

$$\$2000 = \underbrace{\$1000}_{A\,gets\,50\%} + \underbrace{\$1000}_{B\,gets\,50\%}$$

** here we assume that the stocks are infinitely divisible*

Recall that stock A goes up from \$40 to \$42 and stock B drops from \$60 to \$55. Then, the value of our portfolio now becomes,

$$\underbrace{\$1000 \cdot \frac{42}{40}}_{A} + \underbrace{\$1000 \cdot \frac{55}{60}}_{B} = \$1966.66.$$

The portfolio is no longer equal weighted

$$\underbrace{\$1050}_{A:53.39\%} + \underbrace{\$916.66}_{B:46.66\%} = 1966.66$$

In order to restore equal balance between the two stocks we will need to rebalance the 1966.66 in between the two stocks by selling \$66.67 of stock A to buy stock B. Here we assume that there are no transaction costs.

The contrast between the two trading strategies is clearly visible in the example given below. The following example is taken from [4]

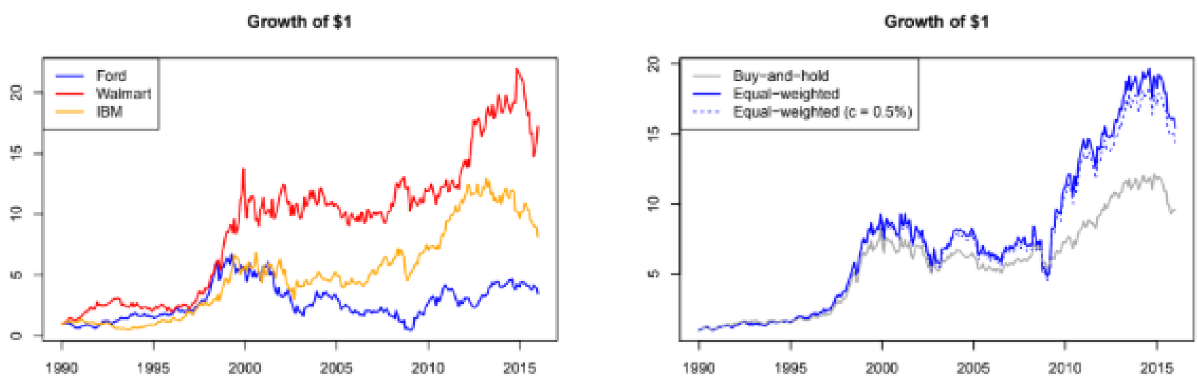

Figure 4.1: Comparing returns of equi-weighted portfolio with 3 stocks

In this example the rebalanced portfolio outperforms the capitalization weighted one. But, it is not the case always. In the following section we study the mathematical conditions required for a rebalanced protfolio to outperform the capitalized weighted one

4.1.3 Two stock case: A simple example

The following example is taken from [4]. Consider 2 stocks, Stock X and stock Y. The price of stock X is constant at \$1 whereas, stock Y goes up by a factor or 2 or drops by factor of $\frac{1}{2}$. A down movement is followed by up movement and vice versa.

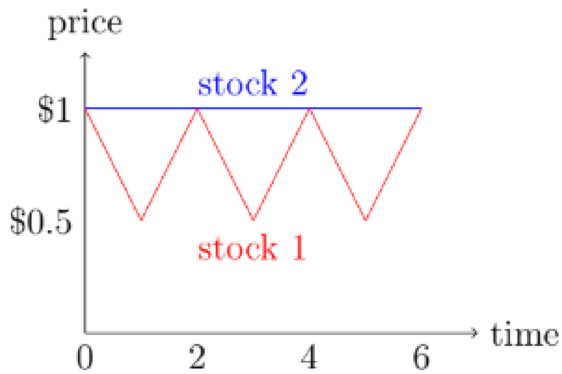

Figure 4.2: Dynamics of two stocks

First consider a buy and hold portfolio. At time 0 investor buys a number of stocks of X and Y, then the portfolio is kept with no further trades. Stock X is constant at \$1 while stock Y fluctuates between \$1 and \$0.5 it is clear that buy and hold will not produce any long term growth.

Now consider equal weighted portfolio. At the begining of each time period the investor rebalances in such a way that equal amount of capital is invested in each stock. After first period we have,

$$\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1 = \frac{3}{4},$$

The simple return is -25%

After the second period the growth is

$$\frac{1}{2} \times 2 + \frac{1}{2} \times 1 = \frac{3}{2}$$

Over these two periods, the portfolio has grown by a factor of $\frac{3}{4} \times \frac{3}{2} = \frac{9}{8} > 1$. If this pattern continues the portfolio will grow exponentially: at every even interval of time the value of the portfolio is $\frac{9}{8}^{\frac{t}{2}}$ times the initial value. Since the portfolio invests equally in two stocks it does not earn money intuitively by anticipating the price movements, rather, profit is generated through rebalancing from volatility of the stocks. The buy-and-hold portfolio on the other hand fails to capture the volatility.

Discussion:

- **Market timing:**
  Since the price of stock Y has a predictable pattern, we should be holding stock X at even times 0,2,4,... and hold stock Y at odd times 1,3,5,... This strategy obviously maximizes the portfolio growth (assuming no shorting is allowed), but the equal weighted portfolio does not aim to time the market. The market timing strategy loses half of its value every time it holds stock Y when it drops, so, it is not a good strategy unless one guesses correctly all or most of the time. **Volatility harvesting strategies does not assume price movements can be predicted correctly**

- **Negative correlation**
  A misconception: rebalancing portfolio is only profitable when the stocks are negatively correlated. This is false. Take the example above and let us derive a formula for our rebalancing portfolio. Whenever the stock Y goes up the portfolio growth is $u = \frac{3}{2}$ and when it goes down $d = \frac{3}{4}$ (we need $ud > 1$). At a given time t the value of equal weighted portfolio is

$$V(t) \;=\; V(0)u^{N_u(t)}d^{N_d(t)} \tag{4.1}$$

  where $N_u(t)$ and $N_d(t) = t - N_u(t)$ are repectively the number of ups an downs of stock Y up untll time t Note that This holds path by path and not by probability (hence, its called pathwise approach). Observe that the value of the portfolio depends only on the cumulative counts of upward and downward movements regardless of their orderings. In particular the, $V(t)$ remains unchanged if we shuffle the upward and downward moves in a given time interval, so it is clear that negative correlation does drive this.

