Monotone Optimal Policies for Quasivariational Inequalities Arising in Optimal Portfolio Liquidation

by

Daniel J. Crawford

B.A.Sc. Engineering Physics, University of British Columbia, 2010

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Electrical and Computer Engineering)

The University of British Columbia
(Vancouver)

December 2014

Abstract

This thesis studies the Hamilton-Jacobi-Ballman quasivariational inequality (HJBQVI), the corresponding optimal value function, and discrete schemes useful for approximating this value function. Moreover, the structural properties of the optimal policy of particular discrete scheme is studied. The motivation is to find a convergent, approximating scheme for the otherwise complicated HJBQVI that has monotone policy structure that can be exploited in a stochastic gradient estimation scheme to approximate optimal policy function parameters. In order to motivate this approach, we consider the problem of optimal liquidation of a single risky asset portfolio as an impulse control problem. The model is defined over continuous time, state, and compact action sets, and the optimal liquidation value and strategy are found from the viscosity solution of a HJBQVI. It is shown that the optimal strategy is monotone in the number of shares owned and the time remaining to liquidation. This structural result is exploited to estimate the optimal policy via a reinforcement learning method based on the simultaneous perturbation stochastic approximation (SPSA) algorithm. The optimal policy can be estimated without knowledge of the parameters of the underlying model.

Preface

This dissertation is original, unpublished, independent work by the author, D. Crawford. Portions of this thesis which use notions from previous work have been indicated with rigorous citation.

Works Submitted for Publication:

D. Crawford, V. Krishnamurthy, "Monotone Optimal Policies for Quasivariational Inequalities Arising in Optimal Portfolio Liquidation".

D. Crawford, V. Krishnamurthy, "Monotone Optimal Policies in Portfolio Liquidation Problems".

Table of Contents

List of Tables

List of Figures

Glossary

HJB	Hamilton-Jacobi-Bellman
QVI	Quasivariational Inequality
HJBQVI	Hamilton-Jacobi-Bellman Quasivariational Inequality
MDP	Markov Decision Process
SPSA	Simultaneous Perturbation Stochastic Approximation
USC	Upper Semi-continuous
LSC	Lower Semi-continuous
WLOG	Without Loss of Generality

Acknowledgments

Special thanks to Ann, Murray, Mark, Peeti, Steph, Bridget, Woody and all of my close friends for the endless support during my degree.

Thank you to all of my laboratory colleagues for numerous insightful talks about mathematics and engineering, and for the less insightful - and equally rewarding - talks about the stresses of studying mathematics and engineering.

A very sincere thank you to my supervisor Dr. Vikram Krishnamurthy - for pushing me to excel in a difficult area of research, and for always being available to discuss the direction of my research.

Dedication

To my grandparents.

Chapter 1

Introduction

In financial frameworks, situations arise where an agent must convert a position in a risky asset into cash. This could be due to sudden unexpected expenses, disinterest in the market, or any foreseeable reason why having a position in the asset is unwanted by the asset holder. Some popular, yet naïve strategies for liquidating this position is to immediately sell the position, uniformly sell portions of shares up to a liquidation deadline, or wait until a liquidation deadline to sell all shares. However, given this liquidation deadline and some a priori knowledge of the price dynamics, the decision maker is given the opportunity to create an optimized strategy in order to maximize the total profit from the process.

This thesis studies optimal liquidation strategies of a single risky asset as an impulse control problem. This problem can be formulated as a Hamilton-Jacobi-Bellman quasivariational inequality (HJBQVI), a continuous time, state, and action optimal stochastic control tool which uses impulse control tools developed by Bensoussan and Lions, 1984 [1] combined with continuous dynamic programming methods. The HJBQVI has since been used in numerous applications, for example in physics, engineering, automatic control and finance. This thesis follows the application of the HJBQVI in optimal impulse control liquidation strategies developed in [2]. In this model, the price of the asset is assumed to follow a geometric Brownian motion and no risk free asset such as a guaranteed interest savings account or optional bond holding is permitted. The model discussed in [2] is extended in this work to include market impact incurred through the selling actions of the agent (see also [3], [4], [5], [6], [7], [2], and [8], although the latter considers a more general geometric Brownian motion whose drift coefficient is a one jump Markov process rather

than constant). The market impact representation used in this work is adapted from [9], and is considered a non-decaying, permanent price impact. Refer to [5], [10] for discussion on price impact (temporary, delayed, permanent). The impact term is included to assist in the structural proofs contained in the following sections. After describing the impulse control HJBQVI, it is shown that the optimal value function arises from an optimal liquidation strategy which is a monotone function function of the number of shares of the asset owned, the price of the asset, and the amount of time remaining until the liquidation horizon. The arguments leading to this conclusion hang on the assumption that the immediate and terminus investor actions give rise to a reward function that is concave in the sell action, and non-decreasing in the asset price. It is shown that this structural result can be derived from a corresponding discretized Markov Decision Process (MDP) (see [11], [12] for background). Further, it is shown in the limiting case of the discretization in time, these properties continue to hold, therefore showing a structural result for the optimal policy of the HJBQVI. Once this method for proving monotone structures of HJBQVI impulse problems is established, it can be adapted for a variety of other problems not necessarily in a financial framework.

The goal of this work is to show that the inherently complicated HJBQVI decision model can be probed to yield a policy structure result. When we have this result, the study of this mathematical construct can be adjusted from the classical optimal control treatment and instead studied from a more modern engineering point of view in the form of applying reinforcement learning techniques.

1.1 Related Work

For a development of risk theory and decision making in finance, see [13] for a basic overview, and [1], [14] for in-depth treatments of optimal stopping time and optimal impulse control processes. Since the primary interest is in analyzing the structure of the optimal policy, our model assumes no bid-ask spread, and no trading frequency penalty. For work closely considering these factors see [2], [3], [15], [16], and [5]. There has been a push to work on more general stochastic processes to model the price of financial instruments, most generally through jump-diffusion processes ([17], [18], [19], [15], [20], [21], [22], [23]) where prices follow discontinuous paths with jumps modeled by Lévy processes such as a compound Poisson process. The ideas of this thesis do not immediately consider jump discontinuities in price, but

the ideas covered here with permanent price impact extend directly to these more general cases as the jump intensities typically do no depend on the investing agent's actions. For an excellent reference on reinforcement learning algorithms, stochastic search, stochastic optimization, stochastic gradient methods, and especially the simultaneous perturbation stochastic approximation algorithm applied in this work, see [24].

1.2 Main Results and Organization

A general background is given in Sec. 2. It is split into four sections. Sec 2.1 introduces the discrete time, space, and action Markov Decision Process (MDP). MDPs are a mathematical tool used to model decision making in situations where outcomes are partially random and partially under the control of an agent. Moreover, the underlying state can also be stochastic. In this sense, MDPs are discrete time, state, and action stochastic control processes. An MDP is used in this work as an approximating tool to its more complex continuous time, state, and action cousin, which, as mentioned above, is formulated as the HJBQVI. This more complicated model is explained in both background and derivation (although not in great depth, as this could take up the majority of this thesis) in Sec. 2.2. In impulse control problems, there are often jumps in the value function solution to the HJBQVI. For this, the concept of viscosity solutions is introduced in Sec. 2.3, and is used later in the convergence proofs when comparing value function of the MDP to the value function of the HJBQVI. Finally, since the major contribution of this work is to incorporate modern engineering reinforcement learning techniques into the complex mathematical discipline of optimal impulse control, Sec. 2.4 introduces the concept of reinforcement learning, it's applications, and various algorithms which are in prominent use today.

