Robust Bounds for Option Prices via Martingale Optimal Transport and its Numerics

Bryan Poh Wei Zhao

Supervisor: Prof. Dr. Julian Sester
Examiner: Prof. Dr. Liu Chun Chun

Department of Mathematics
National University of Singapore

Honours Year Project for Semester 1, AY2024/2025

Abstract

This thesis focuses on robust, model-independent option pricing, a framework that seeks to determine the bounds of option prices without relying on specific probabilistic models of the underlying assets. A key tool in this approach is Martingale Optimal Transport (MOT), which combines martingale theory and optimal transport to derive model-free price bounds based on marginal distributions of asset prices. Building on existing work, this thesis reformulates the dual problem of MOT by leveraging the theory of viscosity solutions, providing a novel mathematical perspective on the problem structure. This reformulation not only enhances theoretical understanding but also facilitates the development of novel numerical algorithms. A key contribution of this thesis is the introduction of a neural network-based numerical method to solve the reformulated dual problem. The proposed method approximates the optimal dual solutions in high-dimensional settings, overcoming computational challenges that traditional approaches may face. Numerical experiments demonstrate the effectiveness and accuracy of the proposed methodology, offering a practical tool for robust option pricing in model-independent settings. The results contribute to the growing literature on MOT and reinforce the utility of neural networks in solving complex financial optimization problems.

Acknowledgment

I would like to acknowledge and express my sincere gratitude to my supervisor, Professor Julian Sester, whose guidance, support, and expertise have made this thesis possible. Despite my lack of a strong background in this topic, his mentorship and patience have been instrumental in helping me navigate and successfully complete this work. His deep knowledge and passion for this field have been both inspiring and motivating throughout this journey.

Finally, I would like to give special thanks to my family and my partner, Eve, for their unending support and love, without which, the completion of this thesis may have been unattainable.

Contents

Chapter 1

Introduction

Since the introduction of the Black-Scholes (BS) model in 1973 by Black, Scholes, and Merton [10], the research area of option pricing has seen tremendous developments. Over the years, more sophisticated option pricing model has been developed. These models include:

1. Jump-diffusion models, which account for sudden jumps in asset prices [22].

2. Volatility models which aim to capture the implied volatility of the underlying asset through market information. The implied volatility could be captured either through a deterministic function obtained from the volatilty surface (local volatility models [7]) or a random variable (stochastic volatility models [1, 15]).

3. Lévy models, which uses lévy processes to model jumps and other non-gaussian behaviours in the dynamics of the asset price [25, Chapter 9]

In principle, these pricing methods tend to be similar: fix an underlying probability space, and assume that the future behaviour of the underlying price will be captured by their probabilistic models. And if the probabilistic model prescribes the martingale property to the underlying price, the model could be used to determine the no-arbitrage price for an option of the underlying asset. However, these models may oversimplify or misspecify the true market model, which could render the option price inaccurate. In extreme cases, as in the financial crisis of 2008, these models are incapable of accounting for black swan events.

One approach to address the risk associated with modelling uncertainty is via a model-independent approach, which we call 'Robust Pricing'. It was initially discussed by Hobson [16], which offered a novel way of pricing financial derivatives via model-independent arbitrage. An important tool used in [16] is the Breeden-Litzenberger formula [4], which allows the risk-neutral distribution of the underlying asset price to be inferred from a continuum of call option prices with different strike prices. Nonetheless, the dependence structure between these marginal distributions poses some uncertainty - that is the joint distribution between different assets, or the same asset at different times is unclear. However, we can make use of this ambiguity, to find bounds for option pricing. That is, we identify a joint distribution, whose marginals are that of the underlying, such that the payoff of the option is optimized[1] at maturity, allowing us to establish bounds of the option price.

For exotic options[2] whose payoff depends on multiple underlying assets at a single future time, one could use the theory of optimal transport (OT) to establish robust bounds as shown in [14, section 2.1.9]. In the case where the payoff depends on a single underlying asset at multiple times, the absence of arbitrage compels the unknown coupling between the known distributions to be a martingale, and the resulting optimization problem with this additional constraint is known as the martingale optimal transport (MOT) problem. In the seminal paper by Beiglböck .[2, Theorem 1.1], the

[1]To find a lower (upper) bound, we minimized (maximized) the expectation of the payoff function with respect to all possible joint distribution of the known underlying marginal distribution.

[2]An exotic option is a type of financial derivative that differs from traditional options due to its more complex features, including unique payoff structures, underlying assets, or expiration conditions, which often tailor it to fit specific hedging or investment needs.

authors used OT theory to establish a dual formulation for the problem of finding bounds for the option price and showed that there is no duality gap.

Since then, there has been increased interest in numerical solutions for MOT problems. [13] used a linear programming (LP) method to solve the primal problem of the multi-marginal MOT problem. By discretizing the marginal distributions of the underlying asset and relaxing the martingale constraint, [13] was able to create a sequence of LP problems, which was shown to converge to the primal problem. However, the number of variables is often large, making the problem computationally intractable for many problems of practical interest. This is especially true for multi-marginal problems, since the number of variables grows exponentially in the number of marginals. In terms of solving the dual problem, [8] approximated the semi-static sub-hedging strategy with a neural network, and introduced a suitable penalty function to impose the constraint that the strategy is sub-hedging.

In this thesis, we will be focusing on numerical solutions that takes advantage of the C-convex duality of robust option pricing bounds. In particular, it has been shown in [2, Proposition 4.4] that the dual problem can be reformulated via convex envelopes which reduces the number of constraints of the problem. Recently, [26] extended the C-convex duality to the general n-marginal case, which would form the basis of our numerical method.

The structure of the paper is as follows: In the first chapter, we go through preliminary topics such as option pricing, optimal transport, and viscosity solution. In the second chapter, we introduce the notion of robust option pricing and its primal problem. Then we discuss how the primal problem maybe reformulated via convex envelopes and as a stochastic control problem in chapter 3. Finally, we discuss a numerical method in chapter 4.

Chapter 2

Preliminaries and Notations

In this section, we will go through the preliminary topics required for robust option pricing with MOT and the notations that we will use for the rest of the paper.

2.1 Optimal Transport

The OT problem was first formulated by Gaspard Monge in 1781 [23]. He was interested in finding the most cost efficient way of transporting a mass of soil from one specified configuration to another (see Figure 2.1). We first introduce the push-forward operator, which is key for the formulation of the OT problem.

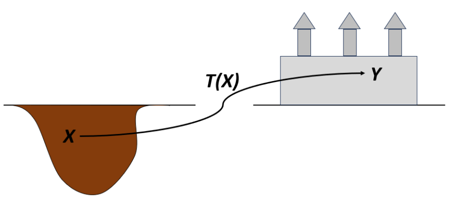

Figure 2.1: A pile of soil with shape or distribution X transported to another location with distribution Y via the transport plan $T(X)$, which induces a cost we want to minimize.

Let (X, Σ) be a measurable space. Denote by $M(X)$ the set of σ-additive functions $\mu : \Sigma \to \mathbb{R}$. Furthermore, we define the following subsets of $M(X)$:

$$M_+(X) := \{\mu \in M(X) : \mu \geq 0\}, \quad \mathcal{P}(X) = \{\mu \in M_+(X) : \mu(X) = 1\}$$

the subsets of non-negative and probability measures respectively.

Definition 2.1.1 (Push forward operator and measure)**.** *Given measurable spaces (X, Σ_x) and (Y, Σ_y). Let $f : X \to Y$ be a measurable function. Then we define the push forward operator $f_\# : M(X) \to M(Y)$ by*

$$f_\#\mu(A) := \mu(f^{-1}(A)) \quad \forall A \in \Sigma_y$$

and call $f_\#\mu$ the push forward measure.

Let (X, Σ_x) and (Y, Σ_y) be measurable spaces. Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be the marginal distributions of X and Y. Denote the cost function by $c(x, y) : X \times Y \to [0, \infty]$.Then we are interested in solving the primal problem:

$$P_M = \inf \left\{ \int_X c(x, T(x))\, d\mu(x) \mid T : X \to Y, T_\#\mu = \nu \right\}.$$

Monge's primal formulation has several restrictions. Namely, it does not allow the transport plan T to split up the mass of X. Furthermore, depending on the nature of the marginal, a solution may not be obtained. For example, if X has a discrete distribution while Y has a continuous distribution, then there is no deterministic transport map T between X and Y.

Now we introduce Kantorovich's formulation [20] of the OT problem. Instead of looking for a deterministic transport map that pushes the source measure μ to ν, [20] considered transport plans, also known as couplings in the probability literature, which allows for probabilistic transportation. A transport plan as follows: Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, define

$$\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(X \times Y) \mid \pi(A \times Y) = \mu(A),\, \pi(X \times B) = \nu(B),\, \forall A \in \Sigma_x,\, B \in \Sigma_y\}.$$

The Kantorovich formulation is as follows:

$$P_K = \inf \left\{ \int_{X \times Y} c(x, y) d\pi(x, y), \middle| , \pi \in \Pi(\mu, \nu) \right\}.$$

A trivial coupling between μ and ν is the tensor product $\mu \otimes \nu$. Therefore, the class of couplings is always non-empty, which means that there will always be a solution to Kantorovich's primal problem, unlike Monge's formulation. Furthermore, Kantorovich's formulation can be extended to n marginals as shown below.

Consider n probability spaces (X_i, μ_i) for $i = 1, 2, \ldots, n$. Let $c(x_1, \ldots, x_n) \colon \mathbb{R}^n \mapsto \mathbb{R}$, be a measurable cost function. We consider a transport plan $\Pi(\mu_1, \ldots, \mu_n)$. Then we have

$$P_K = \inf \left\{ \int_{X_1 \times \cdots \times X_n} c(x_1, \ldots, x_n) d\pi(x_1, \ldots, x_n) \middle| , \pi \in \Pi(\mu_1, \ldots, \mu_n) \right\}.$$

The n-marginal kantorovich formulation of OT, has many applications in areas such as machine learning and physics. For robust option pricing, the Kantorovich formulation is used with a restriction that the transport plan is a martingale. Next we discuss how to formulate the Kantorovich dual problem.