- **Perfect Matching**
  Does rebalancing only work when $N_u(t) = N_d(t)$ i.e; when stock Y reverts back to the starting point? The answer is NO. In fact, this brings up a important point. For each t, let

$$N_{match}(t) = min\{N_u(t), N_d(t)\}$$

  denote the number of pairs of up and down movements up to time t, rearranging the terms in the equation (4.1), the logarithmic growth of the portfolio can be decomposed as following:

$$log\frac{V(t)}{V(0)} = N_{match}(t)\log(ud)\ +\ \delta(t), \tag{4.2}$$

where $ud > 1$ and $\delta(t) = (N_u(t)\ -\ N_{match})\log u$

4.2 Simplex Model

Now, consider a general case with more than two stocks. The main idea here is to represent the state of the market by a probability vector in a unit simplex. This is a highly idealized model of the stock market which has strict assumptions as follows:

- **Assumptions:**
  (a) Time is discrete; $t = 1, 2, 3\ldots$
  (b) The number of stocks involved i.e; n are fixed. $(n \geqslant 2)$
  (c) Each stock has a single outstanding share (infinte divisibility). Thus, it's stock price is equal to its market capitalization. We will denote the capitalization of stock i at time t by $X_i(t)$
  (d) Stocks never die. i.e; $X_i(t) > 0 \quad \forall i, t$
  (e) Shorting of stocks is not allowed.
  (f) Stocks do not pay dividends.
  (g) Trading is immediate and frictionless at each point of time.
  (h) The trading does not influence the price of the stocks, the investor is a 'small' investor.

In this model the data is a path $\big\{X(t) = (X_1(t), X_2(t), \ldots X_n(t)\big\}$ with values in $(0,\infty)^n$. We do not impose any assumptions on dynamics of $\{X(t)\}$; it need not necessarily be a stochastic process, it is indeed some sequence in $(0,\infty)^n$ whose components unfold as time passes by to the investor.

Definition 4.2.1. *(Market weights). The market weight of stock i at a time* $t \geqslant 0$, *is defined by,*

$$\mu(i) := \frac{X_i(t)}{X_1(t)\ + \cdots + X_n(t)}$$

We let $\mu(t) = (\mu_1(t) + \cdots + \mu_n(t))$ be the vector of market weights at time t. $\sum_{i=1}^{n} \mu_i(t) = 1$. Since we assumed that the stocks never die, the market weight vector is an element of open unit simplex.

Definition 4.2.2. *(Unit simplex). For $n \geqslant 2$ the open unit simplex Δ_n in $\mathcal{R}^n$ by*

$$\Delta_n := \left\{ p \ = \ (p_1, \ldots p_n) \in (0,1)^n \ : p_1 + p_2 \cdots + p_n = 1 \right\}$$

The closure of Δ_n in $\mathcal{R}^n$ is denoted by $\overline{\Delta_n}$. The unit simplex can be seen as a probability distributions on a set with n atoms. The evolution of the market with respect to time is given as the sequence $\{\mu(t)\}_{t=0}^{\infty}$. A portfolio vector is an element of closed unit simplex $\overline{\Delta_n}$

Definition 4.2.3. *(Portfolio strategy). The portfolio vector $\pi(t)$ is chosen using the information available at time t. This also includes the history of stock prices upto time t. By a portfolio strategy, we mean the sequence $\pi = \{\pi(t)\}_{t=0}^{\infty}$ taking values in the closed unit simplex $\overline{\Delta_n}$*

4.3 From portfolio to optimal transport

Consider investing in an equity market. At each given time the investor allocates the capital in stocks and forms a portfolio. We will only consider self financing strategies where you are not allowed to borrow or lend any money and we only invest in stocks. The portfolio at any given point is represented as

$$\overline{\Delta^{(n)}} = \left\{ p = (p_1, \ldots, p_n) \in \mathbb{R}, p_i \geqslant 0 \text{ for all i, and } \sum_{i=1}^{n} p_i = 1 \right\} \tag{4.3}$$

Market portfolio is of special importance. The market portfolio is the portfolio with weights $\mu(t) = (\mu_1(t), \ldots, \mu_n(t)$ where, $\mu_i(t)$ and $\mu(t)$ have the meanings as defined above. The market weight takes value in open unit simplex $\Delta^{(n)}$ because of our assumption that no stock dies. The value of this portfolio reflects the growth of entire equity market and is called market index. Market portfolio is used as an investment benchmark and a lot of effort has been put into developing strategies that outperform it. We call a portfolio function to be a map. $\pi : \Delta^{(n)} \to \Delta^{(n)}$, where, if $\mu(t) \in \Delta^{(n)}$

is the current market weight, one chooses the portfolio, $\pi(t) = \pi((\mu(t)) \in \Delta^{(n)}$. This portfolio under the conditions of sufficient volatility and diversity will outperform the market portfolio in large but finite time with probability one. These portfolios are called relative arbitrages with respect to the market. These portfolios are deterministic functions of the current market weights and independent of past and future forecast. We state some definitions used throughout the theory.

4.3.1 Characterizing Pseudo-arbitrage:

Let $K \subset \Delta^{(n)}$. A portfolio function π is called a pseudo-arbitrage on K if the following properties hold true

(a)

Definition 4.3.1. *(Relative value).*

$$V(t) = \frac{\textit{Value at time } t \textit{ of \$1 invested in portfolio } \pi}{\textit{Value at time } t \textit{ of \$1 invested in portfolio } \mu} \tag{4.4}$$

and,

$$V(0) = 1, \qquad V(t+1) = V(t) \sum_{i=1}^{n} \pi_i(\mu(t)) \frac{\mu_i(t+1)}{\mu_i(t)} \tag{4.5}$$

and, $V(t)$ is strictly positive for all t. In order to beat the market we need want to choose a π such that $V(t)$ is large.