In Sec.3, the continuous-time portfolio liquidation problem is formulated in finite-continuous time. The portfolio model contains a three dimensional state space over a positive real cube and includes the price of the asset, the cash in pocket, and the quantity of shares of the asset. The optimal portfolio liquidation value function is shown to satisfy the HJBQVI. The setup closely follows [2] in order to construct the dynamic programming principle. The main result of the work is stated: the value function of the HJBQVI can be approximated by the value function of a denumerable MDP whose optimal policy is monotone in price, shares owned, and time

until liquidation.

Sec. 4 aims to prove the main result. This is done in three steps: A countable state,action, and time MDP is formulated on a discretized grid of the problem variables from Sec. 3. The value function of this MDP is then shown to converge to the value function of the HJBQVI in the limiting case. The value function convergence ultimately uses a viscosity solution argument (see Sec. 2.4), which is in itself an approximating argument. However, we argue that this approximating error is negligible as it is extensively used in the literature. The structure of the optimal liquidation policy of the discrete problem is analyzed and it is shown that there exists monotone characteristics of the optimal liquidation action in time until liquidation, the price of the asset, and the quantity of shares owned. Proof of this structural result is given and follows from concavity of the value function with respect to the asset price and shares owned variables, and the treatment of the HJBQVI as a sequential allocation process. Given this result, combined with the convergence results above, this shows that the HJBQVI has a value function that is equal to the value function under convergence of a discrete MDP with this monotonic action characteristics. We maintain that this is a new result and may be extended to more general decision processes in future work.

In Sec. 5, numerical simulations are provided when assuming that the price evolution distribution is known. A finite difference scheme is introduced along with an iterative method for numerically estimating the optimal value function and optimal liquidation strategy of the HJBQVI. Simulation parameters are chosen to emulate real world scenarios and resemble those used in [2]. The approximating countable state and action MDP is modeled using the same parameters and the results are shown to be consistent with the finite difference approximation. When the price evolution distribution is not known and therefore transition probabilities of the MDP formulation are unknown, the mentioned machine learning approach is used, exploiting the monotone structure results from Sec. 4. More specifically, the reinforcement learning method of simultaneous perturbation stochastic approximation (SPSA) is used to estimate the optimal monotone strategy in this case.

Chapter 2

Background

This introductory section begins by giving an overview of Markov Decision Processes (MDPs) used traditionally in finite state, action, and discrete time optimal control and decision making problems. These mathematical models are heavily used in engineering, robotics, automated control, manufacturing (for instance) when the underlying process which a user bases descisions on is stochastic in nature, and when outcomes are partially random. Following this, a continuous state, action, and time optimal constrol/decision model is introduced in the form of a Hamilton-Jacobi-Bellman quasivariational inequality (HJBQVI). This is a somewhat complex formulation (compared to the previous MDP model) in that we now have to use stochastic calculus, uniqueness/existence theorems, as well as the notion of viscosity solutions to derive optimal control strategies and value functions. Finally, the concept of reinforcement learning algorithms is introduced. These are the types of algorithms the author has been interested in applying to the otherwise complex HJBQVI problem, in an attempt to approximate the optimal control policy when not all paremeters of the underlying model are known, or if the state, time, and/or action discretizations are simply too fine for using classical MDP approximations.

2.1 Discrete Time, State, and Action Optimal Control

This section is adapted from [12] and defines the basic structure of finite horizon Markov Decision Processes in countable state, action, and time. MDPs are models for sequential decision making when the controlled system involves uncertainty. There are five components which describe a MDP:

i. Decision epochs: an increasing time variable, e.g. $n = 0, 1, 2, ...$ in the infinite horizon case or $m \in \mathbb{N} \cap [0, T]$ for some $T < \infty$ in the finite horizon case.

ii. System states: a stochastic process which represents the underlying system state. This could be the condition of some machinery, the price of a stock on a financial exchange, etc. Typically states are given a symbolic ordering s.t. $S = \{s_i\}_{i \in \mathbb{N}}$ where each symbol is understood by the decision maker.

iii. Actions/decisions: a set of functions that constitute a decision maker policy. The application of these action functions is how the decision maker affects the sytem in order to maximize some reward function or minimize some cost function.

iv. Immediate costs/rewards and terminal cost/rewards: These are functions mapping from state and action to some value function which is assessed by the decision maker to be "good" or "bad", depending on the desired outcome of the control process.

v. State Transition Probability Functions: in the MDP formulation, it is assumed that the state can transition according to a Markov Chain with underlying (known or unknown) transition kernels. Moreover, we assume that the state transitions are also affected by the decision maker's actions, such that certain transitions could be considered more or less favorable for the system depending on the user's external input.

The goal of a decision maker is to use a MDP combined with a dynamic programming algorithm to decide which actions are optimal at each decision epoch, in any system state, in order to maximize or minimize some value criterion. In the case of this work, decision epochs $1, ..., N$ are considered, with $N < \infty$. Let $\pi = (d_1, d_2, ..., d_{N-1}), d_i \in \mathcal{T} \subset \mathbb{R}^n$ be a randomized policy for each decision epoch $n = 1, ..., N-1$, where d_i denotes the control decision applied when in epoch i. Let the admissible actions be $\mathcal{D}$. Let the state of the system be donoted $\{X_n\}_{n=1,...,N}$ take values in some set $\mathcal{S} \subset \mathbb{R}^n$. Let $c_n(X_n, d_n) \in \mathbb{R}$ be the immediate cost of applying action d_n when the system is in state X_n at epoch n. Let $C_N(X_N)$ be the terminal cost when the controlled system is left in state X_N at the final epoch. Finally, assume the system is in state $X_i = x \in \mathcal{S}$ at epoch i and that some action $d_i = d \in \mathbb{R}^n$ is applied. At the next epoch, the system will transition to another state $X_j = x' \in \mathcal{S}$ probabilistically with the transition kernel given by $\mathbb{P}(X_j = x' | X_i = x, d_i = d) = \mathbb{P}(x', x, d)$. Given

these components, a total expected cost function can be formulated for policy π for some initial state $x \in \mathcal{S}$ s.t.

$$v^{\pi}(x) = \mathbb{E}\left\{\sum_{n=1}^{N} c(X_n, d_n) + C(X_N)|X_0 = x, \pi\right\}. \tag{2.1}$$

The optimaztion problem is to find $\pi^* = (d_1^*, ..., d_{N-1}^*), d_i^* \in \mathbb{R}^n$ such that

$$v^{\pi^*}(x) = v(x) := \sup_{\pi \in \mathcal{T}} v^{\pi}(x), \tag{2.2}$$

for an initial state $x \in \mathcal{S}$. Given $\pi = \{d_1, ...d_{N-1}\} \in \mathcal{T}$, define the possible actions at each epoch by $d_i \in \mathcal{D}$ This value function, and the coinciding optimal policy π^* have been found to be the solution of the following stochastic dynamic programming recursion called Bellman's equation [12]

$$v_n(x) = \max_{a \in \mathcal{D}} Q_n(x, a), \tag{2.3}$$

$$d_n^*(x) = \arg\max_{a \in \mathcal{D}} Q_n(x, a), \tag{2.4}$$

where

$$Q_n(x, a) = c_n(x, a) + \sum_{x' \in \mathcal{S}} \mathbb{P}(x', x, a) v_{n+1}(x'), \tag{2.5}$$

and $v_N(x) = C_N(x)$.