Assume that there exists μ_i-integrable functions, i.e. $u_i \in L^1(\mu_i)$ for $i = 1, 2, \ldots, n$, such that the following inequality holds pointwise in $\mathbb{R}^n$

$$\sum_{i=1}^{n} u_i(x_i) \leq c(x_1, \ldots, x_n), \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n. \tag{2.1.1}$$

Given that the integral is montonic, we have:

$$\begin{aligned}
P_K &= \inf\left\{\int_{X_1\times\cdots\times X_n} c(x_1,\ldots,x_n)\,d\pi(x_1,\ldots,x_n)\middle|\pi\in\Pi(\mu_1,\ldots,\mu_n)\right\}\\
&\le \int_{X_1\times\cdots\times X_n} c(x_1,\ldots,x_n)\,d\pi(x_1,\ldots,x_n)\\
&\le \sum_{i=1}^{n}\int u_i(x_i)\,d\mu_i\\
&\le \sup_{u_i\in L^1(\mu_i)}\left\{\int u_i(x_i)\,d\mu_i(x_i)\middle|\sum_{i=1}^{n}u_i(x_i)\le c(x_1,\ldots,x_n)\right\}\\
&= D_K.
\end{aligned}$$

By definition, we have $P_K \le D_K$, and by the following theorem, we have $P_K = D_K$:

Theorem 2.1.2. *Let $(X_i,\mu_i)_{i=1,\ldots,n}$ be Polish probability spaces, let $c : X_1,\ldots,X_n \to \mathbb{R}$ be lower semi-continuous with $c(x_1,\ldots,x_n) \ge \sum_{i=1}^{n} u_i(x_i)$ for some $u_i \in L^1(\mu_i)$. Then P_K equals to*

$$\sup_{u_i\in L^1(\mu_i)}\left\{\sum_{i=1}^{n}\int u_i(x_i)d\mu_i(x_i)\,\middle|\sum_{i=1}^{n}u_i(x_i)\le c(x_1,\ldots,x_n)\right\}$$

for all $(x_1,\ldots,x_n)\in X_1\times\cdots\times X_n$.

We refer the reader to [21, Theorem 2.2] for the proof of Theorem 2.1.2.

2.2 Option Pricing with Black-Scholes model

We first consider a financial market that allows the trading of a risk-free asset B, an underlying asset S, and a derivative F of the underlying asset. A financial derivative is defined as follows:

Definition 2.2.1 (Financial derivative)**.** *A derivative of an asset S is a financial in-*

strument whose value depends on the underlying asset S.

In this paper, the financial derivative of interest are exotic options. But for the sake of a preliminary introduction to option pricing, we consider a European-style option with payoff $\Phi : \mathbb{R} \to \mathbb{R}$, at maturity time T.

We denote by B_t, S_t, and $F_t := F(t, S_t)$ the price at time t of the risk-free asset, the underlying asset, and the financial derivative with payoff $\Phi(S_T)$ respectively. The BS model makes the following assumptions about the markets:

1. The market is complete, that is, every financial derivative can be perfectly replicated through trading strategies using existing assets B and S.

2. There are no transaction costs, taxes, or other market friction like bid-ask spreads.

3. S and F can be continuously traded.

4. The market does not allow for arbitrage opportunities, i.e., there is no way to make a risk-free profit with zero net investment.

The risk-free asset B has the dynamics:

$$
\begin{aligned}
dB_t &= B_t r dt \\
B_0 &= 1,
\end{aligned}
$$

where r is a constant risk-free rate. Under the BS model, we assume that the price of the underlying asset S follows a Geometric Brownian Motion (GBM):

$$
\begin{aligned}
dS_t &= S_t(\mu d_t + \sigma dW_t) \\
S_0 &= s,
\end{aligned}
\tag{2.2.1}
$$

where μ is the drift and $\sigma \geq 0$ is the volatility of the asset price, s is the initial asset price, and W is a standard Brownian motion. Then the BS model states that the price of the derivative F_t satisfies the PDE:

$$\begin{aligned} \frac{\partial F}{\partial t} + rS\frac{\partial F}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 F}{\partial S^2} - rF = 0 \\ F(T, S_T) = \Phi(S_T). \end{aligned} \tag{2.2.2}$$

To prove that statement, we first consider a portfolio V that replicates the value of the call option at time T as defined below:

Definition 2.2.2 (Replicating Portfolio)**.** *A portfolio with value V_t at time t replicates the value of the option $\Phi(S_T)$ at time T if and only if $V_T = \Phi(S_T)$. Furthermore, by the no-arbitrage assumption, $V_0 = F_0$, where F_0 is the price of the option at time 0.*

Let the value of our replicating portfolio be $V_t := V(t, S_t)$ at time t. We construct a replicating portfolio by investing in the risk-free asset B and the underlying asset S. Then

$$V_t = \alpha_t B_t + \beta_t S_t\,, \tag{2.2.3}$$

where α_t and β_t are the number of units invested in the bond and the underlying asset respectively. We also require that the portfolio is self-financing as defined below:

Definition 2.2.3. *A portfolio is self-financing if there is no external infusion or withdrawal of money. Therefore, any changes to the value of the portfolio is due to changes in the value of the Bond and the underlying asset, i.e.,*

$$dV_t = \alpha_t dB_t + \beta_t dS_t,$$

where $V_t = \alpha_t B_t + \beta_t S_t$.

Using the PDE for the bond, we have

$$dV_t = \alpha_t r B_t d_t + \beta_t dS_t. \tag{2.2.4}$$

Using Ito's lemma, we have a PDE for F_t:

$$dF_t = \left[\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2\right] dt + \frac{\partial F}{\partial S} dS_t. \tag{2.2.5}$$

As mentioned earlier, our goal is to find a replicating portfolio with such that $dV_t = dF_t$. To this end, we first make the coefficients of dS_t the same in both PDE (2.2.4) and (2.2.5). That is we choose

$$\beta_t = \frac{\partial F_t}{\partial S}. \tag{2.2.6}$$

Substitute (2.2.6) into (2.2.3), we have

$$\alpha_t = \frac{1}{B_t}(V_t - \frac{\partial F_t}{\partial S} S_t). \tag{2.2.7}$$

Rewriting the PDE (2.2.4) for V, we have

$$dV_t = r\left(V_t - \frac{\partial F}{\partial S} S_t\right) dt + \frac{\partial F_t}{\partial S}\, dS_t \tag{2.2.8}$$

Assuming that F follows the BS PDE (2.2.2), we have

$$\begin{aligned} dF_t &= \left(\frac{\partial F}{\partial t} + \frac{1}{2}\frac{\partial^2 F}{\partial S^2}\sigma^2 S_t^2\right) dt + \frac{\partial F}{\partial S} dS_t \\ &= r\left(F_t - \frac{\partial F_t}{\partial S} S_t\right) dt + \frac{\partial F}{\partial S} dS_t. \end{aligned} \tag{2.2.9}$$

Comparing the LHS of (2.2.8) and (2.2.9), we have that $dV_t = dF_t$. If $V_0 = F(0, S_0)$, then

$$V_t = V_0 + \int_0^t dV_s = F(0, S_0) + \int_0^T dF_s = F_t \quad \forall t.$$

In particular, $V_T = F(T, S_T) = \Phi(S_T)$. Since portfolio V replicates the financial derivative, by the no-arbitrage assumption, the price of the call option is V_t.

To solve the BS-PDE (2.2.2), we first apply a transformation to $F(t, S_t)$ via

$$G(t, S_t) := e^{r(T-t)} F(t, S_t),$$

which results in the following dynamics for $G(t, S_t)$:

$$\begin{aligned} \frac{\partial G}{\partial t} + rS\frac{\partial G}{\partial S} + \frac{1}{2}\frac{\partial^2 G}{\partial S^2}\sigma^2 S^2 = 0 \\ G(T, S_T) = \Phi(S_T), \end{aligned} \tag{2.2.10}$$

where

$$\begin{aligned} dS_u &= S_u(rdu + \sigma dW_u) \\ S_t &= s \end{aligned} \tag{2.2.11}$$

. Then applying Feynman-Kac formula [19] to $G(t, S_t)$, we have

$$F(t, s) = e^{-r(T-t)} G(t, s) = e^{-r(T-t)} \mathbb{E}[\Phi(S_T) \mid S_t = s], \tag{2.2.12}$$

as a solution to the BS-PDE. Note that the dynamics (2.2.11) of the asset price differs slightly from that of (2.2.1). In particular, the asset price grows at the risk-free rate r instead of μ.

The Black-Scholes model has many limitations. Notably, the Black-Scholes model assumes that the asset returns follow a normal distribution, which has been empirically shown to have heavy tails instead [3]. Furthermore, a glaring weakness of the BS model is that it assumes that the volatility of the asset to be constant, while empirical studies have shown that volatility implied from the market price of the option tend to be non-constant. Nonetheless, despite the flaws of the model, its simplicity made it a widely used model [17].

2.3 Viscosity Solution

Viscosity solutions are a class of generalized solutions to nonlinear partial differential equations (PDEs), particularly those arising in the fields of optimal control and differential games. Introduced in the early 1980s by mathematicians Michael Crandall and Pierre-Louis Lions [6], this concept addresses the limitations of classical solutions when the PDEs are not smooth enough or when classical solutions do not exist.

We first consider a parabolic nonlinear second order PDE:

$$F(t,x,w(t,x),\frac{\partial w}{\partial t}(t,x),D_x w(x),D^2_{xx}w(x)) = 0 \quad ,(t,x) \in [0,T) \times \mathcal{O}, \tag{2.3.1}$$

where $F : [0,T) \times \mathcal{O} \times \mathbb{R} \times \mathbb{R} \times, \mathbb{R}^N \times \mathcal{S}^N \to \mathbb{R}$ is a continuous function and $\mathcal{O}$ is an open set in $\mathbb{R}^N$. Here $\mathcal{S}^N$ denotes the space of all $N \times N$ real valued symmetric matrices. The function F is assumed to be degenerate elliptic, i.e.:

$$M \leq \tilde{M} \implies F(t,x,r,p_t,p,M) \geq F(t,x,r,p_t,p,\tilde{M}),$$

and parabolic, i.e.:

$$p_t \leq \hat{p}_t \implies F(t,x,r,p_t,p,M) \geq F(t,x,r,\hat{p}_t,p,M),$$

where $\tilde{M} \geq M$ means that $\tilde{M} - M$ is a positive semi-definite matrix. In particular, the parabolic condition, implies that we working with PDEs that evolve forward in time. Next we introduce, for a locally bounded function w on $[0,T]\times\mathcal{O}$, its lower and upper semicontinuous envelope w_* and w^* respectively. That is:

$$w_*(t,x)=\liminf_{(t',x')\to(t,x)} w(t',x') \quad \text{and} \quad w^*(t,x)=\limsup_{(t',x')\to(t,x)} w(t',x').$$

Remark 2.3.1. *By definition,* $w_*(t,x)\le w^*(t,x)$*.*

Remark 2.3.2. w^* *is the smallest upper semicontinuous function that is greater than* w*. Similarly,* w_* *is the biggest lower semicontinuous function that is smaller than* w*.*

Definition 2.3.3 (Viscosity solution)**.** *Let* w *be a locally bounded function on* $\mathcal{O}$*.*

(i) w *is a viscosity subsolution of (2.3.1) if:*

$$F\left(t,x,\varphi(t,x),\frac{\partial\varphi}{\partial t}(t,x),D_x\varphi(t,x),D_x^2\varphi(t,x)\right)\le 0,$$

for any $(t,x)\in[0,T]\times O$ *and smooth test function* $\varphi\in C^2([0,T]\times O)$ *such that* (t,x) *is a maximum of* $w^*-\varphi$ *with* $0=(w^*-\varphi)(t,x)$*.*