(b) There exists a constant $C = C(K, \pi) \geq 0$ such that for all sequences of market weight $\{\mu(t)\}_{t=0}^{\infty}$ taking values in K, we have, $\log V(t) \geq -C$ for all $t \geq 0$

(c) There exists some sequence $\{\mu(t)\}_{t=0}^{\infty} \subset K$ along which $\lim_{t\to\infty} V(t) = \infty$

The first condition states that under generalized diversity condition the portfolio has bounded downside risk regardless of market movement. The second condition states that these is a possibility of unbounded upward gain. These conditions pose a restriction on the portfolio map.

4.3.2 Characterization of pseudo-arbitrages as portfolios that are functionally generated.

Theorem 4.3.1. *The portfolio π is a pseudo arbitrage on a convex subset $K \subset \Delta^{(n)}$ if and only if there exists a concave function $\Phi : \Delta^{(n)} \to [0, \infty)$, satisfying*

(a) the restriction of Φ on K is not affine.

(b) there exists $\epsilon > 0$ that $\inf_{p \in K} \Phi(p) \geq \epsilon$

(c) for any $p \in K$, the vector $\frac{\pi(p)}{p}$ of co-ordinate ratios define a super-gradient of concave funtion $\log \Phi$ at p

if π is continuous then on K it is necessarily given by,

$$\pi_i(p) = p_i\big(1 + D_{e(i)-p} \log \Phi(p)\big) \qquad \text{for } i = 1, 2 \ldots, n \tag{4.6}$$

Here $D_{e(i)-p}$ is the one-sided directional derivative in the direction $e(i) - p$. $e(i)$ is the vector having all zeros except at ith position.

In the above system we say that π is generated by Φ

4.3.3 Pseudo-arbitrage via optimal transport problem:

Here we describe pseudo-arbitrages via solutions to Monge-Kantorovich optimal transport problem: Given a pair of two probability measures P and Q on $\mathbb{R}^{n-1}$ the optimal coupling of these two probability measures with respect to above cost can be expressed in terms of portfolio map of an exponentially concave function. We consider the Monge-Kantorovich optimal transport problem with cost c by:

$$\inf_{R \in \Pi(P,Q)} \mathbb{E}[c(X, Y)] \tag{4.7}$$

A optimal coupling solves Monge problem if we have $Y = F(X)$is deterministic function of X. Specialize to case $\mathcal{X} = \overline{\Delta^{(n)}}$ and $\mathcal{Y} = [-\infty, \infty)^n$. Where P is the probability measure on $\Delta^{(n)}$ and Q is the probability measure on $\mathbb{R}^n$, together with the cost function:

$$c(\mu, h) = \log\left(\sum_{i=1}^{n} e^{h_i} \mu_i\right) \tag{4.8}$$

Interpretation is that μ represents the market weight and h represents the deviation of the portfolio vector from the market weight vector. For a given $\mu \in \Delta^{(n)}$ and $h \in [-\infty, \infty)^n \setminus \{(-\infty, \dots, -\infty)\}$, we can define the portfolio vector π corresponding to μ via the change of measure

$$\frac{\pi_i}{\mu_i} = \frac{1}{\mathbb{E}_\mu[e^h]} e^{h_i}, \qquad i = 1, \dots, n. \tag{4.9}$$

where $\mathbb{E}_\mu[e^h] := \sum_{i=1}^n e^{h_i} \mu_i$.

Theorem 4.3.2. *Let $K \subset \Delta^{(n)}$ and suppose $F : K \to [-\infty, \infty)^n \setminus \{(-\infty, \dots, -\infty)\}$ is a map such that $(\mu, F(\mu)) \in \mathcal{R}\, \forall \mu \in K$. Define portfolio with $h = F(\mu)$. Then there exists a concave function $\Phi : \Delta^{(n)} \to [0, \infty)$ such that part c of theorem 4.3.1 holds true. Then π is a pseudo-arbitrage on K whenever K is an open convex subset and condition (a), (c) of theorem 4.3.1 hold true.*
Conversely, if π is a pseudo-arbitrage over a subset $K \subset \Delta^{(n)}$. Suppose that $\{\log \frac{\pi(p)}{p}, p \in K\}$ is co-ordinatewise bounded below. Define $h = T(\mu)$ as a function of μ via $h_i = \log \frac{\pi_i(\mu)}{\mu_i}$ and consider the coupling (μ, h). For any probability measure P on K and let Q be the distribution of $h(\mu)$ when $\mu \sim P$. Then the coupling (μ, h) solves the transport problem.

4.3.4 Optimal transport problem in term of exponential functions:

We formulate optimal transport in terms of exponential co-ordinate systems of the unit simplex $\Delta^{(n)}$. The advantage of doing so is now we have the transport problem in Euclidean space with strictly convex cost.

Definition 4.3.2. *(Exponential Co-ordinate system). We define the exponential co-ordinates for $p \in \Delta_n$ by the co-ordinates $\theta = (\theta_1, \dots, \theta_{n-1}) \in \mathbb{R}^n$ and is given by*

$$\theta_i = \log \frac{p_i}{p_n} \qquad i = 1, \dots, n-1. \tag{4.10}$$

We denote this map by $\theta : \Delta_n \to \mathbb{R}^{n-1}$. While the inverse transformation $\mathbf{p} = \theta^{-1}$ is given by:

$$p_i = \mathbf{p}_i(\theta) = e^{\theta_i - \psi(\theta)}, \qquad 1 \leqslant i \leqslant n, \tag{4.11}$$

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

By changing co-ordinate systems, any function on Δ_n can be expressed as a function on $\mathbb{R}^{n-1}$ and vice versa. Any function ϕ on Δ_n can be expressed in exponential co-ordinates by $\theta \to \phi(\mathbf{p}(\theta))$. To simplify the notations we will use, $\phi(p)$ or $\phi(\theta)$ depending on the co-ordinate system.