Since this work is aimed to probe the structural properties of the decision process, the monotonic structure of an optimal policy is described. A policy π^* is monotonic in a certain variable if policy elements d_i^* are nonincreasing (or nondecreasing) in that variable, for all $i \in \{1, ..., N-1\}$. For example, π^* is monotonically nonincreasing in x if $d_n^*(x) \leq d_n^*(x + \delta)$ for some $\delta \geq 0$.

There are two methods for proving this structural result in this work so far: one uses a result in a sequential allocation framework, and the other uses super- and submodularity arguments. For the latter, define supermodularity as follows:

Definition 2.1.1. *A function $F(x, y) : X \times Y \to \mathbb{R}$ is supermodular in (x, y) if $F(x_1, y_1) + F(x_2, y_2) \geq F(x_1, y_2) + F(x_2, y_1) \quad \forall x_1, x_2 \in X$ and $y_1, y_2 \in Y$ with $x_1 > x_2, y_1 > y_2$. If the inequality is reveresed, the function $F(\cdot, \cdot)$ is called submodular.*

As will be shown in the Work Completed section, the methodology used to prove the monotonic structure of a policy is done in two steps:

1. The monotonicity of the optimal value function $v_n(x)$ is shown.

2. The state-action reward function $Q_n(\cdot,\cdot)$ define in (2.5) is shown to be supermodular in (x,n), and when combined with 1) yields the monotonic structure of the optimal policy.

This structural result is shown for a MDP which simulates the Hamilton-Jacobi-Bellman quasivariational inequality (discussed in the next section). A MDP that simulates a HJBQVI is a MDP whose value function converges to the value function of the HJBQVI, but not necessarily the policy. In essense the summary of this work is that given a HJBQVI, we can formulate a MDP with specific optimal policy structure that converges in value to the original problem. This structure can then be exploited in numerical schemes by utilizing reinforcement learning techniques such as the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm, giving a machine learning approach to solving the HJBQVI when certain state properties are unknown.

2.2 Continuous Time, State, and Action Optimal Control

This work utilizes the notion of a quasivariational inequality (QVI) originally described by [1] in 1982, but since has been used in many areas including finance, engineering, and operations research. The formal framework for QVIs is given as follows:

Definition 2.2.1. *(Formal QVI Framework)*
Let $\mathcal{A}$ denote a second order elliptical or parabolic differential operator on a set $\mathcal{O} \subset \mathbb{R}^n$ or in $\mathcal{Q} = \mathcal{O} \times]0,T[$. Consider the following non-linear operator on a measurable function v:

$$v \to M(v) \tag{2.6}$$

with the following increasing property

$$v_1 \leq v_2 \Rightarrow M(v_1) \leq M(v_2) \tag{2.7}$$

The goal is to find a function $u(x,t), x \in \mathcal{O}, t \in]0,T[$ *such that*

$$\begin{cases} \mathcal{A}u - f \leq 0, \quad u - M(u) \leq 0, \\ (\mathcal{A}u - f)(u - M(u)) = 0 \quad \text{in } \mathcal{O}\times]0,T[, \end{cases} \tag{QVI}$$

and the addition of limit conditions, and if $\mathcal{A}$ *is evolutionary, initial conditions.*

Next, given the formal definition of this mathematical tool, it is shown how we can utilize the QVI in an impulse control problem in a dynamical system. Consider the state of some system of interest given by the stochastic differential equation (SDE)

$$dX_t = g(X_t,t)dt + \sigma(X_t,t)dB_t, \tag{2.8}$$

where $g(\cdot,\cdot)$ is called the *drift* of process X and $\sigma(\cdot,\cdot)$ is called the *diffusion* of process X. It is assumed that the coefficients $b(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ are Lipschitz, such that $\exists K > 0$ s.t.

$$|\sigma(t,x) - \sigma(t,y)| \leq K|x-y|, \qquad |b(t,x) - b(t,y)| \leq K|x-y|,$$

for all $t \geq 0$ and $x, y \in \mathcal{O}$. If the Lipschitz condition holds, and for each constant $R > 0$ there is some C_R such that

$$|\sigma(s,0)| + |b(s,0)| \leq C_R$$

for all $s \leq R$, then SDE (2.8) is an exact SDE and $\exists!$ solution $X_t : [0,T] \times \mathcal{O} \mapsto \mathbb{R}^n$. To control process X with an *impulse control*, it is assumed that at certain *impulse instances* the state undergoes jumps refered to as *impulses*. The decision process is then the collection of pairs of impulse instances and impulse intensities, called an impulse control. Denote the impulse control strategy as $\alpha = \{\tau_n, \zeta_n\}_{n=0,1,\ldots}$, and given a finite horizon $T < 0$,

$$\mathbb{P}\Big(\lim_{n\to\infty} \tau_n \geq T\Big) = 0. \tag{2.9}$$

The state dynamics under impulse strategy α becomes

$$\begin{aligned} dX_t &= g(X_t,t)dt + \sigma(X_t,t)dB_t, \quad \text{for } \tau_i \le t < \tau_{i+1}, \\ X_{\tau_i} &= X_{\tau_i^-} + \zeta_i, \\ X_0 &= x, \end{aligned}$$

where X_{τ_i} is given by

$$X_{\tau_i} = X_{\tau_{i-1}} + \int_{\tau_{i-1}}^{\tau_i} g(X_s,s)ds + \int_{\tau_{i-1}}^{\tau_i} \sigma(X_s,s)dB_s.$$

The goal is to maximize some economic function that is scenario specific. The economic function is tailored by specifying rewards dependent on the impulse control and system state throughout the problem time interval $[0,T]$, with the option of also specifying a terminal reward dependent on the state and final action. In general the economic criterion for impulse control is written

$$J^\alpha(x) = \mathbb{E}\left[\int_0^T f(X_s,s)ds + \sum_{i:\tau_i \le T} c(\zeta_i, X_{\tau_i})\mathrm{e}^{-\beta\tau_i} + C_T(X_T)\right], \tag{2.10}$$

where $f(\cdot,\cdot)$ is the continuous running cost of the problem, $c(\cdot,\cdot)$ is the cost/reward of a single impulse action, $\beta > 0$ is some time value discount factor, and $C_T(\cdot)$ is the terminal reward/cost dependent on the final state of the system. The decision maker wants to either minimize (in the case of $c(\cdot,\cdot)$ being a detriment), or maximize (in the case of $c(\cdot,\cdot)$ being favorable) the economic function J s.t.