(ii) w *is a viscosity supersolution of (2.3.1) if:*

$$F\left(t,x,\varphi(t,x),\frac{\partial\varphi}{\partial t}(t,x),D_x\varphi(t,x),D_x^2\varphi(t,x)\right)\ge 0,$$

for any $(t,x)\in[0,T]\times O$ *and smooth test function* $\varphi\in C^2([0,T]\times O)$ *such that* (t,x) *is a minimum of* $w_*-\varphi$ *with* $0=(w_*-\varphi)(t,x)$*.*

(iii) w *is a viscosity solution of (2.3.1) if it is a viscosity subsolution and supersolution.*

Remark 2.3.4. *By defintion, if* w *is a viscosity subsolution (resp. supersolution) of* F*, then its upper (resp. lower) semicontinuous envelope* w^* *(resp.* w_**) is a viscosity subsolution (resp. supersolution) of* F *as well.*

Remark 2.3.5. *If w is a viscosity solution, then by Remark 2.3.4, w^* and w_* is a viscosity subsolution and supersolution respectively. Then by a comparison principle [5][Theorem 3.3], we have that $w* \leq w_*$, which, together with Remark 2.3.1, implies that $w^* = w_*$.*

Chapter 3

Robust Option Pricing

3.1 Primal Problem

Consider an exotic option whose payoff depends only on the value of a single asset S at times $t_1 < \ldots < t_n$. Denote its payoff by $\Phi(S_1, \ldots, S_n)$, where Φ is a measurable function. Let $\mathbb{Q}$ be the joint probability measure of the underlying asset prices $S_1, \ldots, S_n$. Under the no-arbitrage framework, $\mathbb{Q}$ is required to be a martingale measure, i.e.

$$\mathbb{E}_{\mathbb{Q}}[S_{i+1} \mid S_i] = S_i.$$

By [4], we can infer the marginal distribution μ_i of the underlying asset price S_i at time t_i using a continuum of call option prices $\mathcal{C}(t_i, K)$ with different strikes $K \in \mathbb{R}$. Then we impose that the one-dimensional marginals of $\mathbb{Q}$ satisfy

$$\mathbb{Q}^i = \mu_i \, , \, \forall i = 1, 2, \ldots, n.$$

That is, we want $\mathbb{Q}$ to be calibrated to market information. Given the payoff $\varphi_{i,K}(S_i) = (S_i - K)^+$ at each date t_i, the price of the call option is

$$\mathcal{C}(t_i, K) = \mathbb{E}_{\mathbb{Q}}[\varphi(S_i)] = \mathbb{E}_{\mu_i}[\varphi(S_i)].$$

Then the fair value of the exotic option should be the expectation of its payoff $\mathbb{E}_{\mathbb{Q}}[\Phi]$.

Note that $\mathbb{Q}$, as defined above, does not tell us anything about the relationship (or dependence structure) between the underlying asset price at different times. To find a model-independent lower bound, we consider the primal problem:

$$P = \inf\{\mathbb{E}_{\mathbb{Q}}[\Phi] \colon \mathbb{Q} \in \mathcal{M}(\mu_1, \cdots, \mu_n)\},$$

where $\mathcal{M}(\mu_1, \ldots, \mu_n)$ denotes the set of all martingale measures on $\mathbb{R}^n$ that have marginals $\mu_1, \ldots, \mu_n$ and mean s_0, where s_0 is the spot price of the underlying asset at time t_0.

Note that $\mathcal{M} \neq \emptyset$. By [27], if the marginals $\mu_1, \ldots \mu_n$ are increasing in the convex order, i.e. for every convex function $\psi : \mathbb{R} \to \mathbb{R}$, the map

$$t \mapsto \int \psi(x) \mu_t \, d(x)$$

is increasing, and if the marginals have mean s_0, then there exist a martingale measure such that the marginal distributions are $\mu_1, \ldots, \mu_n$. Indeed the marginals of the underlying asset price is increasing in the convex order. In financial terms, given a strike price K, the price of the call option $\mathcal{C}(t, K)$ increases with t, due to the potential for higher payoffs at a later expiry. Therefore, the set of martingale measures $\mathcal{M}(\mu_1, \ldots, \mu_n)$ is non-empty.

3.2 Dual Problem

To formulate an equivalent dual problem, we note that the dual problem D of a minimization problem has to satisfy $D \leq P$. To this end, [2] considered a sub-replicating portfolio and showed that the dual problem consists of maximizing a self-financing portfolio of vanilla call options. Let

$$\Psi_{(u_i),(\Delta_j)}(s_1,\dots,s_n) = \sum_{i=1}^{n} u_i(s_i) + \sum_{j=1}^{n-1} \Delta_j(s_1,\dots,s_j)(s_{j+1}-s_j), \tag{3.2.1}$$

denote the value of the self-financing portfolio, where $s_1,\dots,s_n \in \mathbb{R}$, $u_i : \mathbb{R} \to \mathbb{R}$ are μ_i-integrable ($i = 1,\dots,n$) and the functions $\Delta_j : \mathbb{R}^j \to \mathbb{R}$ are assumed to be bounded measurable ($j = 1,\dots,n-1$).

The financial interpretation of this portfolio is as follows: $u_i(s_i)$ represent the payoff of the vanilla call options that expire at time t_i, and Δ_j represent the self-financing strategy at time t_j. As mentioned, we have $\Psi_{(u),(\Delta_j)} \leq \Phi$. Then taking the expectation of $\Psi_{(u_i),(\Delta_j)}$,

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[\Psi_{(u),(\Delta_j)}(S_1,\cdots S_n)] &= \mathbb{E}_{\mathbb{Q}}\big[\sum_{i=1}^{n} u_i(S_i)\big] + \mathbb{E}_{\mathbb{Q}}\big[\sum_{j=1}^{n-1} \Delta_j(S_1,\dots,S_j)(S_{j+1}-S_j)\big] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mu_i}\big[u_i(S_i)\big] + \mathbb{E}_{\mathbb{Q}}\left[\mathbb{E}_{\mathbb{Q}}\left[\Delta(S_1,\cdots,S_j)(S_{j+1}-S_j)\right] \mid S_1,\cdots,S_j\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mu_i}\big[u_i(S_i)\big] + \mathbb{E}_{\mathbb{Q}}\left[\Delta(S_1,\cdots,S_j)\mathbb{E}_{\mathbb{Q}}\left[S_{j+1}-S_j \mid S_1,\cdots,S_j\right]\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mu_i}\big[u_i(S_i)\big] + \mathbb{E}_{\mathbb{Q}}\left[\Delta(S_1,\cdots,S_j)\times 0\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{\mu_i}\big[u_i(S_i)\big],
\end{aligned}$$

where the second term of the RHS of Equation 3.2.1 vanishes due to the martingale property. Then the dual is obtained by maximizing the expected value of the sub-replicating portfolio:

$$D = \sup \left\{ \sum_{i=1}^{n} \mathbb{E}_{\mu_i}[u_i] : \exists \Delta_1, \ldots, \Delta_{n-1} \text{ s.t. } \Psi_{(u,\Delta_j)} \leq \Phi \right\}. \tag{3.2.2}$$

3.3 Existence of Solution

We state the main result of [2].

Theorem 3.3.1 ([2, Theorem 1.1])**.** *Assume that $\mu_1, \ldots, \mu_n$ are Borel probability measures on $\mathbb{R}$ such that $\mathcal{M}(\mu_1, \ldots, \mu_n)$ is non-empty. Let $\Phi \colon \mathbb{R}^n \to (-\infty, +\infty]$ be a lower semi-continuous function such that*

$$\Phi(s_1, \cdots, s_n) \geq -K(1 + |s_1| + \cdots + |s_n|)$$

on $\mathbb{R}^n$ for some constant K. Then $P = D$. Moreover, the primal value P is attained, i.e., there exists a martingale measure $\mathbb{Q} \in \mathcal{M}(\mu_1, \ldots, \mu_n)$ such that $P = \mathbb{E}_{\mathbb{Q}}[\Phi]$.

The result of Theorem 3.3.1 can be extended to upper bounds for option pricing via [2, Corollary 1.2].

3.4 Proving Theorem 3.3.1

Proving Theorem 3.3.1, requires showing that $D \geq P$ and that $\mathcal{M}(\mu_1, \ldots, \mu_n)$ is compact. The former was rigorously proven using OT theory in [2]. However, the proof of the latter was abbreviated in [2]. It is the interest of this thesis to prove that $\mathcal{M}(\mu_1, \cdots, \mu_n)$ is compact in the weak topology [2, Proposition 2.4]. To this end we require the following lemmas:

Lemma 3.4.1 ([2], Lemma 2.2)**.** *Let $(\mathcal{S}_1,\mu_1),\dots,(\mathcal{S}_n,\mu_n)$ be probability spaces. Let $c\colon \mathbb{R}^n \to \mathbb{R}$ be continuous and assume that that there exists a constant K such that*

$$|c(s_1,\cdots,s_n)| \leq K(1+|s_1|+\cdots+|s_n|)$$

for all $s_1 \in \mathcal{S}_1,\cdots,s_n \in \mathcal{S}_n$. Then the mapping

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

is continuous on the weak topology of $\Pi(\mu_1,\cdots,\mu_n)$.

Lemma 3.4.2 ([2], Lemma 2.3)**.** *Let $\pi \in \Pi(\mu_1,\cdots,\mu_n)$. Then the following are equivalent:*

1. *$\pi \in \mathcal{M}(\mu_1,\cdots,\mu_n)$.*
2. *For $1 \leq j \leq n-1$ and for every continuous bounded function $\Delta\colon \mathbb{R}^j \to \mathbb{R}$, we have*
$$\int_{\mathbb{R}^\kappa} \Delta(s_1,\cdots,s_j)(s_{j+1}-s_j)\,d\pi(s_1,\cdots,s_n)=0$$

As a consequence of Prokhorov's Theorem [28, Lemma 4.4], the set of transport plans $\Pi(\mu_1,\dots,\mu_n)$ is compact in the weak topology. Since $\mathcal{M}(\mu_1,\dots,\mu_n) \subseteq \Pi(\mu_1,\dots,\mu_n)$, then it is sufficient to prove that $\mathcal{M}(\mu_1,\dots,\mu_n)$ is closed in order to prove it is compact in the weak topology as well.

Before proving Lemma 3.4.1 and 3.4.2, we briefly explain why proving the compactness of $\mathcal{M}(\mu_1,\dots,\mu_n)$ is required for the proof of optimality. We first define what it means for a set of measures to be compact in the weak topology.