Formulating transport problem:

Let $X = Y = \mathbb{R}^{n-1}$ be equipped with standard Euclidean metric topology. Let P and Q be Borel probability measures on X and Yrespectively. Let $\Pi(P, Q)$ be probability measures on $X \times Y$ whose marginals are P and Q respectively. We define the cost function for $(\theta, \phi) \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ by

$$c(\theta, \phi) := \psi(\theta - \phi), \quad \text{where} \quad \psi(x) := \log\left(1 + \sum_{i=1}^{n-1} e^{x_i}\right) \tag{4.12}$$

The advantage of this formulation is that now the transport is on an Euclidean space with strictly convex cost function. The function $\phi(\theta)$ represents the negative shift in exponential co-ordinates to go from μ to π. Given prior beliefs regarding the possible market weights in future represented by P and a collection of portfolio weights to be chosen from given by Q investor can solve the optimal transport problem to obtain a functionally generated portfolio. The solutions are characterized by Theorem 4.3.1 and Theorem 4.3.2. Both characterizations follow from a property called multiple cyclic monotonicity. Intuitively, this property requires that portfolio does not under perform the market if the market weight goes over any discrete cycle in unit simplex.

Definition 4.3.3. *(Multiplicative cyclic monotonicity). Let* $\pi : \Delta^{(n)} \to \Delta^{(n)}$ *be a portfolio. We say that* π *satisfies MCM if over any cycle* $\{\mu(t)\}_{t=0}^{m+1} \subset \Delta^{(n)}$ *we have* $V(m+1) \geq 1$ *i.e;*

$$\prod_{t=0}^{m} \left(1 + \left\langle \frac{\pi(\mu(t))}{\mu(t)},\, \mu(t+1) - \mu(t) \right\rangle\right) \geq 1 \tag{4.13}$$

Proposition: Let $\pi : \Delta^{(n)} \to \Delta^{(n)}$ be a portfolio map. Then, *pi* satisfies MCM if and only if there is a concave function $\Phi : \Delta^{(n)} \to (0, \infty)$ such that

$$1 + \left\langle \frac{\pi(p)}{p}, q - p \right\rangle \geq \frac{\Phi(q)}{\Phi(p)}, \qquad \text{for all } p, q \in \Delta^{(n)} \tag{4.14}$$

Theorem 4.3.3. *Consider the cost function*

$$c_1(p,q) = log\left(\sum_{i=1}^{n} p_i q_i\right) = \log(\mathbf{p} \cdot \mathbf{q}) \tag{4.15}$$

where $p \in \Delta^{(n)}$ and $q \in \overline{\Delta^{(n)}}$. Given a portfolio mao $\pi : \Delta^{(n)} \to \Delta^{(n)}$. The map defined by $w : \Delta^{(n)} \to \Delta^{(n)}$ via normalized weight ratios:

$$w(p) := \left(\frac{\frac{\pi_i(p)}{p_1}}{\frac{\sum_{j=1}^{n} \pi_j(p)}{p_j}}, \ldots, \frac{\frac{\pi_n(p)}{p_n}}{\frac{\sum_{j=1}^{n} \pi_j(p)}{p_j}} \right) \tag{4.16}$$

then the graph of w is c-cyclically monotone if and only if π is multiplicatively cyclically monotone.

Theorem 4.3.4. *Using the notations in the previous theorem, suppose the exponential co-ordinates of p is θ and the exponential co-ordinate of q is $-\phi$. Then, the cost function takes the form*

$$c_2(p,q) = \log(p \cdot q) = \psi(\theta - \phi) - \psi(\theta) - \phi(\phi) \qquad \theta, \phi \in \mathbb{R}^{n-1} \tag{4.17}$$

thus in exponential co-ordinates, the cost function is equivalent to $\psi(\theta - \phi)$. In particular, the given portfolio map π, the exponential co-ordinates of $\pi(p)$ is $\theta - \phi$ and the exponential co-ordinates of $w(p)$ is $-\phi$. Moreover we have,

$$\phi_i = \theta_i - \log \frac{\pi_i(p)}{\pi_n(p)}, \qquad 1 = 1, \ldots, n-1. \tag{4.18}$$

Via the exponential co=ordinate system we may regard the portfolio map as the function of θ

Chapter 5

Martingale Optimal Transport

The goal of this chapter is prove a Monge-Kantorovich type duality for discrete martingale measures. Such duality plays an important role mathematical finance and has been widely studied very recently. See for instance [1, 2, 3] and references therein.

5.1 Martingale measures

Definition 5.1.1. *(Martingale Measure.) We denote by $\mathcal{M}(\mu_1, \ldots, \mu_n)$ the set of all martingale measures $\mathcal{Q}$ on $\mathbb{R}^n$ having marginals $\mathcal{Q}^1 = \mu_1, \ldots, \mathcal{Q}^n = \mu_n$ and mean s_0. We have, $\mathcal{Q} \in \mathcal{M}(\mu_1, \ldots, \mu_n)$ if and only if $\mathbb{E}_Q[S_i|S_1, \ldots, S_{i-1}] = S_{i-1}$ for $= 2, \ldots, n$. Here $(S_i)_{i=1}^n$ is a co-ordinate process*

$$S_i : \mathbb{R}^n \to \mathbb{R}, \quad S_i(s_1, \ldots, s_n) = s_i \, , \, i = 1, \ldots, n.$$

5.2 Probability measures with given marginals

Let X and Y be complete separable metric spaces. Let $\hat{y}$ and $\hat{z}$ be positive and not necessarily bounded above continuous functions on S and T respectively, bounded away from 0. Put

$$\hat{x} = \hat{y} \circ p_S + \hat{z} \circ p_T,$$

where p_s and p_T are the projections of $S \times T$ on S and T respectively. Let X be the Banach space of continuous functions of x on $S \times T$ such that,

$$||x|| = \sup\{|x(s,t)| \setminus \hat{x}(s,t) \, : \, s \in S, t \in T\} \leq \infty$$

Similarly, let Y and Z be the Banach space of continuous functions y on S and z on T respectively. And, $||y|| = \sup(|y| \setminus \hat{y})(S) \leq \infty$ and $||z|| = \sup(|z| \setminus \hat{z})(T) \leq \infty$. Then, e.g., $y \in Y$ is equivalent to $y \circ p_S \in X$.