$$v = \sup_{\alpha\in\mathbb{A}} J^\alpha, \tag{2.11}$$

$$\alpha^* = \arg\sup_{\alpha\in\mathbb{A}} J^\alpha, \tag{2.12}$$

where $\mathbb{A}$ is a system dependent set of admissible impulse actions which the decision maker defines to be allowable at each point in $\mathcal{O}\times[0,T]$, v is the optimal value function (called just the value function), and α^* is the optimal impulse strategy. The goal of this research is to study this optimal impulse strategy through various discretizations and convergences, and say that the value function v corresponds to a discrete MDP that has an optimal policy which has an exploitable structure. More

details about this structure is described in the next section, outlining the work done thus far. To relate this maximization/minimization problem to (QVI), define a dynamic programming approach to finding the value function v. [Bensoussan and Lions, 1982] show that under certain assumptions the value function v is the unique solution to the quasivariational inequality

$$\begin{cases} \langle \mathcal{A}v, u-v\rangle \geq \langle f, u-v\rangle, \quad \forall v : \mathcal{O}\times[0,T], \quad u \leq \mathcal{M}u \\ 0 \leq v \leq \mathcal{M}v, \end{cases} \tag{2.13}$$

where $\mathcal{M}u(t,x) : \mathcal{O}\times[0,T] \mapsto \mathcal{O}\times[0,T]$ is an intervention operator defined as

$$\mathcal{M}u(t,x) = k + \inf_{(\tau_i,\zeta_i)\in\mathbb{A}, x+\zeta\in\mathcal{O}} u(t^-, x+\zeta) + c(\zeta_i, X_{\tau_i}), \tag{2.14}$$

where $k \geq 0$ is a fixed impulse cost, and $c(\cdot,\cdot)$ is the immediate reward/cost as described above. The operator $\mathcal{A}u$ can be thought of as an infinitessimal, no-impulse, continuation regime, in which the state evolves without interruption according to the state SDE. Therefore we define $\mathcal{A}$ as follows

$$\mathcal{A}u(x) = \lim_{t\downarrow 0} \frac{\mathbb{E}^{x,y,p}[u(t,X_t)] - u(t,x)}{t},$$

and for the general SDE described in (2.8), for some function $f \in C^2(\mathbb{R}^n \times [0,T])$, by an application of Itô's formula

$$\mathcal{A}f(t,x) = \sum_i g_i(t,x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j}(\sigma\sigma^T)_{i,j}(t,x)\frac{\partial^2 f}{\partial x_i \partial x_j}. \tag{2.15}$$

Given these control systems definiton, many authors have shown that the value function v is the unique viscosity solution the the following

$$\max\left\{\frac{\partial v}{\partial t} - \mathcal{A}v, \mathcal{M}v - v\right\} = 0 \quad \text{ on } [0,T)\times\mathcal{O}, \tag{2.16}$$

together with the terminal condition

$$\max\{v - C_T(X_T), \mathcal{M}v - v\} = 0 \quad \text{ on } \{T\}\times\mathcal{O}. \tag{2.17}$$

Equations (2.16) and (2.17) are referred to as the Hamilton-Jacobi-Bellman quasi-variational inequality (HJBQVI) for the impulse control problem (2.11) and (2.12).

2.3 Viscosity Solutions

The main result of this work relies on the concept of *viscosity solutions*, a topic thoroughly discussed in [25]. Viscosity solutions are used when uniqueness, existence, and stability results are sought after for partial differential equations where singularities, discontinuities or regions of non-differentiability may occur. Further, we are considering applications of impulse control strategies with the Hamilton-Jacobi-Bellman PDE which is inherently discontinuous in regions of impulse. The formal definition of a viscosity solution, a *weak solution* to a PDE is as follows. Take the equation

$$F(x, u(x), \tfrac{\partial u}{\partial x}, \nabla u(x)) = 0 \tag{2.18}$$

over some domain Ω. Here, $\nabla u(x)$ is the Hessian matrix. $F(\cdot,\cdot,\cdot)$ is called degenerate elliptic if for any two symmetric matrices X and Y, $X - Y$ is positive definite, and $\forall x \in \Omega$ and $p \in \mathbb{R}^n$,

$$F(x, u(x), p, X) \geq F(x, u(x), p, Y). \tag{2.19}$$

For a given upper semicontinuous function $\bar{u}(x) : \Omega \mapsto \mathbb{R}^n$, $u(\cdot)$ is called the *viscosity subsolution* if at any point $\hat{x} \in \Omega$, and for any C^2 function $\phi(\cdot) : \Omega \mapsto \mathbb{R}^n$ s.t. $\phi(\hat{x}) = u(\hat{x})$ and $\phi(x) \geq u(x)$ for $x \in B(\epsilon, \hat{x})$, a ball of arbitrary radius $\epsilon > 0$ around point $\hat{x}$, the following holds

$$F(\hat{X}, \phi(\hat{x}), \tfrac{\partial \phi(\hat{x})}{\partial x}, \nabla \phi(\hat{x})) \leq 0. \tag{2.20}$$

Conversely, given a lower semicontinuous function $\underline{u}(x) : \Omega \mapsto \mathbb{R}^n$, $u(\cdot)$ is called the *viscosity supersolution* if at any point $\hat{x} \in \Omega$, and for any C^2 function $\phi(\cdot) : \Omega \mapsto \mathbb{R}^n$ s.t. $\phi(\hat{x}) = u(\hat{x})$ and $\phi(x) \leq u(x)$ for $x \in B(\epsilon, \hat{x})$, a ball of arbitrary radius $\epsilon > 0$ around point $\hat{x}$, the following holds

$$F(\hat{X}, \phi(\hat{x}), \tfrac{\partial \phi(\hat{x})}{\partial x}, \nabla \phi(\hat{x})) \geq 0. \tag{2.21}$$

Finally, a continuous function $u(x) : \Omega \mapsto \mathbb{R}^n$ is a *viscosity solution* of the degenerate elliptical equation (2.18) if $u(\cdot)$ is both a viscosity supersolution and a viscosity subsolution. In the next sections, the concept of viscosity solutions to the derived HJBQVI will be exploited to show an approximate convergence result of

the discretization scheme used to show the monotone structure of the "ϵ"-optimal liquidation policy, where "ϵ"-optimal indicates the optimal policy found from the viscosity-convergent value function of the corresponding MDP.

2.4 Reinforcement Learning

The ultimate goal of this work is to apply a reinforcement learning technique to the optimal liquidation impulse control problem, an approach that has yet to be attempted for an impulse control represented by the HJBQVI. This section gives a broad defintion of reinforcement learning, its application, and the use of the SPSA algorithm for reasons which will be pointed out in the following.

Reinforcement learning is an area of machine learning intent on learning how to take action/control in a given environment, in such a way as to maximize or minimize some given reward or cost structure. When compared to dynamic programming, where a decision making agent has knowledge of the underlying (possibly stochastic) system process, a reinforcement learning technique is considered *approximate dynamic programming*, and is a situation where either the state space is too large, or the system dynamics are not completely known.