Definition 3.4.3 (Compactness in the Weak Topology)**.** *A set $\mathcal{K}$ of probability measures is compact in the weak topology if every sequence $(Q_k)_{k\geq 1} \subset \mathcal{K}$ has a subsequence*

$(Q_{k_j})_{j\geq 1}$ *that converges weakly to some probability measure* $Q \in \mathcal{K}$ *That is, for every* $f \in C_b(\mathbb{R})$, *we have*

$$\int f\, dQ_{k_j} \to \int f\, dQ \quad \textit{as } j \to \infty.$$

A key result in general transport theory for a lower semi-continuous function [28, Lemma 4.3], which follows from the compactness in the weak topology, is stated as follows: if there exists a sequence of measures $(\pi_k)_{k\in\mathbb{N}} \in \Pi(\mu_1, \ldots, \mu_n)$ that converges to $\pi \in \Pi(\mu_1, \ldots, \mu_n)$, then the mapping

$$\pi \mapsto \int \Phi\, d\pi$$

is lower semi-continuous. That is

$$\int \Phi\, d\pi \leq \liminf_{k\to\infty} \int \Phi\, d\pi_k.$$

This result guarantees the existence of a minimizer in the following way. Suppose that there exists a sequence $(\pi_k)_{k\in\mathbb{N}} \subset \Pi(\mu_1, \ldots, \mu_n)$ such that

$$\lim_{k\to\infty} \Phi\, d\pi_k = \inf_{\pi\in\Pi} \int \Phi\, d\pi.$$

By the compactness of $\Pi(\mu_1, \ldots, \mu_n)$ in the weak topology, we can extract a subsequence $(\pi_{k_j})_{j\in\mathbb{N}}$ that converges weakly to some measure $\pi^* \in \Pi(\mu_1, \ldots, \mu_n)$. Using the lower semi-continuity of the functional $\pi \mapsto \int \Phi\, d\pi$, it follows that:

$$\int \Phi\, d\pi^* \leq \liminf_{j\to\infty} \int \Phi\, d\pi_{k_j}.$$

Since (π_k) is a minimizing sequence, we have:

$$\int \Phi \, d\pi^* = \inf_{\pi \in \Pi(\mu_1, \ldots, \mu_n)} \int \Phi \, d\pi.$$

Thus, π^* is a minimizer of the problem, and the infimum is attained. If $\mathcal{M}(\mu_1, \ldots, \mu_n)$ is compact in the weak topology, then this result can be applied to the proof of Theorem 3.3.1. Therefore, the weak compactness of $\mathcal{M}(\mu_1, \ldots, \mu_n)$ is required.

3.4.1 Proving Lemma 3.4.1

Proof. Let $\{\pi_m\} \in \Pi(\mu_1, \cdots, \mu_n)$ be a sequence such that $\lim_{m \to \infty} \pi_m = \pi$. We want to show that

$$\lim_{m \to \infty} \int_{\mathbb{R}^n} c \, d\pi_m = \int_{\mathbb{R}^n} c \, d\pi$$

However, $c(s_1, \cdots, s_n)$ is not necessarily bounded. Define

$$c_L(s_1, \cdots, s_n) = \begin{cases} c(s_1, \cdots, s_n) & \text{if } |c(s_1, \cdots, s_n)| \leq L, \\ L & \text{if } c(s_1, \cdots, s_n) > L, \\ -L & \text{if } c(s_1, \cdots, s_n) < -L. \end{cases}$$

Then $\lim_{L \to \infty} c_L(s_1, \cdots, s_n) = c(s_1, \cdots, s_n)$. By definition, $|c_L| \leq L$, i.e. c_L is bounded. Note that c_L is also continuous as well[1]. Since $c_L \in C_b(\mathbb{R}^n)$,

$$\lim_{m \to \infty} \int_{\mathbb{R}^n} c_L \, d\pi_m = \int_{\mathbb{R}^n} c_L \, d\pi$$

Let $g(s) = k(1 + |s_1| + \cdots + |s_n|)$. Note that

$$c(s_1, \cdots, s_n) \leq g(s) \implies c_L(s_1, \cdots, s_n) \leq g(s).$$

[1]Given two continuous functions f and g, $\max(f, g)$ is continuous.

Since S_i's have finite first moments, $g(s)$ is π-integrable. By Dominated Convergence Theorem,

$$\lim_{L\to\infty} \int_{\mathbb{R}^n} c_L \, d\pi = \int_{\mathbb{R}^n} c \, d\pi$$

To show the continuity of the integral with respect to c, we note that for sufficiently large L and all m large enough,

$$\begin{aligned}\left| \int_{\mathbb{R}^n} c \, d\pi_m - \int_{\mathbb{R}^n} c \, d\pi \right| \leq \left| \int_{\mathbb{R}^n} (c - c_L) \, d\pi_m \right| + \left| \int_{\mathbb{R}^n} c_L \, d\pi_m - \int_{\mathbb{R}^n} c_L \, d\pi \right| \\ + \left| \int_{\mathbb{R}^n} (c_L - c) \, d\pi \right|\end{aligned}$$

The second term on the right-hand side goes to zero as $m \to \infty$ for each fixed L due to weak convergence. The first and third term can be made arbitrarily small by choosing L sufficiently large enough because c_L converges to c pointwise and the dominating function g(x) bounds $|c(s_1, \cdots, s_n) - c_L(s_1, \cdots, s_n)|$. Thus, by choosing L large enough, we can ensure that for all sufficiently large m,

$$\left| \int_{\mathbb{R}^n} (c - c_L) \, d\pi_m \right| \leq \frac{\epsilon}{3}$$

and

$$\left| \int_{\mathbb{R}^n} (c - c_L) \, d\pi \right| \leq \frac{\epsilon}{3}.$$

Since the integral of c_L with respect to π_m converges to the integral of c_L with respect to π, we can choose M such that for all $m \geq M$,

$$\left| \int_{\mathbb{R}^n} c_L \, d\pi_m - \int_{\mathbb{R}^n} c_L \, d\pi \right| \leq \frac{\epsilon}{3}.$$

Adding these inequalities, we get that for all $m \geq M$,

$$\left| \int_{\mathbb{R}^n} c \, d\pi_m - \int_{\mathbb{R}^n} c \, d\pi \right| \leq \epsilon,$$

which proves that the mapping $\pi \mapsto \int_{\mathbb{R}^n} c \, d\pi$ is continuous with respect to the weak topology on Π. $\square$

3.4.2 Proving Lemma 3.4.2

Proof. Prove ($\mathbf{1} \implies \mathbf{2}$): Suppose $\pi \in \mathcal{M}(\mu_1, \cdots, \mu_n)$. Due to the martingale property, the following holds

$$\begin{aligned} \mathbb{E}_\pi \left[S_{j+1} \mid S_1, \cdots, S_j \right] &= S_j \\ \mathbb{E}_\pi \left[S_{j+1} \mid S_1, \cdots, S_j \right] &= \mathbb{E}_\pi \left[S_j \mid S_1, \cdots, S_j \right] \\ \mathbb{E}_\pi \left[S_{j+1} - S_j \mid S_1, \cdots, S_j \right] &= 0, \end{aligned}$$

for $1 \leq j \leq n-1$. Then

$$\begin{aligned} &\int_{\mathbb{R}^n} \Delta(S_1, \cdots, S_j)(S_{j+1} - S_j) \, d\pi(S_1, \cdots, S_n) \\ &= \mathbb{E}_\pi \left[\Delta(S_1, \cdots, S_j)(S_{j+1} - S_j) \right] \\ &= \mathbb{E}_\pi \left[\mathbb{E}_\pi \left[\Delta(S_1, \cdots, S_j)(S_{j+1} - S_j) \right] \mid S_1, \cdots, S_j \right] \\ &= \mathbb{E}_\pi \left[\Delta(S_1, \cdots, S_j) \mathbb{E}_\pi \left[S_{j+1} - S_j \mid S_1, \cdots, S_j \right] \right] \\ &= 0 \end{aligned}$$

for $1 \leq j \leq n-1$. Therefore, $\mathbf{1} \implies \mathbf{2}$.

Prove ($\mathbf{2} \implies \mathbf{1}$): Let $\Delta \colon \mathbb{R}^j \to \mathbb{R}$ be a continuous bounded function. Then using the tower property of expectations, rewrite the integral

$$
\begin{aligned}
&\int_{\mathbb{R}^n} \Delta(s_1, \cdots, s_j)(s_{j+1} - s_j)\, d\pi(s_1, \cdots, s_n) \\
&= \mathbb{E}_\pi \left[\Delta(s_1, \cdots, s_j)(s_{j+1} - s_j)\right] \\
&= \mathbb{E}_\pi \left[\mathbb{E}_\pi \left[\Delta(s_1, \cdots, s_j)(s_{j+1} - s_j)\right] \mid s_1, \cdots, s_j\right] \\
&= \mathbb{E}_\pi \left[\Delta(s_1, \cdots, s_j)\mathbb{E}_\pi \left[s_{j+1} - s_j \mid s_1, \cdots, s_j\right]\right] \\
&= \int_{\mathbb{R}^n} \Delta(s_1, \cdots, s_j)\mathbb{E}_\pi \left[s_{j+1} - s_j \mid s_1, \cdots, s_j\right] \, d\pi(s_1, \cdots, s_n)
\end{aligned}
$$

. Given that the integral equals to zero and that $\Delta(x_1, \cdots, x_j) \in \mathbb{C}_b(\mathbb{R}^n)$, by the Fundamental Lemma of the Calculus of Variations,

$$\mathbb{E}_\pi \left[s_{j+1} - s_j \mid s_1, \cdots, s_j\right] = 0.$$

Therefore, ($\mathbf{2} \implies \mathbf{1}$). □

3.4.3 Proof of Compactness of $\mathcal{M}(\mu_1, \ldots, \mu_n)$

Proof. Let $f\colon \mathbb{R}^j \to \mathbb{R}$ be a continuous bounded function. Define the set X_j as:

$$X_j = \left\{\pi \in \Pi(\mu_1, \ldots, \mu_n) \colon \int_{\mathbb{R}^n} f(s_1, \ldots, s_j)(s_{j+1} - s_j)\, d\pi(s_1, \ldots, s_n) = 0\right\}.$$

Then X_j is the pre-image of the set $\{0\}$ under the mapping defined by:

$$\Pi(\mu_1, \cdots, \mu_n) \ni \pi \mapsto \int_{\mathbb{R}^n} f(s_1, \ldots, s_j)(s_{j+1} - s_j)\, d\pi(s_1, \ldots, s_n).$$

We first show that X_j is closed for $j = 1, \cdots, n-1$. Since $f(s_1, \cdots, s_j)$ is bounded, $|f(s_1, \cdots, s_j)| \leq K$ for some $K \in \mathbb{R}$. It follows that

$$|f(s_1, \cdots, s_j)(s_{j+1} - s_j)| \leq K|s_{j+1} - s_j|.$$

Then by Lemma 2.2, the mapping

$$\pi \mapsto \int_{\mathbb{R}^n} f(s_1, \cdots, s_j)(s_{j+1} - s_j)\, d\pi(s_1, \cdots, s_n).$$

is continuous with respect to the weak topology on $\Pi(\mu_1, \cdots, \mu_n)$. Since the image of the mapping is the closed set $\{0\}$, the pre-image of the mapping, i.e. X_j, is a closed set as well. By Lemma 2.3,

$$\mathcal{M}(\mu_1, \cdots, \mu_n) = \bigcap_{j=1}^{n-1} X_j\,,$$

where $\bigcap_{j=1}^{n-1} X_j$ is a closed set. Since $\mathcal{M}(\mu_1, \cdots, \mu_n)$ is a closed subset of $\Pi(\mu_1, \cdots, \mu_n)$ which is compact, $\mathcal{M}$ is compact in the weak topology. $\square$

Chapter 4

Dual Reformulation

This chapter is organized as follows. We discuss the reformulation of the 2-marginal dual problem via convex envelopes (Section 4.1) and how it is related to the value function of a stochastic control problem (Section 4.2). Then we will show how the result of Section 4.1 and Section 4.2 can be extended to the n-marginal case in Section 4.3 and Section 4.4.