Let Let π be the set of all Borel probability measures π in $S \times T$ such that $\hat{x}$ is $\pi-$ integrable, endowed with topology τ generated by the functional $\pi \to \int x d\pi$ for $x \in X$. Let Λ be a non-empty closed convex subset of Π and μ and ν be Borel probability measures in S and T respectively such that $\hat{y}$ and $\hat{z}$ are $\mu-$ and $\nu-$ integrable, repectively

Theorem 5.2.1. *A necessary and sufficient condition for the existence of a probability measure λ inΛ such that ,*

$$\mu = \lambda \circ p_S^{-1} \text{ and } \nu = \lambda \circ p_T^{-1} \tag{5.1}$$

is that

$$\int y d\mu + z d\nu \leq \sup\left\{ \int (y \circ p_s + z \circ p_T) d\gamma \colon \gamma \in \Lambda \right\} \tag{5.2}$$

for all $y \in Y$, $z \in Z$

Theorem 5.2.2. *Let $(\mu_n)_{n\geqslant 1}$ be a sequence of probability measures in $\mathbb{R}^k$. Then a necessary and sufficient condition for the existence of a k-dimensional martingale measure $(\xi_n)_{n\geqslant 1}$ such that the distribution of ξ_n is μ_n for all n is that all μ_n have finite means and for any concave function φ on $\mathbb{R}^k$ the sequence $(\int \varphi \, d\mu_n)_{n\geqslant 1}$ is non increasing.*

Outline of proof. The necessity of the condition is well known from Jensen's inequality. The sufficiency it is enough to prove that if two probability measures μ and ν in $\mathbb{R}^k$ have finite means and satisfy $\int \varphi d\mu \geq \int \varphi d\nu$ for every concave function φ then there is a probability measure λ in $\mathbb{R}^k \times \mathbb{R}^k$ with marginals μ and ν such that the expectation of the first k co-ordinates given the last k co-ordinates is the first k co-ordinates. If this is known to us we can construct a martingale $(\eta_n)_{n\geq 1}$ as a Markov process. To prove the existence of a measure λ we use theorem 5.2.1, set $S = T = \mathbb{R}^k$,

$$\hat{y}(t) = \hat{z}(t) = 1 + |t|$$

Where $|t|$ is the Euclidean length of t and, let Λ be a set of all probability measures in $S \times T$ which are the joint distribution of some $k-$ dimensional martingale

$$\Lambda = \Big\{\lambda : \lambda \in \Pi \text{ and for all bounded continuous functions y on S}$$
$$\int p_T(y \circ p_s)d\lambda = \int p_S(y \circ p_S)d\lambda\Big\}.$$

Because the assumptions before Theorem 5.2.1 are satisfied proof will be complete if we cans show equation 5.1 holds true. Let z_0 be the smallest concave function $\geq z$, i.e; z_0 is the infimum of the set of the affine function on $\mathbb{R}^k$ which are $\geq z$ if this set is non-empty and $z_0 = \infty$ otherwise. Then,

$$\int yd\mu + \int zd\nu \leq \int yd\mu + \int z_0 d\nu \leq \int (y + z_0)d\mu$$
$$\leq \sup_{s \in S}(y(s) + z_0(s)),$$

where $y + z_0 = \infty$ if $z_0 = \infty$. Let r be any real number $< \sup\{y(s) + z_0(s) : s \in \mathcal{S}\}$. Then for some $s \in \mathcal{S}$, $r < y(s) + z_0(s)$. We have to show,

$$r < \sup\left\{\int (y \circ p_S + z \circ p_T)d\gamma \quad \gamma \in \Lambda\right\} \tag{5.3}$$

For any $t \in T$ let Λ_t be the set of probability measures in T with expectation t. The function z_1 on T defined by

$$z_1(t) = \sup\left\{\int zd\alpha \quad \alpha \in \Lambda_t\right\}$$

is concave and $\geq z$, so that $z_1 \geq z_0$ and hence $r < y(s) + z_1(s)$. By the definition of $z_1(s)$ there is an $\alpha \in \Lambda_s$ with,

$$r < y(s) + \int zd\alpha = \int (y \circ p_s + z \circ p_T)d\gamma,$$

where $\gamma = \delta_s \times \alpha$. This proves the theorem. □

5.3 Duality: Martingale Optimal Transport in two dimensions.

5.3.1 Primal Formulation

In the current setting, we fix a exotic option depending only on the value of a single asset S at discrete times $t_1, \ldots, t_n$ and denote by $\Phi(S_1, \ldots, S_n)$ its payoff, where Φ is some measurable function. In no arbitrage framework, the standard approach is to postulate a probability maeasure $\mathcal{Q}$ on $\mathbb{R}^n$ under which the co-ordinate process as defined in Definition 5.1.1 is a discrete martingale in its own filtration. The fair value of Φ is given as the expectation of the payoff

$$\mathbb{E}_{\mathcal{Q}}[\Phi].$$

Additionally, we impose that our model is calibrated to continuum of call options with payoffs $\Phi_{i,K}(S_i) = (S_i - K)_+$, $K \in \mathbb{R}$ at each date t_i and price

$$C(t_i, K) = \mathbb{E}_{\mathcal{Q}}[\Phi_{i,K}] = \int_{\mathbb{R}^+} (s - K)_+ dLaw_{s_i}(s).$$

One dimensional marginals of $\mathcal{Q}$ satisfy

$$\mathcal{Q}^i = Law_{s_i} = \mu_i \text{ for all } i = 1, \ldots, n.$$

We consider the primal problem by considering the lower bound:

$$P = \inf \left\{ \mathbb{E}_Q[\Phi] \ : \ \mathcal{Q} \in \mathcal{M}(\mu_1, \ldots, \mu_n) \right\}. \tag{5.4}$$