The SPSA algorithm is an algorithm belonging to the reinforcement learning class. This overview is adapted from [26] and [24]. SPSA is a method for optimizing stochastic processes when one or more of the underlying processes are governed by unknown parameters. First and foremost, SPSA is an algorithm capable of global maxima or minima to optimization problems of the form

$$d^* = \arg\max_{d \in \mathcal{D}} v(d), \tag{2.22}$$

where $v(\cdot)$ is some cummulative cost or reward functional dependent on parameter d. Notice how this formulation compares to the optimal control defined in Sec. 2.1. The algorithm uses an iterative scheme as follows:

$$d_{n+1} = d_n - \gamma_n \nabla v(d_n), \tag{2.23}$$

where γ_n is a sequence of positive values eventually converging to zero, $d_{n+1} = ((d_{n+1})_1, ..., (d_{n+1})_k)$ is a k-dimensional vector which could, for example, represent an action process. If we define a random perturbation vector Δ_n and a scaling value c_n, the SPSA algorithm estimates the gradient of the i-the entry of cummulative

value function via

$$\nabla v(d_n)_i = \frac{v(d_n + c_n\Delta_n) - v(d_n - c_n\Delta_n)}{2c_n(\Delta_n)_i}, \tag{2.24}$$

which means we only need to simulate two ranom variables per entry of the value function. Comparing this to classical estimation algorithms such as the Kiefer-Wolfowitz algorithm, which requires k simulations per value function entry. As the state and action spaces of the MDP (Sec. 2.1) uses small enough discretizations in order to approximate the HJBQVI value function (Sec. 2.2), it is easy to see that the SPSA algorithm remains tractable while other stochastic gradient estimation algorithms such as Kiefer-Wolfowitz become impractical. This is the motivation behind using SPSA, and more details involving convergence and performance can be read in the cited reference and are not discussed in this thesis. In the numerical results portion (Sec. 5) the algorithm is written out explicitly to give a more sophisticated look at the methodology.

This concludes the brief introduction of MDPs, the Hamilton-Jacobi-Bellman quasivariational inequality, viscosity solutions, and the SPSA reinforcement learning / stochastic gradient estimation algorithm. The next section beings the main work by the author on the thesis, and In the following section a detailed financial application is developed and explored, and the main result of this work is described.

Chapter 3

Finite Horizon Portfolio Liquidation Model

All processes are defined on a standard probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Assume the price of a single risky asset evolves according to a geometric brownian motion studied in, for example, [6],[2],[27], over the time interval $t \in [0,T]$, where $T > 0$ is a terminal portfolio liquidation horizon. Define a sigma algebra generated by a one-dimensional Brownian motion B_t as $\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t)$. Consider an impulse control strategy $\alpha = \{\tau_n, \zeta_n\}_{n=0,1,\ldots}$, with each τ_n an $\mathcal{F}_t$-stopping time and ζ_n an $\mathcal{F}_{\tau_n}$-measurable random variable, where $\mathcal{F}_{\tau_n} = \sigma(B_s : 0 \leq s \leq \tau_n)$. Each τ_n indicates a trading impulse instant, and each ζ_n indicates a trade amount. In this work only sell actions are considered as the agent is trying to liquidate a position and therefore $\zeta_n > 0$ for all n. Given an impulse control strategy α, the impacted price process we use to study the liquidation problem is written as a controlled scalar-valued price process P_t s.t.

$$dP_t = \mu(P_t)dt + \sigma(P_t)dB_t - \sum_{\tau_i \leq T} m(\zeta_i), \quad P_0 = p, \tag{3.1}$$

where B_t is a standard one-dimensional Brownian motion, $p > 0$ is a given constant. The market impact function $m(\cdot)$ is a bounded function described in [9]. Assume the following about the price impact function

Assumption 3.0.1. *(Market Impact)*

1. $m(\zeta)$ *is a nondecreasing function of trading action* $\zeta > 0$
2. $m(\zeta_1) \geq m(\zeta_2)$ *for all* $\zeta_1 \geq \zeta_2 > 0$

3. $m(0^+) = 0$

4. $m(\zeta) \geq 0 \quad \forall \zeta \geq 0$

5. *The application of any market impact term cannot exceed the asset price s.t.* (3.1) *is always non-negative*

Well-definedness of (3.1) follows from the coefficients of (3.1) meeting the standard Lipschitz and linear growth conditions and the price impact function $m(\zeta_i), i = 0, 1, ...$ is bounded for all ζ_i.
For now, given any impulse instant $\tau_n \in [0,T]$, assume the following simple instantaneous reward structure

$$c(X_{\tau_n^-}, Y_{\tau_n^-}, P_{\tau_n^-}, \zeta_n) = c(P_{\tau_n^-}, \zeta_n). \tag{3.2}$$

In general, take the following assumptions on $c(\cdot,\cdot)$

Assumption 3.0.2. *(Instantaneous reward function.)*

1. $\inf_{\zeta \in \mathbb{R}_+} c(p,\zeta) \sim L\zeta > 0$, *for some* $L > 0$,

2. $c(\cdot,\cdot)$ *is an increasing function in the asset price* P_t *and an increasing, concave function in the sell action* ζ.

3. $c(\cdot,\cdot)$ *is sufficiently bounded in the price variable to ensure future expectations are bounded.*

This assumption will be exploited in later sections to prove structural results of the optimal liquidation policy.

So far the state of the system involves only the controlled price process. Motivated by the portfolio liquidation problem, we now introduce additional state processes. The portfolio description closely follows [2] up to the formulation of the HJBQVI for the impulse control problem. These states will allow the definition of admissible impulse strategies and allow formulation of applicable solution algorithms.

3.1 Liquidation Dynamics

Beyond the asset price dynamics, the amount of shares owned by the trading agent as a function of time will need to be considered in order to impose realistic constraints

on the liquidation problem. Call the shares process Y_t taking values in $\mathbb{R}_+ \cup \{0\}$. Typically we consider integer numbers of shares of an asset tradable through some exchange, but for the purposes of this work we consider values in some compact subset of the real line. Given the impulse control definition for $\{\tau_n, \zeta_n\}_{n=0,1,\ldots}$, the dynamics of Y_t is given by

$$Y_s = Y_{\tau_n} \text{ for } \tau_n \leq s < \tau_{n+1} \text{ and } Y_{\tau_{n+1}} = Y_{\tau_n} - \zeta_{n+1}, \tag{3.3}$$

and for $Y_s = 0, s \leq T$, $Y_t = 0$ for all $s \leq t \leq T$ such that when the initial position in the asset has been liquidated, it will remain liquidated (no purchasing action is possible).

Next define the amount of cash the investor has in pocket at time $t \in [0,T]$. Let this value be X_t taking values in $\mathbb{R}_+$. The cash in pocket is static between trading times so that

$$X_t = X_{\tau_n}, \tau_n \leq t < \tau_{n+1}, \quad n \geq 0. \tag{3.4}$$

Given an impulse strategy $\alpha = \{\zeta_k, \tau_k\}$ such that $\Delta Y_t = -\zeta_{n+1}$ occurs at time $t = \tau_{n+1}$, then $\Delta X_t \equiv X_t - X_{t^-} = -\int P(s) dY_s$. As such, we clearly ignore illiquidity effects, frequency trading penalties, and bid-ask spreads in this formulation. So for this impulse control we have

$$X_{\tau_n} = X_{\tau_n^-} + \zeta_n P_{\tau_n} = X_{\tau_{n-1}} + \zeta_n P_{\tau_n}, \quad n \geq 1. \tag{3.5}$$

It is clear now that we are only considering impulsive control without a continuous control, as is evident from our control definition and the price dynamics definition (no continuous control terms appear in either the drift or diffusion terms).