4.1 2-Marginal C-convex Duality

First we define the convex envelope[1] as follows:

Definition 4.1.1 (Convex Envelope)**.** *Let $f : \mathbb{R} \to \mathbb{R}$. Then $g : \mathbb{R} \to \mathbb{R}$ is a convex envelope of f if g is the largest convex function such that $g \leq f$, i.e.*

$$g = \sup \{h(x) \mid h(x) \textit{ is convex}, \, h(x) \leq f(x), \, \forall x \in \mathbb{R}\} .$$

We denote the convex envelope of f as f^{**}. In the one-time step setting, that is the

[1]The concave envelope of f can also be defined similarly as the smallest concave function that majorizes f.

2-marginal case, [2] reformulated the dual problem as follows:

Proposition 4.1.2 ([2], Proposition 4.4)**.** *Let $\Phi : \mathbb{R}^2 \to (-\infty, \infty]$ be a lower semi-continuous function such that $\Phi(s_1, s_2) \geq -K(1 + |s_1| + |s_2|)$, $s_1, s_2 \in \mathbb{R}$, and assume that there is some $Q \in \mathcal{M}(\mu_1, \mu_2)$ satisfying $\mathbb{E}_Q[\Phi] < \infty$. Then*

$$P = \sup_{u_2 : \mathbb{R} \to \mathbb{R}, \int |u_2| d\mu_2 < \infty} \left(\mathbb{E}_{\mu_1} \left[(\Phi(S_1, \cdot) - u_2(\cdot))^{**} (S_1) \right] + \mathbb{E}_{\mu_2}[u_2(S_2)] \right).$$

We refer the reader to [2, Proposition 4.4] for the proof of Proposition 4.1.2. As stated in [2, section 4.4], the benefit of the dual reformulation lies in the removal of the inequality constraint of the semi-static sub-hedging strategy, $\Psi_{(u_i, \Delta_j} \leq \Phi$, which simplifies the construction of the sub-hedging strategy.

4.2 2-Marginal Case and HJB Equations

We first introduce a control diffusion problem in finite horizon, whose viscosity solution is a convex envelope.

4.2.1 Control Diffusion Problem in Finite Horizon

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $T > 0$ a finite time, $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ a filtration satisfying the usual conditions[2], and B a 1-dimensional Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, P)$. We consider the controlled process $X_t \in \mathbb{R}$:

$$dX_t = \sigma_t \, dB_t, \tag{4.2.1}$$

[2]We say that $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ satisfies the usual condition if $(\Omega, \mathcal{F}, P)$ is complete, $\mathcal{F}_0$ contains all the P-null sets in $\mathcal{F}$, and $(\mathcal{F}_t)_{t \geq 0}$ is right continuous. See [18, Chapter 1] for a detailed introduction.

where the control $\sigma = (\sigma_t)_{0\leq t\leq T}$ is a progressively measurable process valued in $[0,\infty]$. Given $(t,x) \in [0,T]\times\mathbb{R}$, and a square-integrable control process σ, we denote by $\left\{X_T^{t,x},\, t\leq s\leq T\right\}$ as the strong solution to (4.2.1). Now consider a loss function

$$J(t,x,\sigma) = \mathbb{E}[g(X_T^{t,x})], \tag{4.2.2}$$

which we want to minimize. For the expectation to be well defined, we impose that g has a linear growth condition. Then, the value function for this stochastic control problem is

$$v(t,x) = \inf_{\sigma\in[0,\infty]} \{J(t,x,\sigma)\} = \inf_{\sigma\in[0,\infty]} \mathbb{E}\left[g(X_T^{t,x})\right].$$

Our goal is to find a control process σ that satisfies $v(0,x)$.

4.2.2 HJB Equation and Viscosity Solution

We start by assuming that $V \in C^{1,2}([0,T]\times\mathbb{R})$. Then the HJB equation[3] associated with the value function is

$$\frac{\partial V}{\partial t}(t,x) + \inf_{\sigma\in[0,\infty]} \left\{\frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial x^2}(t,x)\right\} = 0, \tag{4.2.3}$$

with terminal condition:

$$V(T,x) = g(x). \tag{4.2.4}$$

From the HJB equation (4.2.3),

$$\inf_{\sigma\in[0,\infty]} \left\{\frac{1}{2}\sigma^2\frac{\partial^2 V}{\partial x^2}(t,x)\right\} = \begin{cases} 0 & \text{if } \frac{\partial^2 V}{\partial x^2}(t,x) \geq 0, \\ -\infty & \text{if } \frac{\partial^2 V}{\partial x^2}(t,x) < 0. \end{cases} \tag{4.2.5}$$

[3]We refer the reader to [18, Chapter 3] for a detailed derivation of the HJB equation for a general control diffusion problem in finite horizon.

From (4.2.5), we have the associated HJB variational inequality:

$$\min\left\{\frac{\partial V}{\partial t}(t,x), \quad \frac{\partial^2 V}{\partial x^2}(t,x)\right\} = 0. \tag{4.2.6}$$

Then by [24, Theorem 3.1], V is a visocsity solution[4] to the HJB variational inequality (4.2.6).

4.2.3 Convex Envelope as a Viscosity Solution

The HJB variational inequality (4.2.6) asserts that the value function V is convex in x. The implication is that

$$\frac{\partial V}{\partial t}(t,x) = 0 \implies V(T,x) = V(t,x), \quad \forall t \in [0,T].$$

Using terminal condition (4.2.4),

$$V(0,x) = V(T,x) = g(x). \tag{4.2.7}$$

In general, however, V may be discontinuous at T such that the terminal condition (4.2.4) does not hold. One of the reasons is that g may not be convex in x. This is where viscosity solution play a crucial role in providing a weak solution to the HJB variational inequality (4.2.6).

In order to characterize the value function, the relevant terminal condition is to check if $V(T^-,x) := \lim_{t\nearrow T} V(t,x)$ exists. By [24, Theorem 3.2], $\hat{V} := V(T^-,\cdot)$ exist and is a viscosity solution to

$$\min[g(x) - \hat{V}(x), \quad \frac{\partial^2 \hat{V}}{\partial x^2}] = 0 \tag{4.2.8}$$

[4] We refer the reader to [24] for the proof of the aforementioned statement.

Next let $\tilde{g}$ be the lower $\mathcal{G}$-envelope of g, defined as the largest function below g and viscosity supersolution to

$$\mathcal{G}(\tilde{g}) := \frac{\partial^2 \tilde{g}}{\partial x^2} = 0, \tag{4.2.9}$$

By definition, $\hat{V}$ is the viscosity supersolution to (4.2.8), which means that $\hat{V} \leq g$[5]. It is also a viscosity supersolution to (4.2.9). Then by definition of $\tilde{g}$, $\tilde{g} \geq \hat{V}$. On the other hand, $\tilde{g}$ is also the viscosity supersolution to the variational inequality (4.2.8). By the comparison principle (see Remark 2.3.5), we have

$$\tilde{g} \leq \hat{V} \implies \hat{V} = \tilde{g} = V(T^-, x).$$

By definition of $\tilde{g}$, $\tilde{g}$ is the convex envelope of g, and this implies that $\hat{V} = \tilde{g}$ is convex as well. Since the continuity of V at T is ensured by the viscosity solution, from (4.2.7), we have

$$V(t, x) = \hat{V}, \quad \forall t \in [0, T].$$

Therefore, the viscosity solution of the HJB equation (4.2.6) is the convex envelope of g. Since $V(0, x) = V(T^-, x)$, we do not have to solve $V(t, x)$ for all $t \in [0, T]$. Therefore, we reformulate HJB variational inequality (4.2.6) as follows:

$$\min[g(x) - V(x), \quad \frac{\partial^2 V}{\partial x^2}] = 0, \tag{4.2.10}$$

and solving (4.2.10) will yield the same viscosity solution as (4.2.6).

4.2.4 2-marginal MOT

As mentioned previously, the convex envelope of a value function is a viscosity solution to (4.2.10). [14, Corollary 2.2] made use of this fact, to give an alternative characteri-

[5] This follows from the defintion of viscosity supersolutions 2.3.3 and Remark 2.3.2

zation of the 2-marginal C-convex duality.

Corollary 4.2.1 ([14, Corollary 2.2])**.**

$$D_{HJB} = \sup_{u_2(\cdot)\in L^1(\mu_2)} \{\mathbb{E}_{\mu_1}[V(0,S_1,S_1)] + \mathbb{E}_{\mu_2}[u_2(S_2)]\} = P,$$

where

$$V(t,s_1,s) = \inf_{\sigma\in[0,\infty]} \mathbb{E}\left[\Phi(s_1,X_T) - u_2(X_T) \mid X_t = s\right]$$

with $dX_t = \sigma_t dB_t$. *B is a Brownian motion and σ is an adapted (with respect to the filtration of B) unbounded control process.*

Proof. For fixed s_1, $V(0,s_1,s)$ corresponds to the value function of the stochastic control problem which consists in minimizing the expectation of the loss function)

$$g(s) = \Phi(s_1, X_T^{0,s}) - u_2(X_T^{0,s}),$$

with respect to a control $\sigma \in [0,\infty]$, where

$$dX_t = \sigma_t\, dB_t, \quad t\in[0,T],$$
$$X_0 = s.$$

Then by the discussion in Section 4.2.3, the convex envelope of $g(\cdot) = (\Phi(s_1,\cdot)-u_2(\cdot))(x)$ is the viscosity solution to the HJB variational inequality

$$\min[\Phi(s_1,s) - u_2(s) - V(0,s_1,s), \quad \frac{\partial^2 V}{\partial x^2}] = 0. \tag{4.2.11}$$

In particular, $V(0,S_1,S_1) = (\Phi(S_1,\cdot) - u_2(\cdot))^{**}(S_1)$. Then by Proposition 4.1.2, we have $D_{HJB} = P$. $\square$

4.3 N-marginal C-convex Duality

The C-convex duality in the 2-marginal case can be extended to the n-marginal case as shown by Sester [26]:

Proposition 4.3.1 ([26, Proposition 2.1]). *Let $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ be lower semi-continuous and $\Phi(s_1, \dots, s_n) \geq -K(1 + |s_1| + \cdots + |s_n|)$ for all $s_1, \dots, s_n \in \mathbb{R}$ and some $K \in \mathbb{R}$. Additionally assume there exists some $\mathbb{Q} \in \mathcal{M}(\mu_1, \dots, \mu_n)$ such that $\mathbb{E}_{\mathbb{Q}}[\Phi(S_1, \dots, S_n)] < \infty$. We set*

$$c_n : L^1(\mu_2) \times \cdots \times L^1(\mu_n) \times \mathbb{R}^n \to \mathbb{R}$$

$$(u_2, \dots, u_n, s_1, \dots, s_n) \mapsto \Phi(s_1, \dots, s_n) - \sum_{i=2}^{n} u_i(s_i),$$

and define inductively for $i = n-1, \dots, 1$

$$c_i : L^1(\mu_2) \times \cdots \times L^1(\mu_i) \times \mathbb{R}^i \to \mathbb{R}$$

$$(u_2, \dots, u_n, s_1, \dots, s_i) \mapsto \left(c_{i+1}(u_2, \dots, u_n, s_1, \dots, s_i, \cdot)\right)^{**}(s_i).$$

Then, we have

$$\inf_{\mathbb{Q} \in \mathcal{M}(\mu_1, \dots, \mu_n)} \mathbb{E}_{\mathbb{Q}}[c(S_1, S_2, \dots, S_n)] = \sup_{\substack{u_{(i)} \in L^1(\mu_i) \\ i=2,\dots,n}} \mathbb{E}_{\mu_1}\left[c_1(u_2, \dots, u_n, S_1)\right] + \sum_{i=2}^{n} \mathbb{E}_{\mu_i}[u_i(S_i)].$$

Proof. We refer the reader to [26, Section 3] for the proof. □

The C-convex formulation of the dual could give rise to new numerical schemes to solve MOT problems. The crux now lies in the computation of convex biconjugates (or envelopes). To this end, [26, Section 2.1] makes use of the representation of the convex biconjugate in terms of applying the convex conjugate $f*$ twice to a function f

as shown below:

$$\mathbb{R} \ni x \mapsto f^{**}(x) = (f^*)^*(x) = \sup_{m \in \mathbb{R}} \left\{ m \cdot x - \sup_{y \in \mathbb{R}} \{ y \cdot m - f(y) \} \right\}$$

. A practical implementation[6] of this is provided by [26] for a 3-marginal example from [9, Section 5.4].

4.4 N-marginal Case and HJB Equations

To extend the 2-marginal HJB characterization of the MOT problem, we use the recursive scheme of [26] to generate $n-1$ convex envelopes, each of which are viscosity solutions to a HJB equation. We require the following corollary:

Corollary 4.4.1. *Let* $\Phi : \mathbb{R}^n \to (-\infty, \infty]$ *be lower semi-continuous and* $\Phi(s_1, \ldots, s_n) \geq -K(1 + |s_1| + \cdots + |s_n|)$ *for all* $s_1, \ldots, s_n \in \mathbb{R}$ *and some* $K \in \mathbb{R}$*. Set*

$$c_n(u_2, \ldots, u_n, s_1, \ldots, s_n) = \Phi(s_1, \ldots, s_n) - \sum_{i=2}^{n} u_i(s_i),$$

and define inductively for $i = n-1, \ldots, 1$

$$c_i(u_2, \ldots, u_n, s_1, \ldots, s_i) = c_{i+1} \left(u_2, \ldots, u_n, s_1, \ldots, s_i, \cdot \right))^{**} (s_i).$$

Then for $i = n-1$ *to* 1,

$$c_i(u_2, \ldots, u_n, s_1, \ldots, s_i) = \inf_{\sigma^{(i)} \in [0,\infty]} \left\{ \mathbb{E} \left[c_{i+1} \left(u_2, \ldots, u_n, s_1, \ldots, s_i, X_{i+1} \right) \mid X_i = s_i \right] \right\},$$

[6] The complete code is publicly available and can be accessed through the GitHub repository https://github.com/juliansester/C-Convex

with a controlled process:

$$dX_t = \sigma_t^{(i)}\, dB_t, \quad t \in [i, i+1], \tag{4.4.1}$$

where B represents a one-dimensional Brownian Motion and $\sigma^{(i)}$ is a control process, adapted to the filtration $\mathcal{F}_t$. Then c_i is the viscosity solution of the HJB equation:

$$\min\left\{ c_{i+1} - c_i, \quad \frac{\partial^2 c_i}{\partial x^2}(x) \right\} = 0.$$

Proof. The proof follows from Section 4.2 and Corollary 4.2.1. □

For notational convenience we remove the dependence of each c_i's from $u_2, \ldots, u_n$. We reformulate the Dual Problem via the following lemma:

Lemma 4.4.2. *We have that*

$$\begin{aligned}
&\inf_{\mathbb{Q}\in\mathcal{M}(\mu_1,\ldots,\mu_n)} \mathbb{E}_{\mathbb{Q}}[\Phi(S_1, S_2, \ldots, S_n)] \\
&= \sup_{\substack{u_i \in L^1(\mu_i)\\ i=2,\ldots,n}} \mathbb{E}_{\mu_1}\Bigg[\inf_{\sigma^{(1)}\in[0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2)}\in[0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(3)}\in[0,\infty]} \mathbb{E}\Bigg[\cdots \inf_{\sigma^{(n)}\in[0,\infty]} \\
&\quad \mathbb{E}\left[c_n\left(S_1, S_1 + \int_1^2 \sigma_s^{(1)} dB_s, \ldots, S_1 + \sum_{i=1}^{n-1} \int_i^{i+1} \sigma_s^{(i)} dB_s \right) \right] \cdots \Bigg]\Bigg]\Bigg] + \sum_{i=2}^{n} \mathbb{E}_{\mu_i}[u_i(S_i)].
\end{aligned}$$

Proof. From Lemma 4.4.1,

$$c_1(s_1) = \inf_{\sigma^{(n-1)}\in[0,\infty]} \Big[\mathbb{E}\left[c_2\left(s_1, X_2\right) \mid X_1 = s_1 \right] \Big].$$

Then iteratively using the definition of c_2 we have,

$$\begin{aligned}
c_1(s_1) &= \inf_{\sigma^{(1)}\in[0,\infty]} \mathbb{E}[c_2(s_1, X_2) \mid X_1 = s_1] \\
&= \inf_{\sigma^{(1)}\in[0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2}\in[0,\infty]} \mathbb{E}\Bigg[c_3(s_1, X_2, X_3) \mid X_2 \Bigg] \mid X_1 = s_1 \Bigg] \\
&= \inf_{\sigma^{(1)}\in[0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2}\in[0,\infty]} \mathbb{E}\Bigg[c_3(s_1, X_2, X_3) \Bigg] \mid X_1 = s_1 \Bigg].
\end{aligned}$$

By recursively using the definition of c_i's for $i = 3$ to $n - 1$,

$$\begin{aligned}
c_1(s_1) = \inf_{\sigma^{(1)}\in[0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2)}\in[0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(3)}\in[0,\infty]} \mathbb{E}\Bigg[\cdots \inf_{\sigma^{(n)}\in[0,\infty]} \\
\mathbb{E}\Bigg[c_n\Bigg(s_1, X_2, \ldots, X_n \Bigg) \Bigg] \cdots \Bigg] \Bigg] \mid X_1 = s_1 \Bigg].
\end{aligned}$$

By the definition of the controlled process 4.4.1,

$$X_{i+1} = X_i + \int_i^{i+1} \sigma_s^{(i)}\, dB_s, \quad \text{for } i = 1 \text{ to } n - 1.$$

Then we have

$$\begin{aligned}
X_{i+1} &= X_i + \int_i^{i+1} \sigma_s^{(i)}\, dB_s \\
&= X_{i-1} + \int_{i-1}^{i} \sigma_s^{(i-1)}\, dB_s + \int_i^{i+1} \sigma_s^{(i)}\, dB_s \\
&\vdots \\
&= X_1 + \sum_{i=1}^{n-1} \int_i^{i+1} \sigma_s^{(i)}\, dB_s
\end{aligned} \tag{4.4.2}$$

To remove the conditional expectation, we use the alternative definition of the controlled process (4.4.2), with $X_1 = s_1$, which yields

$$c_1(s_1) = \inf_{\sigma^{(1)} \in [0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2)} \in [0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(3)} \in [0,\infty]} \mathbb{E}\Bigg[\cdots \inf_{\sigma^{(n)} \in [0,\infty]} \\ \mathbb{E}\Bigg[c_n\Bigg(s_1, s_1 + \int_1^2 \sigma_s^{(1)} dB_s, \ldots, s_1 + \sum_{i=1}^{n-1} \int_i^{i+1} \sigma_s^{(i)} dB_s \Bigg) \Bigg] \cdots \Bigg] \Bigg] \Bigg]. \tag{4.4.3}$$

Then by Proposition 4.3.1, we have the reformulated dual as stated in Lemma 4.4.2. $\square$

Chapter 5

Numerics

In this chapter, we use the reformulated dual problem as stated in Lemma 4.4.2, to construct a numerical method to solve the primal problem. In particular, our method uses neural networks to approximate the control process σ and the dual functions u. This chapter is organized as follows: we introduce the numerical method with a 2-marginal problem in Section 5.1, and show how the method can be extended to the 3-marginal case in Section 5.2.

5.1 2-Marginal Problem

The 2-marginal problem we want to solve are as follows:

$$P = \inf_{\mathcal{M}(\mu_1,\mu_2)} \mathbb{E}[\Phi(S_1, S_2)], \tag{5.1.1}$$

where $\Phi(s_1, s_2) = |s_2 - s_1|$, and the marginals μ_1, μ_2 are defined by the respective densities (where λ is the Lebesgue measure)

$$\frac{d\mu_1}{d\lambda}(s_1) = \frac{1}{2}\mathbf{1}_{[-1,1]}(s_1), \quad \frac{d\mu_2}{d\lambda}(s_1) = \frac{2+s_1}{3}\mathbf{1}_{[-2,-1]}(s_1) + \frac{1}{3}\mathbf{1}_{[-1,1]}(s_1) + \frac{2-s_1}{3}\mathbf{1}_{[1,2]}(s_1).$$

[2, Section 4.2] showed that $P = \frac{1}{3}$.