5.3.2 Dual Formulation

Dual formulation corresponds to the construction of a semi-static hedging strategy consisting of the static vanilla portfolio and a delta strategy. More precisely, we are interested in the payoffs of the form:

$$\Psi_{(u_i),(\Delta_j)}(s_1,\ldots,s_n) = \sum_{i=1}^{n} u_i(s_i) + \sum_{j=1}^{n-1} \Delta_j(s_1,\ldots,s_j)(s_{j+1}-s_j), \qquad s_1,\ldots,s_n \in \mathbb{R}^n, \tag{5.5}$$

where $u_i : \mathbb{R} \to \mathbb{R}$ are μ_i- integrable for $(i = 1,\ldots,n)$ and the functions $\Delta_j : \mathbb{R}^j \to \mathbb{R}$ are assumed to be bounded mesurable for $j = 1,\ldots,n)$
If these functions lead to a sub-hedging strategy in the sense

$$\Phi \geq \Psi_{(u_i)(\Delta_j)}, \tag{5.6}$$

we have for every pricing measure $\mathcal{Q}$ the inequality

$$\mathbb{E}_{\mathcal{Q}}[\Phi] \geq \mathbb{E}_{\mathcal{Q}}[\Psi_{(u_i)(\Delta_j)}] = \mathbb{E}_{\mathcal{Q}}\left[\sum_{i=1}^{n} u_i(s_i)\right] = \sum_{i=1}^{n} \mathbb{E}_{\mu_i}[u_i]. \tag{5.7}$$

This leads us to consider the dual problem,

$$D = \sup\left\{\sum_{i=1}^{n} \mathbb{E}_{\mu_i}[u_i] \,:\, \Delta_1,\ldots,\Delta_{n-1}\, s.t.\, \Psi_{(u_i),(\Delta_j)} \leqslant \Phi\right\} \tag{5.8}$$

Which by (5.7) satisfies,

$$P \geqslant D \tag{5.9}$$

(The dual problem can be realized by holding a static position in the European options with respective maturity date t_i and payoff u_i.) We will prove that there does not exist a duality gap for martingale measures for two dimensional case. For now, we will state few corollaries and propositions needed to prove the duality for martingale measures. We turn to multi-dimensional version of optimal transport which will be the main tool for the duality proof of martingale measures. Consider probability measures $\mu_1,\ldots,\mu_n$ on real line, with finite moments. We have the set $\Pi(\mu_1,\ldots,_n)$ the set of transport plans of probability measures with marginals $\mu_1,\ldots,\mu_n$. The cost function $\Phi : \mathbb{R}^n \to (-\infty,\infty]$ which is bounded from below by μ_i integrable functions in the sense that

$$\Phi \geq u_1 \oplus \ldots, u_n \tag{5.10}$$

where $u_1\oplus,\ldots,\oplus u_n := u_1(x_1) + \cdots + u_n(x_n)$ Given a cost function Φ and a transport plan π the cost functional is defined as

$$I_\pi(\Phi) = \int_{\mathbb{R}^n} \Phi d\pi \tag{5.11}$$

This would be our primal Monge-Kantorovich problem where we minimize $I_\pi(\Phi)$ over the set of all transport plans $\pi \in \Pi(\mu_1, \ldots, \mu_n)$. Given μ_i integrable functions u_i's such that $\Phi \geq u_1 \oplus \ldots, u_n$ holds true, the dual part of Monge-Kantorovich problem would be to maximize the right hand side of the following equation over a suitable class of functions which satisfy (5.10),

$$\int \Phi d\pi \geq \int (u_1 \oplus \cdots \oplus u_n) d\pi = \int u_1 d\mu_1 + \cdots + \int u_n d\mu_n$$

For the purpose of our application we are going to restrict the u_i's to the following class of functions $\mathcal{S}$. Here, $\mathcal{S}$ is the linear combination of call options known to us and defined as follows:

$$\mathcal{S} = \{u : \mathbb{R} \to \mathbb{R} : u(x) = a + bx + \sum_{i=1}^{m} c_i(x - k_i)_+, a, b, c, k_i \in \mathbb{R}\}$$

Corollary 5.3.1. *Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be a lower semi-continuous function satisfying*

$$\Phi(s_1, \ldots, s_n) \geqslant -K \cdot (1 + |s_1| + \cdots + |s_n|) \tag{5.12}$$

on $\mathbb{R}^n$ for some constant K and let $\mu_1, \ldots, \mu_n$ br probability measurs on $\mathbb{R}$ having finite moments. Then,

$$\begin{aligned} P_{MK}(\Phi) &= \inf \left\{ I_\pi(\Phi) : \pi \in \Pi(\mu_1, \ldots, \mu_n) \right\} \\ &= \sup \left\{ \sum_{i=1}^{n} \int u_i d\mu_i : u_1 \oplus \cdots \oplus u_n \leqslant \Phi, u_i \in \mathcal{S} \right\} \\ &= D_{MK}(\Phi) \end{aligned} \tag{5.13}$$

Lemma 2. *Let $c : \mathbb{R}^n \to \mathbb{R}$ be a continuous cost function and assume that there exists a constant K such that*

$$|c(x_1 + \cdots + x_n)| \leqslant K(1 + |x_1| + \cdots + |x_n|)$$

for all $x_1 \in X_1, \dots x_n \in X_n$. *Then the mapping*

$$\pi \mapsto \int_{\mathbb{R}^n} c d\pi$$

is continuous on $\Pi(\mu_1, \dots, \mu_n)$

Proof. We assumed that the $\mu_1, \dots, \mu_n$ have finite moments, $\int_{\mathbb{R}^n \setminus [-a,a]^n} c d\pi$ converges to 0 uniformly in $\pi \in \Pi(\mu_1, \dots, \mu_n)$ as $a \to \infty$. Then the proof follows from the weak convergence of measure. □