In order to apply a procedure to find an impulse control strategy in a realistic market scenario, a solvency constraint requires defining. Consider a liquidation function $L(x,y,p) = x + yp$, representing a full sell action such that no shares of the asset remain. The first constraint is to require the portfolio's liquidation value to satisfy $L(X_t, Y_t, P_t) \geq 0$ for all $t \in [0,T]$. Given this criterion, define the solvency region

$$\mathcal{S} = \{(x,y,p) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ : L(x,y,p) > 0\}, \tag{3.6}$$

with boundary and closure $\partial\mathcal{S} = \partial_y\mathcal{S} \cup \partial_L\mathcal{S}$ and

$$\bar{\mathcal{S}} = \mathcal{S} \cup \partial\mathcal{S}, \tag{3.7}$$

respectively. Here we have

$$\begin{aligned}
\partial_y\mathcal{S} &= \{(x,y,p) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ : L(x,y,p) \geq 0, y = 0\}, \quad \text{and} \\
\partial_L\mathcal{S} &= \{(x,y,p) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ : L(x,y,p) = 0\}.
\end{aligned}$$

Given $(t,x,y,p) \in [0,T] \times \bar{\mathcal{S}}$, we define an impulse control α as follows

Definition 3.1.1. *(Conditions for an Admissible Impulse Control) The couple process $\{\zeta_k, \tau_k\}_{k\geq 1}$ is admissible if the following properties are satisfied:*

1. *$\{X_s, Y_s, P_s\}$ follow state laws (3.1), (3.3), (3.4), and (3.5) and remain in (3.7) for all liquidation impulses, for all time $t \leq s \leq T$*
2. *$0 < \tau_i < \tau_{i+1}, \quad i \geq 1$,*
3. *τ_i is an $\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq k)$ stopping time for $t \geq 0$.*
4. *ζ_i is $\mathcal{F}_t$-measurable, and $\zeta_i \leq Y_{\tau_i}, \quad i \geq 1$, where $\mathcal{F}_t$ is the right-continous filtration,*
5. *$\mathbb{P}\left[\lim_{i\to\infty} \tau_i \leq T\right] = 0$ for all $T \in [0,\infty)$,*

Given an initial state $S_t = (X_t, Y_t, P_t) = (x,y,p)$, call the set of all admissible strategies $\mathcal{A}(x,y,p)$.

3.2 HJBQVI Characterization of Optimal Policy

Associate with the above controlled price process (3.1) an objective function

$$v^\alpha(t,x,y,p) = \mathbb{E}^{x,y,p}\left[\sum_{\tau_i} c(P^\alpha_{\tau_i^-}, \zeta_i)\right]. \tag{3.8}$$

Our goal is to maximize our expected return and find a function $v^*(t,x,y,p)$ such that

$$v^*(t,x,y,p) = \sup_{\alpha \in \mathcal{A}(x,y,p)} v^\alpha(t,x,y,p) \tag{3.9}$$

and the corresponding optimal impulse strategy

$$\alpha^* = \arg \sup_{\alpha \in \mathcal{A}(x,y,p)} v^\alpha(t,x,y,p). \tag{3.10}$$

The standard derivation (see for example [28], [29], [4], or [14]) of the corresponding dynamic programming principle shows that the optimal value function (3.9) is the unique viscosity solution $v(t,x,y,p)$ of the following Hamilton-Jacobi-Bellman quasivariational inequality (HJBQVI)

$$\max\left\{\frac{\partial v}{\partial t} + \mathcal{L}v, \mathcal{M}v - v\right\} = 0, \quad \text{in } [0,T) \times \bar{\mathcal{S}}, \tag{3.11}$$

with terminal condition

$$\max\{v - L(X_T, Y_T, P_T), \mathcal{M}v - v\} = 0, \quad \text{on } \{T\} \times \hat{\mathcal{S}}. \tag{3.12}$$

As a note on the existence of the sup mentioned in (3.10), first define the following theorem:

Theorem 3.2.1. *(Arzelá-Ascoli Theorem)*
Given a compact metric space Ω, a subset $F \subset \mathcal{C}(\Omega)$ is the space of continuous complex valued functions on Ω. F is compact if and only if F is closed, bounded, and equicontinuous.

Proof. The proof is a standard result in functional analysis. □

Since $(x,y,p) \in \bar{\mathcal{S}}$ is compact, and we have placed bounds on α s.t. at all instances $\alpha_t = (\tau_i, \zeta_i) \Rightarrow \zeta_i \leq y_t$ and we can reasonably state that $y_t < y_{max}$ for some finite real valued constant y_{max}, then α is a closed and bounded function over a compact space. Showing that α is equicontinuous allows us to say that α is a continous function, and thus the sup exists for all state combinations $[0,T] \times \bar{\mathcal{S}}$.

Lemma 3.2.1. *(Equicontinuity of Optimal Action Function) The function*

$$\alpha^*(x,y,p) = \arg \sup_{\alpha \in \mathcal{A}(x,y,p)} v^\alpha(t,x,y,p)$$

is equicontinous, i.e., as $\alpha^* : [0,T] \times \bar{\mathcal{S}} \mapsto \mathbb{R}_+$, α^* is said to be equicontinuous at point

$x \in [0,T] \times \bar{\mathcal{S}}$ *if for every* $\epsilon > 0$*,* x *has a neghborhood* U_x *such that*

$$d_{R_+}(\alpha^*(x), \alpha^*(u)) < \epsilon$$

for all $u \in U_x$ *and* $\alpha^* \in \mathbb{R}_+$*. Here* $d_\cdot(\cdot,\cdot)$ *is the distance function.*

Proof. The proof follows from the application of the Uniform Boundedness Principle when acknowledging that the optimal value function α^* is a set of continuous linear operators. □

In (3.11), the infinitesimal generator $\mathcal{L}$ and intervention operator $\mathcal{M}$ are defined in (3.14) and (3.16) below. Let $L(X_T, Y_T, P_T) = C_T(Y_T, P_T)$ be the terminal liquidation reward function, which assumes the following

Assumption 3.2.1. *(Terminal reward function).* $C_T(\cdot)$ *is a concave function of the terminal number of shares* Y_T*, and an increasing function of the terminal price state* P_T *state s.t.*

$$\begin{aligned}
&C_T(X_T, Y_T, P_T) = C_T(Y_T, P_T), C_T(\cdot,\cdot) : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}, \\
&C_T(Y_T + \delta_y, P_T) \geq C_T(Y_T, P_T), \quad \forall Y_T, P_T \in \mathbb{R}_+, \delta_y \geq 0 \text{ and } C_T(0, P_T) = 0, \\
&C_T(Y_T, P_T + \delta_p) \geq C_T(Y_T, P_T), \quad \forall Y_T, P_T \in \mathbb{R}_+, \delta_p \geq 0 \text{ and } C_T(Y_T, 0) = 0.
\end{aligned} \tag{3.13}$$

This can be thought of as a price-dependent, final trade for what has not been sold at the end of the trading interval.