5.1.1 Neural Network Approximation

To show how neural networks may be used to solve problem (5.1.1), we first apply Lemma 4.4.2,

$$P = \sup_{u_2} \mathbb{E}_{\mu_1}\left[\inf_{\sigma^{(1)}\in[0,\infty]} \left\{ \mathbb{E}[c_2(S_1, S_1 + \int_1^2 \sigma_s^{(1)} dB_s)] \right\} \right] + \mathbb{E}_{\mu_2}[u_2(S_2)]$$
$$= \sup_{u_2} \mathbb{E}_{\mu_1}\left[\inf_{\sigma^{(1)}\in[0,\infty]} \left\{ \mathbb{E}[\Phi(S_1, S_1 + \int_1^2 \sigma_s^{(1)} dB_s) - u_2(S_1 + \int_1^2 \sigma_s^{(1)} dB_s)] \right\} \right] + \mathbb{E}_{\mu_2}[u_2(S_2)].$$

Then we assume that there exists an optimal $\widetilde{\sigma^{(1)}}$ such that

$$\inf_{\sigma^{(1)}\in[0,\infty]} \left\{ \mathbb{E}[\Phi(S_1, S_1 + \int_1^2 \sigma_s^{(1)} dB_s) - u_2(S_1 + \int_1^2 \sigma_s^{(1)} dB_s)] \right\}$$
$$= \mathbb{E}[\Phi(S_1, S_1 + \int_1^2 \tilde{\sigma}_s^{(1)} dB_s) - u_2(S_1 + \int_1^2 \tilde{\sigma}_s^{(1)} dB_s)].$$

We have

$$P = \sup_{u_2} \mathbb{E}_{\mu_1}\left[\mathbb{E}[\Phi(S_1, S_1 + \int_1^2 \tilde{\sigma}_s^{(1)} dB_s) - u_2(S_1 + \int_1^2 \tilde{\sigma}_s^{(1)} dB_s)] \right] + \mathbb{E}_{\mu_2}[u_2(S_2)]. \quad (5.1.2)$$

As mentioned earlier, we use a neural network to approximate both $\sigma^{(1)}$ and u_2, which we denote by $\tilde{\sigma}^{(1)}$ and $\tilde{u}_2$ respectively[1]. We use (5.1.2) as the loss function for both neural networks. The essence of our numerical method is to obtain the optimal solution by simultaneously minimizing and maximizing the loss function with respect to the parameters of $\tilde{\sigma}^{(1)}$ and that of $\tilde{u}_2$ respectively.

[1] We defer the discussion of the inputs to the model to Section 5.1.2

5.1.2 Loss Function

Define

$$L_1 := \Phi(S_1, S_1 + \int_1^2 \tilde{\sigma}_s^{(1)} dB_s) - \tilde{u}_2(S_1 + \int_1^2 \tilde{\sigma}_s^{(1)} dB_s),$$

and

$$L_2 := \mathbb{E}_{\mu_2}[\tilde{u}_2(S_2)].$$

In order to numerically compute the loss function (5.1.2) we use the sample mean to approximate the expectations and the Euler-Maruyama method to discretize the SDE in the loss function. That is:

$$L_1 \approx \frac{1}{N}\sum_{i=1}^{N} \frac{1}{N_p}\sum_{j=1}^{N_p} \Phi(S_1^{(i)}, S_1^{(i)} + \sum_{k=1}^{N_b} \tilde{\sigma}^{(1)} dB_{t_k}) - \tilde{u}_2(S_1^{(i)} + \sum_{k=1}^{N_b} \tilde{\sigma}^{(1)} dB_{t_k}), \tag{5.1.3}$$

and

$$L_2 \approx \frac{1}{N}\sum_{i=1}^{N} \tilde{u}(S_2^{(i)}), \tag{5.1.4}$$

where N is the number of S_1 and S_2 samples drawn from μ_1 and μ_2 respectively, N_p is the number of paths simulated for each starting point $X_1 = S_1$, N_b is the number of time steps in the interval $[1, 2)$, and $dB_{t_k} \sim \mathcal{N}(0, \sqrt{\frac{1}{N_b}})$ is the Brownian increment between each timesteps. Then the loss function that will be simultaneously minimized and maximized is $L := L_1 + L_2$.

5.1.3 Inputs of the Model

With regards to $\tilde{u}_2$, there will only be one input variable which is either $S_2^{(i)}$ or $X_2^{(i)} \approx S_1^{(i)} + \sum_{k=1}^{N_b} \tilde{\sigma}^{(1)} dB_{t_k}$. Since $\tilde{\sigma}^{(1)}$ is adapted to the filtration generated by the Brownian motion B, we let the cumulative sum of the Brownian increments dB_{t_k},

$$W_{t_{k-1}} := \sum_{i=1}^{k-1} dB_{t_i}$$

be one of the inputs to $\tilde{\sigma}^{(1)}$ and let $t_k \in [1, 2)$ to be the second input variable. The pseudocode to compute the loss is as shown in Algorithm 1.

Algorithm 1 Compute Loss

Require: batch size N, number of simulated paths N_p, number of time steps N_b, and neural networks $\tilde{\sigma}^{(1)}$ and $\tilde{u}_2$

```
Initialize momentum v_0 = 0
Sample N number of S_1 and S_2 samples respectively
for i = 1 to N do
    for j = 1 to N_p do
        for k = 1 to N_b do
            Sample dB_k^(j) from N(0, sqrt(1/N_b))
            Compute W_k^(j) = sum_{i=0}^{k-1} dB_k^(j), where dB_0^(j) = 0
        end for
        Compute path p_j^(i) = sum_{k=1}^{N_b} σ̃^(1)(t_k, W_{k-1}^(j)) dB_k^(j)
        Compute final state X_2^((i),j) = S_1^(i) + p_j^(i)
        Compute Φ_j^(i) = Φ(S_1^(i), X_2^((i),j)) − ũ_2(X_2^((i),j))
    end for
    Compute L_1^(i) = (1/N_p) sum_{j=1}^{N_p} Φ_j^(i)
    Compute L_2^(i) = ũ_2(S_2^(i))
end for
Compute L = (1/N) sum_{i=1}^{N} L_1^(i) + (1/N) sum_{i=1}^{N} L_2^(i)
```

5.1.4 Optimization

The main challenge our of our numerical method is that our problem is saddle-point problem and ensuring convergence of the loss function is non-trivial. Furthermore, the computed gradients are noisy as they are computed using sampled data rather than fixed data. An optimization method that would fit this scenario is the stochastic extragradient method (**SEG**) [12]. SEG consists of two steps: (**a**) an extrapolation step that computes a gradient update from the current iteration, and (**b**) an update step that updates the current iterate using the value of the vector field at the extrapolated point. To further reduce the variance of the loss function, momentum [11, Section 8.3.2] is used to smooth the optimization path by averaging gradients over successive iterations, thereby mitigating the effects of noisy updates. The pseudocode is given in Algorithm 2. For our case, we use equal step sizes[2] for each gradient step (i.e. $\eta_1 = \eta_2$).

Algorithm 2 Stochastic Extragradient Method with Momentum

Require: Neural network $f(\theta)$ with parameters θ_0, step sizes η_1, η_2, momentum coefficient β, number of epochs T

for $t = 0$ to $T - 1$ **do**

Sample a mini-batch of data $\mathcal{D}_t$

Compute stochastic gradient of the loss with respect to the network parameters:

$$g_t = \begin{cases} \nabla_\theta L(f(\theta_t; \mathcal{D}_t)), & \text{if minimizing the loss} \\ -\nabla_\theta L(f(\theta_t; \mathcal{D}_t)), & \text{if maximizing the loss} \end{cases}$$

Extrapolation step: Compute intermediate update:

$$\tilde{\theta}_t = \theta_t - \eta_1 g_t$$

Sample a new mini-batch of data $\mathcal{D}_{t+1}$

Compute stochastic gradient of the loss at the extrapolated parameters:

$$\tilde{g}_t = \begin{cases} \nabla_\theta L(f(\tilde{\theta}_t; \mathcal{D}_{t+1}), \mathcal{D}_{t+1}), & \text{if minimizing the loss} \\ -\nabla_\theta L(f(\tilde{\theta}_t; \mathcal{D}_{t+1}), \mathcal{D}_{t+1}), & \text{if maximizing the loss} \end{cases}$$

Update step: Perform final update using momentum:

$$v_t = \beta v_{t-1} + (1 - \beta)(-\eta_2 \tilde{g}_t)$$

$$\theta_{t+1} = \theta_t + v_t$$

end for

return final parameters θ_T

---

[2] Also known as the learning rate, η_1 can be chosen to be much larger than η_2 to emphasize more exploration of the loss landscape whilst conservatively updating the parameters.

5.1.5 Results

The loss function obtained via our algorithm is shown in Figure 5.1, with an average loss of 0.323 for the final 500 iterations.

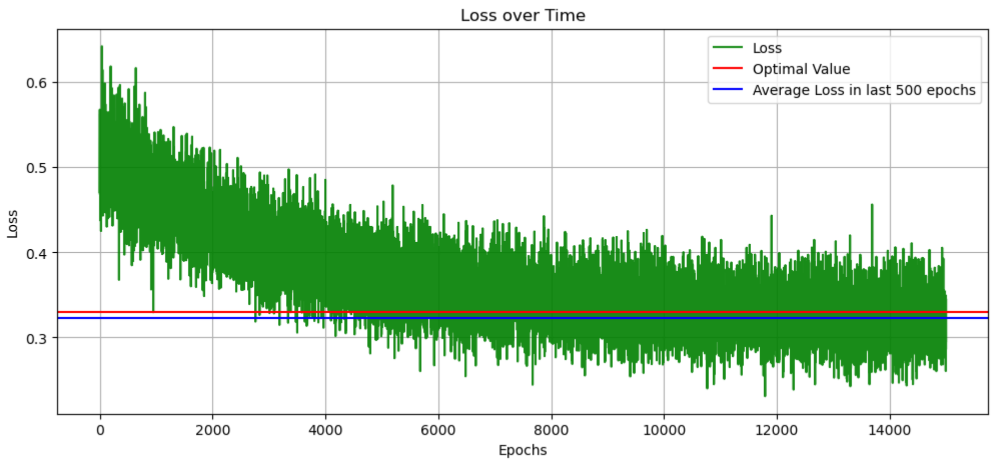

Figure 5.1: Loss vs Epochs

The aforementioned result was obtained using the following hyper-parameters (Table 5.1). In particular, a low number of simulated paths N_p is used to compute the loss

N	N_p	N_b	η	Epochs
64	2	50	0.0001	15000

Table 5.1: Detailed settings for the model training.

function. From Table 5.2[3], the variance of the loss function decreases as N_p increases. Indeed, it may be desirable to choose a higher N_p value to further reduce the variance of the loss function, though at the cost of higher computational complexity.

[3]Loss estimates are computed using the average loss of the last 500 training epochs.

N_p	**1**	**2**	**5**
Loss Estimates	0.3379	0.3234	0.345
Variance of Loss	0.000979	0.000764	0.000591
Absolute Error	0.00791	0.0099	0.015

Table 5.2: Variance of loss function for each N_p

5.2 3-Marginal Problem

We consider the 3-marginal problem from [9]:

$$\Phi(s_1, s_2, s_3) = (s_3 - \frac{1}{2}(s_2 + s_1))^+,$$

$$S_i = \exp(\sigma\sqrt{t_i} \cdot X - \sigma^2 \cdot \frac{t_i}{2}),$$

where $X \sim \mathcal{N}(0,1)$, with $\sigma = 0.25$, $t_i = i$ for $i = 1, 2, 3$. The result mentioned in [9] is

$$P \approx 0.059.$$

We will adapt the method from Section 5.1 for the 3-marginal case.