Lemma 3. *Let* $\pi \in \Pi(\mu_1, \dots, \mu_n)$. *Then the following statements are equivalent.*

(a) $\pi \in \mathcal{M}(\mu_1, \dots, \mu_n)$

(b) For $1 \leq j \leq n-1$ *and for every continuous bounded function* $\Delta : \mathcal{R}^j \to \mathcal{R}$ *we have*

$$\int_{\mathcal{R}^n} \Delta(x_1, \dots, x_n)(x_{j+1} - x_j) d\pi(x_1, \dots, x_n) = 0$$

Proof. (a) asserts that whenever $A \subseteq \mathbb{R}^j, i = 1, \dots, (n-1)$ is Borel measurable, then,

$$\int_{\mathbb{R}^n} I_A(x_1, \dots, x_n))(x_{j+1} - x_j) d\pi(x_1, \dots, x_n) = 0.$$

Using standard approximation we can obtain that this is equivalent to (b). □

Lemma 4. *The set* $\mathcal{M}(\mu_1, \dots, \mu_n)$ *is compact in weak topology.*

Proof. We know that $\mathcal{M}(\mu_1, \dots, \mu_n)$ is contained in compact set $\Pi(\mu_1, \dots, \mu_n)$, hence it is sufficient to prove that it is closed. $\mathcal{M}(\mu_1, \dots, \mu_n)$ is the intersection of sets

$$\left\{ \pi \in \Pi(\mu_1, \dots, \mu_n) : \int_{\mathbb{R}^n} f(x_1, \dots, x_n)(x_{j+1} - x_j) d\pi(x_1, \dots, x_n) = 0 \right\}, \tag{5.14}$$

where $j = 1, \dots, n-1$ and $f : \mathbb{R}^j \to \mathbb{R}$ runs through all continuous bounded support functions. By lemma 5, the sets in (5.14) are closed. □

Theorem 5.3.2 (Min-Max). *Let* A *and* B *be convex subsets of vector spaces* V_1 *and respectively* V_2, *where* V_1 *is locally convex and let* $f : A \times B \to \mathbb{R}$. *If,*

(a) A *is compact,*

(b) $f(\cdot, y)$ *is continuous and convex on* A *for every* $y \in B$*,*

(c) $f(x, \cdot)$ *is concave on* B *for every* $x \in A$ *then,*

$$\sup_{y \in B} \inf_{x \in B} f(x, y) = \inf_{x \in A} \sup_{y \in B} f(x, y).$$

Theorem 5.3.3. *(No Duality gap). Assume that* $\mu_1, \ldots, \mu_n$ *are Borel probability measures on* $\mathbb{R}$ *such that* $\mathcal{M}(\mu_1, ..., \mu_n)$ *is a non-empty set. Let* $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ *be a lower semi-continuous function such that*

$$\Phi(s_1, \ldots, s_n) \geqslant -K \cdot (1 + |s_1| + \cdots + |s_n|)$$

on $\mathbb{R}^n$ *for some constant* K*. Then there is no duality gap, i.e.* $P = D$*. Moreover the primal value* P *is attained, i.e, there exists a martingale measure* $\mathcal{Q} \in \mathcal{M}(\mu_1, \ldots \mu_n)$ *such that* $P = \mathbb{E}_Q[\Phi]$

For the existence of a martingale measure having marginals $\mu_1, \ldots, \mu_n$ it is necessary and sufficient that these measures possess the same finite first moments and increase in convex order, i.e.,

$$\mathbb{E}_{\mu_1} \phi \leqslant \ldots \mathbb{E}_{\mu_n} \phi,$$

for each convex function $\phi : \mathbb{R} \to \mathbb{R}$.

Proof. Since we are interested in showing that subhedging portfolios can be formed using just call options, we will restrict the dual candidates $\Psi_{(u_i)(\Delta_j)}$ such that $u_i \in \mathcal{S}$, $\forall i = 1, \ldots, n$ and $\Delta_j \in \mathcal{C}_b(\mathbb{R}^j)$, $j = 1, \ldots, n-1$ and let Φ be a continuous function. We construct convex sets $A = \Pi(\mu_1, \ldots, \mu_n)$ and $B = \mathcal{C}_b(\mathbb{R}) \times \cdots \times \mathcal{C}_b(\mathbb{R}^{n-1})$ of $(n-1)$ tuples so that we can apply Theorem 5.3.2 on the following function:

$$f(\pi, (\Delta_i)) = \int \Phi(x_1, \ldots, x_n) - \sum_{j=1}^{n-1} \Delta_j(x_1, \ldots, x_n)(x_{j+1} - x_j) d\pi(x_1, \ldots, x_n).$$

The assumptions of Theorem 5.3.2 are satisfied and continuity of $f(\cdot, (\Delta_j))$ on $\Pi(\mu_1, \ldots, \mu_n)$ being a consequence of Lemma 2 .

Then, we have,

$$D \geq \sup_{u_i \in \mathcal{S},\, \Delta_j \in \mathcal{C}_b(\mathbb{R}^j),\, \Psi_{(u_i)(\Delta_j)} \leq \Phi} \sum_{i=1}^{n} \int u_i d\mu_i \tag{5.15}$$

$$= \sup_{\Delta_j \in \mathcal{C}_b(\mathbb{R}^j)} \sup_{u_i \in \mathcal{S},\ \sum_{i=1}^n u_i(x_i) \leq \Phi(x_1,\ldots x_n) - \sum_{j=1}^{n-1} \Delta_j(x_1,\ldots,x_n)(x_{j+1}-x_j)} \sum_{i=1}^{n} \int u_i d\mu_i \tag{5.16}$$

$$= \sup_{\Delta_j \in \mathcal{C}_b(\mathbb{R}^j)} \inf_{\pi \in \Pi(\mu_1,\ldots,\mu_n)} \int \Phi(x_1,\ldots,x_n) - \sum_{j=1}^{n-1} \Delta_j(x_1,\ldots,x_n)(x_{j+1}-x_j) d\pi(x_1,\ldots,x_n) \tag{5.17}$$

$$= \inf_{\pi \in \Pi(\mu_1,\ldots,\mu_n)} \sup_{\Delta_j \in \mathcal{C}_b(\mathbb{R}^j)} \int \Phi(x_1,\ldots,x_n) - \sum_{j=1}^{n-1} \Delta_j(x_1,\ldots,x_n)(x_{j+1}-x_j) d\pi(x_1,\ldots,x_n) \tag{5.18}$$

$$= \inf_{\mathcal{Q} \in \mathcal{M}(\mu_1,\ldots,\mu_n)} \int \Phi(x_1,\ldots,x_n) d\mathcal{Q} = P \tag{5.19}$$