$\mathcal{L}\phi(t,x,y,p)$ is the differential infinitesimal generator function for the uncontrolled price process (3.1) which follows the following definition

Definition 3.2.1. *Let* P_t *be a drift-diffusion process in* $\mathbb{R}_+$ *corresponding to* (3.1)*. The infinitesimal generator* $\mathcal{L}$ *of for a function of* P_t *is given by*

$$\mathcal{L}\phi(t,x,y,p) = \lim_{t \downarrow 0} \frac{\mathbb{E}^{x,y,p}[v(t, X_t, Y_t, R_t)] - v(t,x,y,p)}{t}$$

for all $p \in \mathbb{R}_+$*. Define the set of functions* $f : \mathbb{R}_+ \to \mathbb{R}$ *such that the limit exists at* p *as* $\mathcal{D}_{\mathcal{L}}(p)$ *while* $\mathcal{D}_{\mathcal{L}}$ *denotes the set of all functions in which the limit exists for all* $p \in \mathbb{R}_+$*.*

Notice that $v(t,x,y,p) \in \mathcal{D}_{\mathcal{L}}$ from (3.8), (3.9), and (3.1). Since the impulse actions and immediate/terminal rewards are bounded, $v(\cdot,\cdot,\cdot,\cdot)$ is bounded for all

$(t,x,y,p) \in [0,T] \times \bar{\mathcal{S}}$. Given Definition 3.2.1 the explicit generator function for $v(t,x,y,p)$ exists and is given in general as

$$\mathcal{L}v(t,x,y,p) = \frac{1}{2}\sigma(p)^2 \frac{\partial^2}{\partial p^2} v(t,x,y,p) + \mu(p)\frac{\partial}{\partial p} v(t,x,y,p).$$

Since a geometric Brownian motion process is sought from the process (3.1), the generator function is given by

$$\mathcal{L}v(t,x,y,p) = \frac{1}{2}\sigma^2 p^2 \frac{\partial^2}{\partial p^2} v(t,x,y,p) + \mu p \frac{\partial}{\partial p} v(t,x,y,p). \tag{3.14}$$

$$\tag{3.15}$$

Remark: The structural results in Sec. 3 are not dependent on this particular formulation. The functions $\mu(P_t) = \mu P_t$ and $\sigma(P_t) = \sigma P_t$ are chosen specifically to match the literature consensus on using the geometric Brownian motion to represent the price evolution of a risky asset. Other interpretations are valid, as long as for $\mu(P_t)$ and $\sigma(P_t)$ there exists constants $C_\mu, C_\sigma > 0$ such that for each $x, y \in \mathbb{R}_+$

$$|\mu(x) - \mu(y)| \leq C_\mu |x-y|, |\sigma(x) - \sigma(y)| \leq C_\sigma |x-y|,$$

and $m(\cdot)$ is bounded, so that SDE (3.1) admits a solution. The geometric Brownian motion trivially satisfies this requirement.

The operator $\mathcal{M}$ is an intervention operator such that $\mathcal{M}$ maps a function $v : [0,T] \times \bar{\mathcal{S}} \mapsto [0,T] \times \bar{\mathcal{S}}$ to the function $\mathcal{M}v : [0,T] \times \bar{\mathcal{S}} \mapsto [0,T] \times \bar{\mathcal{S}}$ given by

$$\begin{aligned} \mathcal{M}v(t,x,y,p) = \sup_{e \in \mathcal{C}(x,y,p)} & v(t, \Gamma(x,y,p,e)), \\ & (x,y,p) \in \bar{\mathcal{S}}, e \in \mathbb{R}, t \in [0,T]. \end{aligned} \tag{3.16}$$

and terminal condition $v(T,x,y,p) = C_T(y,p)$. This is the result of the instantaneous sell order ζ at some time t, so that the state process jumps from $S_{t^-} = (x,y,p)$ to $S_t = \Gamma(x,y,p,\zeta) = (x + \zeta p, y - \zeta, p - m(\zeta)) \in \bar{\mathcal{S}}$. The admissible impulse action set $\mathcal{C}(x,y,p)$ is given by

$$\mathcal{C}(x,y,p) = \left\{ e \in \mathbb{R}_+ : \left(\Gamma(x,y,p,e) \in \bar{\mathcal{S}} \right) \right\}.$$

3.3 Main Result

Theorem 3.3.1. *Consider the HJBQVI* (3.11)-(3.12)*. Assume 3.0.1, 3.0.2 and 3.2.1 hold. The optimal value function* $v^*(t,x,y,p)$ *defined in* (3.9) *coincides with the optimal value function under time, state, and action convergence of a denumerable Markov Decision Process with the same cost functions and price dynamics. Furthermore, this MDP has an optimal policy that is nondecreasing with respect to shares owned* Y_t*, the price of the underlying asset* P_t*, and the time accumulated during the liquidation process* t*. Note that the optimal policy of the discrete scheme is not necessarily convergent to the optimal policy of the HJBQVI.*

Proof. The proof of Theorem 3.3.1 is given in the following steps $1-3$:

1. A countable state, action, and time MDP exists with transition probabilities derived from the distribution of (3.1), and with identical immediate and terminal reward functions as (3.2) and (3.13).

2. The truncated trade decision epochs defined by the MDP from Step A approach the HJBQVI impulse trade instances when the truncation is removed, and the optimal value function of the MDP from Step A converges to the optimal value function of HJBQVI (3.11) under time, state, and action discretization.

3. The optimal policy of the MDP is monotone increasing in the shares owned Y_t, price of the underlying asset P_t, and time elapsed t, thus giving the desired structural result.

□

The details of each step $1-3$ of the proof of Theorem 3.3.1 is given in Sec. 4.

Showing the HJBQVI has a monotone result is useful in application because it can be exploited numerically when searching for the optimal policy via a Simultaneous Perturbation Stochastic Approximation (SPSA) type algorithm, explored in Sec. 4.

Chapter 4

Proof of Main Result

4.1 The Denumerable MDP Representation of the HJBQVI

Before defining the MDP, consider the following state and time discretization for the solvency region (3.7) and the HJBQVI (3.11).

This section adapts a discrete derivation and convergence analysis from [30]. Consider a time discretization to the HJBQVI (3.11) as follows. Let the time step be

$$h = T/m, m \in \mathbb{N}\backslash\{0\} \tag{4.1}$$

and let $\mathbb{T}_m = \{t_i = ih, i = 0, ..., m\}$ be the uniform gird over the interval $[0, T]$. So the time discretization of (3.11) leads to the following backward scheme:

$$\begin{aligned} v^h(t_i, x, y, p) &= \max\left\{\mathbb{E}\left[v^h(t_{i+1}, X_{t_{i+1}}, Y_{t_{i+1}}, P_{t_{i+1}})\right], \mathcal{M}v^h(t_i, \Gamma(x, y, p, \zeta))\right\} \\ &= \max\left\{\mathbb{E}\left[v^h(t_{i+1}, X_{t_{i+1}}, Y_{t_{i+1}}, P_{t_{i+1}})\right], \sup_{\zeta \in \mathcal{C}(x,y,p)} v^h(t_i, \Gamma(x, y, p, \zeta))\right\} \end{aligned} \tag{4.2}$$