5.2.1 Loss Function

Recall that

$$P = \sup_{\substack{u_{(i)} \in L^1(\mu_i) \\ i=2,3}} \mathbb{E}_{\mu_1}[c_1(S_1)] + \sum_{i=2}^{3} \mathbb{E}_{\mu_i}[u_i(S_i)],$$

where

$$\begin{aligned} c_1(S_1) &= \inf_{\sigma^{(1)} \in [0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2)} \in [0,\infty]} \mathbb{E}\Bigg[c_3(S_1, S_1 + \int_1^2 \sigma_s^{(1)} dB_s, S_1 + \int_1^2 \sigma_s^{(1)}\, dB_s + \int_2^3 \sigma_s^{(2)}\, dB_s)\Bigg]\Bigg] \\ &= \inf_{\sigma^{(1)} \in [0,\infty]} \mathbb{E}\Bigg[\inf_{\sigma^{(2)} \in [0,\infty]} \mathbb{E}\Bigg[\Phi(S_1, S_1 + \int_1^2 \sigma_s^{(1)} dB_s, S_1 + \int_1^2 \sigma_s^{(1)}\, dB_s + \int_2^3 \sigma_s^{(2)}\, dB_s) - \\ &\qquad u_2(S_1 + \int_1^2 \sigma_s^{(1)} dB_s) - u_3(S_1 + \int_1^2 \sigma_s^{(1)}\, dB_s + \int_2^3 \sigma_s^{(2)}\, dB_s)\Bigg]\Bigg] \end{aligned}$$

for the 3-marginal case. As in the 2-marginal case, we approximate the control processes and dual functions via neural networks:

$$\tilde{\sigma}_1 := \tilde{\sigma}_1(t_k, W_{k-1}), \quad \tilde{\sigma}_1 := \tilde{\sigma}_2(t_k, W_{k-1}), \quad \tilde{u}_2, \quad \tilde{u}_3.$$

Then have the loss functions as defined below:

$$\begin{aligned} L_1 &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{N_p^{(1)}}\sum_{j=1}^{N_p^{(1)}}\frac{1}{N_p^{(2)}}\sum_{j=1}^{N_p^{(2)}}\Phi(S_1^{(i)}, X_2^{(i)}, X_3^{(i)}) - \tilde{u}_2(X_2^{(i)}) - \tilde{u}_3(X_3^{(i)}), \\ L_2 &= \frac{1}{N}\sum_{i=1}^{N}\tilde{u}_2(S_2^{(i)}) + \tilde{u}_3(S_3^{(i)}), \end{aligned}$$

where

$$X_2^{(i)} = S_1^{(i)} + \sum_{k=1}^{N_b}\tilde{\sigma}^{(1)}dB_{t_k}, \quad X_3^{(i)} = X_2^{(i)} + \sum_{k=1}^{N_b}\tilde{\sigma}^{(2)}dB_{t_k}.$$

The hyper-parameters[4] are similar to the 2-marginal case in Section 5.1.2, with the addition of one new parameter $N_p^{(2)}$, which is the number of simulated paths starting from X_2. Despite the addition of new hyper-parameters and neural networks, the optimization process follows closely to that of the 2-marginal case.

5.2.2 Results

For the 3-marginal problem, we used the same parameters as in the 2-marginal case in Table 5.1, with $N_p^{(1)} = N_p^{(2)} = 2$. The results are shown in table 5.3, and the graph of loss versus training epochs is given in Figure 5.2.

N_p	Estimate	Absolute Error	Variance of Loss
2	0.0629	0.0039	0.000440

Table 5.3: Results for 3-Marginal Problem

[4]Note that $N_p^{(1)}$ is just the number of simulated paths starting from S_1.

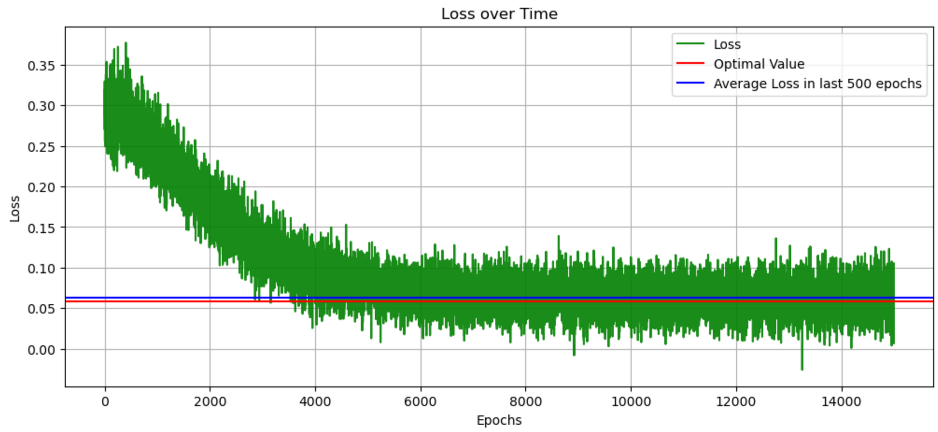

Figure 5.2: Loss vs Epochs for 3-Marginal Problem

5.2.3 Evaluation

For both the 2-marginal and 3-marginal case, the model is able to give good estimates of the optimal solution. However, as the problem is a saddle-point problem, the hyper-parameters of the method has to be tuned carefully to ensure convergence of the loss function. In both cases, the learning rates were scheduled to decay as the training progress. Figure 5.3 shows the variance of the loss function increasing after obtaining a loss close to the optimal answer of 0.059. Going from the 2-marginal to 3-marginal

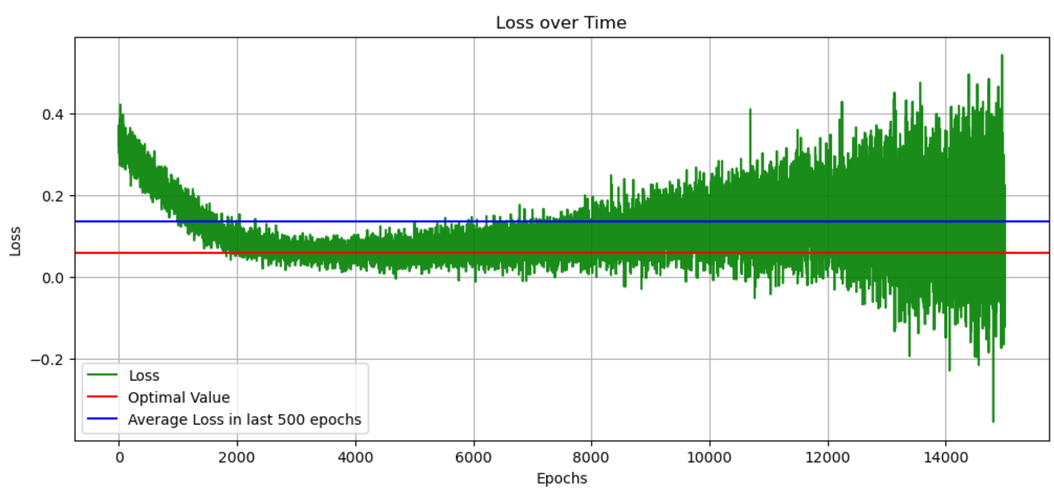

Figure 5.3: Variance of loss function increases after hitting minima.

case, the number of neural networks required increased from 2 to 4, which means that the number of trainable parameters has increased. In general, the number of neural networks required for the n-marginal case is $2(n-1)$. Furthermore, the number of simulated paths required will increase as well. For the n-marginal case, the number of simulated paths is

$$\prod_{i=1}^{n-1} N_p^{(i)}.$$

If we choose $N_p^{(i)} = N_p^{(j)}$, for $i, j = 1, 2, \ldots n-1$ then we have N_p^{n-1} number of simulated paths. The computational bottleneck would then be the exponential increase in the number of simulated paths. However, as demonstrated in Table 5.2 with the 2-marginal case, using a small N_p value, such as 1, can still yield an estimate close to the optimal value. This approach, however, comes at the cost of increased variance of the loss function. In balance, a smaller N_p value maybe used in conjunction with careful tuning of the learning rates to ensure that the variance of the loss function is not large, as demonstrated by the method for the 3-marginal problem.

Bibliography

[1] Ole E. Barndorff-Nielsen and Neil Shephard. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001.

[2] Mathias Beiglböck, Pierre Henry-Labordère, and Friedrich Penkner. Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17(3):477–501, 2013.

[3] Szymon Borak, Adam Misiorek, and Rafał Weron. *Models for heavy-tailed asset returns*, pages 21–55. Springer Berlin Heidelberg, 2011.

[4] Douglas T. Breeden and Robert H. Litzenberger. Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–651, 1978.

[5] Ishii-H. Lions P. L. Crandall, M.G. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American Mathematical Society*, 27(1):1–67, 1992.

[6] Pierre-Louis Crandall, Michael G.; Lions. Viscosity solutions of hamilton-jacobi equations. *Transactions of the American Mathematical Society*, 277:1–42, 1983.

[7] Bruno Dupire. Pricing with a smile. *Risk*, 7(1):18–20, 1994.

[8] Stephan Eckstein and Michael Kupper. Computation of optimal transport and related hedging problems via penalization and neural networks. *Applied Mathematics & Optimization*, 83:639–667, 2021.

[9] Stephan Eckstein and Michael Kupper. Martingale transport with homogeneous stock movements. *Quantitative Finance*, 21(2):271–280, 2021.

[10] Myron Scholes Fischer Black. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[11] Ian J. Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[12] Eduard A. Gorbunov, Hugo Berard, Gauthier Gidel, and Nicolas Loizou. Stochastic extragradient: General analysis and improved rates. In *International Conference on Artificial Intelligence and Statistics*, 2021.

[13] GaoYue Guo and Jan Obłój. Computational methods for martingale optimal transport problems. *The Annals of Applied Probability*, 29(6):3311–3347, 2019.

[14] Pierre Henry-Labordère. *Model-free hedging: A martingale optimal transport viewpoint*. CRC Press, 2017.

[15] Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The review of financial studies*, 6(2):327–343, 1993.

[16] David G Hobson. Robust heding of the lookback option. *Finance and Stochastics*, 2(4):329–347, 1998.

[17] Zuzana Janková. Drawbacks and limitations of black-scholes model for options pricing. *Journal of Financial Studies and Research*, pages 1–7, 2018.

[18] Xun Yu Zhou Jiongmin Yong. *Stochastic Controls - Hamiltonian Systems and HJB Equations*. Springer New York, NY, 1999.

[19] Mark Kac. On distributions of certain weiner functionals. *Transactions of the American Mathematical Society*, 65:1–13, 1949.

[20] Leonid Vitalyevich Kantorovich. On the translocation of masses. *Dokl. Akad. Nauk. USSR (NS)*, 37:199–201, 1942.

[21] Hans G. Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 67(4):399–432, 1984.

[22] Steven G. Kou. A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101, 2002.

[23] Gaspard Monge. M´emoire sur la th´ eorie des d´ eblais et des remblais. *Histoire de l'Acad´ emie Royale des Sciences de Paris*, 1781.

[24] Huyên Pham. On some recent aspects of stochastic control and their applications. *Probability Surveys [electronic only]*, 2:506–549, 2005.

[25] Wim Schoutens. *Lévy processes in finance*. Wiley, 2003.

[26] Julian Sester. A multi-marginal c-convex duality theorem for martingale optimal transport. *Statistics Probability Letters*, 210, 2024.

[27] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[28] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.