Here in step 2, Corollary 5.3.1 is applied on the cost function

$$\int \Phi(x_1,\ldots,x_n) - \sum_{j=1}^{n-1} \Delta_j(x_1,\ldots,x_n)(x_{j+1}-x_j) d\pi(x_1,\ldots,x_n)$$

to show the inequality between (5.16) and (5.17). Theorem 5.3.2 is used to show the inequality between (5.17) and (5.18) . The inequality in the final step is shown by the fact that (by using Lemma 3), for some j there is a function Δ_j such that

$$B = \int \Phi(x_1,\ldots,x_n) - \sum_{j=1}^{n-1} \Delta_j(x_1,\ldots,x_n)(x_{j+1}-x_j) d\pi(x_1,\ldots,x_n)$$

does not vanish if $\mathcal{Q}$ is not a martingale measure then by appropriately scaling Δ the value of B can be made arbitrarily large.

Next, since $\Phi : \mathbb{R}^n \to [0,\infty)$ be a lower-semi continuous function. We can approximate it by a increasing set of functions such that $\Phi_1 \leqslant \Phi_2 \leq \ldots$ and $\Phi = sup_{k\geq 0}\Phi_k$ In the following section we write $P(\Phi), D(\Phi), P(\Phi_k)$ and $D(\Phi_k)$ to emphasize on cost function. For each k pick $\mathcal{Q}_k \in \Pi(\mu_1,\ldots,\mu_n)$ such that,

$$P(\Phi_k) \geq \int \Phi d\mathcal{Q}_k - \frac{1}{k},$$

Passing to a subsequence if necessary, we may assume that $(\mathcal{Q}_k)$ converges weakly to some $\mathcal{Q} \in \Pi(\mu_1, \ldots, \mu_n)$. Then,

$$\begin{aligned}
P(\Phi) \leq \int \Phi d\mathcal{Q} &= \lim_{m\to\infty} \int \Phi_m d\mathcal{Q} \\
&= \lim_{m\to\infty} (\lim_{k\to\infty} \int \Phi_m d\mathcal{Q}_k) \\
&\leq \lim_{m\to\infty} (\lim_{k\to\infty} \int \Phi_k d\mathcal{Q}_k) \\
&= \lim_{k\to\infty} P(\Phi_k)
\end{aligned} \tag{5.20}$$

Since $P(\Phi_k) \leq P(\Phi)$ it follows that $D(\Phi) \geq D(\Phi_k) = P(\Phi_k) \uparrow P(\Phi)$

To show that the optimal value of the primal problem is indeed attained we use the lower semi-continuity of $\int \phi d\pi$ on $\Pi(\mu_1, \ldots, \mu_n)$: if the sequence π_k in $\Pi(\mu_1, \ldots, \mu_n)$ converges weakly to π then,

$$\int \phi d\pi = \lim_{k\to\infty} \int \Phi_k d\pi = \lim_{k\to\infty} (\lim_{m\to\infty} \int \Phi_k d\pi_m) \leq \liminf_{m\to\infty} \int \Phi d\pi_m \tag{5.21}$$

If $P = \infty$, the infimum is trivially attained. So assume, $P < \infty$ and pick a sequence $\mathcal{Q}_k$ in $\mathcal{M}(\mu_1, \ldots, \mu_n)$ such that $P = \lim_k \int \Phi d\mathcal{Q}_k$. As $\mathcal{M}(\mu_1, \ldots, \mu_n)$ is compact $(\mathcal{Q}_k)$ will converge to some measure $\mathcal{Q}$ along a subsequence and $\mathcal{Q}$ is a primal minimizer by (5.21). This completes the proof. $\square$

Bibliography

[1] M. Beiglböck and N. Juillet (2016) On a problem of optimal transport under marginal martingale constraints. Ann. Probab., 44(1):42–106

[2] Y. Dolinsky and H. M. Soner (2014). Martingale optimal transport and ro- bust hedging in continuous time. Probab. Theory Related Fields 160 391–427.

[3] D. Hobson, (2011). The Skorokhod embedding problem and model-independent bounds for option prices. In ParisPrinceton Lectures on Mathematical Finance 2010. Lecture Notes in Math. 2003 267318. Springer, Berlin.

[4] Tim-Kam Leonard Wong. and Soumik Pal, Portfolio Theory, optimal transport and Information Geometry, lecture notes prepared for a topic course MATH599 given at the University of Southern California in Fall 2017.

[5] Mathias Beiglböck, Pierre Henry-Labordere and Friedrich Penker. Model Independent Bounds for Option prices-Mass Transport approach. Finance Stoch. 17 (2013), no. 3, 477-501.

[6] Soumik Pal and Tim-Kam Leonard Wong. Geometry of relative arbitrage. arXiv:1402.3720v5 [q-fin.PM] 28 Jul 2015, Mathematics and Financial Economics 10(3) 263-293, 2015

[7] Soumik Pal and Tim-Kam Leonard Wong. Exponentially concave functions and a new Information Geometry. arXiv:1605.05819v5 [math.PR], 31 May 2017.

[8] V. Strassen. Existence of probability measures with given marginals. The Annals of Mathematical Statistics, Volume 36, Number 2 (1965), 423-439.

[9] Cedric Villani. Topics in Optimal transport, Graduate Text in Mathematics, Volume 58.