$$v^h(t_m, x, y, p) = \max\left\{C_T(x, y, p), \sup_{\zeta \in \mathcal{C}(x,y,p)} C_T(y - \zeta, p - m(\zeta))\right\}. \tag{4.3}$$

for $i = 0, ..., m-1$, $t_m = T$, and $(x, y, p) \in \bar{\mathcal{S}}$. Since this definition is implicit due

to the nonlocal obstactle $\mathcal{M}$, the usual way to remedy this issue is to introduce an iterative scheme by considering a sequence of optimal stopping problems (see mentions of this, for example, in [30], [31], [2]) s.t. for $k \in \mathbb{N}\backslash\{0\}$:

$$v^{h,k+1}(t_i,x,y,p) = \max\left\{\mathbb{E}\left[v^{h,k+1}(t_{i+1},X_{t_{i+1}},Y_{t_{i+1}},P_{t_{i+1}})\right], \sup_{\zeta\in\mathcal{C}(x,y,p)} v^{h,k}(t_i,\Gamma(x,y,p,\zeta))\right\} \quad (4.4)$$

$$v^{h,k+1}(t_m,x,y,p) = \max\left\{C_T(x,y,p), \sup_{\zeta\in\mathcal{C}(x,y,p)} C_T(y-\zeta,p-m(\zeta))\right\}. \quad (4.5)$$

This iterative scheme is used temporarily to handle the implicit scheme by approximating the discretized value function, and later discarded by taking many iterations. A more detailed use of this approach is discussed in the proofs section (Appendix B).

For state discretization, define a finite, localized subset of the admissible state space as a uniform grid, denoted by

$$\bar{\mathcal{S}}_{loc} = \bar{\mathcal{S}} \cap \mathcal{X}\times\mathcal{Y}\times\mathcal{P}, \quad (4.6)$$

where

$$\mathcal{X} = [x_{min},x_{max}], \quad \mathcal{Y} = [y_{min},y_{max}], \quad \mathcal{P} = [p_{min},p_{max}]. \quad (4.7)$$

Here, $\mathcal{X}$ has increments of size $(x_{max}-x_{min})/l, l \in \mathbb{N}\backslash\{0\}$, and similar for $\mathcal{Y}$, $\mathcal{P}$. Each state variable does not need to be have the same discretization parameter l, but for ease in notation and implementation assume they are equivalent. Define the regular grid

$$\mathcal{Z}_l = \{(x,y,p) \in \mathcal{X}\times\mathcal{Y}\times\mathcal{P} : (x,y,p) \in \bar{\mathcal{S}}_{loc}\}. \quad (4.8)$$

Define a similar grid for the admissible controls

$$\mathcal{C}_{M,R}(x,y,p) = \{\zeta_i = \zeta_{min} + \frac{i}{M}(\zeta_{max}-\zeta_{min}) : 0 \le i \le M, \Gamma(x,y,p,\zeta_i) \in \bar{\mathcal{S}}_{loc}\}, \quad (4.9)$$

where $\zeta_{min} < \zeta_{max} \in \mathbb{R}_+$ and $M \in \mathbb{N}\backslash\{0\}$ are fixed constants, and

$$R := \min\left(|x_{min}|, |x_{max}|, |y_{min}|, |y_{min}|, |p_{max}|\right). \tag{4.10}$$

We assume the projection on the price variable

$$\Pi_{[0,p_{max}]} : \mathbb{R}_+ \to [0, p_{max}] \tag{4.11}$$

$$p \to p\mathbf{1}_{[0,p_{max}]} + p_{max}\mathbf{1}_{]p_{max},+\infty[} \tag{4.12}$$

in order to constrain the price to $p_{min} = 0$ and p_{max} to some fixed positive constant.

Next, approximate the expectation value from (4.4) and by the following Karhunen-Loève quantizer method (again, adapted from [30])

$$\begin{aligned} &\mathbb{E}\left[v^h(t_{i+1}, X_{t_{i+1}}, Y_{t_{i+1}}, P_{t_{i+1}})\right] \sim \\ &\mathcal{E}^{N,R}[v^h(t_{i+1}, X_{t_{i+1}}, Y_{t_{i+1}}, P_{t_{i+1}}] = \sum_{i=1}^{N_d} \mathbb{P}_{i_1,\ldots,i_{d(N)}} v^{h,n}(t, Z_{N,R}^{0,s,x,y,p}(t)) \quad \forall s \le t \end{aligned} \tag{4.13}$$

where

$$Z_{N,R}^{0,s,x,y,p}(t) = \left(x, y, \prod_{p\in[p_{min},p_{max}]} (p\, \exp\{(\mu - \tfrac{\sigma^2}{2})(t-s) + \sigma W_{i_1,\ldots,i_{d(N)}}^N(t-s)\})\right),$$

$$W_{i_1,\ldots,i_{d(N)}}^N(t) = \sum_{k=1}^{d(N)} \frac{\sqrt{2T}}{\pi(k-\frac{1}{2})} x_{i_k} \sin(\frac{\pi t}{T}(k-\frac{1}{2})),$$

$$1 \le i_k \le N_k, \quad \prod_{k=1}^{d(N)} N_k = N,$$

and

$$\mathbb{P}_{i_1,\ldots,i_{d(N)}} = \prod_{j=1}^{d(N)} \mathbb{P}(x_{i_j}).$$

Here (x_{i_j}) is the N_k quantizer of the standard normal distribution.$W_{i_1,\ldots,i_{d(N)}}^N(t)$ is the quantizer of size N that is the optimal decomposition of $N = N_1 \times N_2 \times \ldots \times N_{d(N)}$ giving the optimal $d(N)$ and (N_k) for the given N. $\mathbb{P}_{i_1,\ldots,i_{d(N)}}$ is the weight associated with the quantizer x_{i_k} and is the product of the weights associated with the

quantization of the normal distribution. The optimal grid (x_{i_n}) and weights $\mathbb{P}_{i_1,\ldots,i_d}$ can be found online at http://www.quantize.maths-fi.com/downloads.

In summary, we have the following notation for the approximating scheme to the HJBQVI (3.11)

1. h : the time step
2. n : iterative scheme index
3. N : the Brownian motion discretization parameter
4. M : the number of admissible transactions
5. R : the boundaries of the localized, discretized solvency region $\bar{\mathcal{S}}_{loc}$.

It is now possible to write the final iterative approximation scheme for the HJBQVI using the new expectation value approximation definition in (4.4) and (4.5)

$$v^{h,n+1}(t_i,x,y,p) = \max\left\{\mathcal{E}^{N,R}\left[v^{h,n+1}(t_{i+1},X_{t_{i+1}},Y_{t_{i+1}},P_{t_{i+1}})\right], \sup_{\zeta\in\mathcal{C}(x,y,p)} v^{h,n}(t_i,\Gamma(x,y,p,\zeta))\right\} \tag{4.14}$$

$$v^{h,n+1}(t_m,x,y,p) = \max\left\{C_T(x,y,p), \sup_{\zeta\in\mathcal{C}(x,y,p)} C_T(y-\zeta,p-m(\zeta))\right\}. \tag{4.15}$$

starting from $v^{h,0} = \mathcal{E}^{N,R}[C_T(Y_T^{0,t,x,y,p},P_T^{0,t,x,y,p})]$.

Rewrite the discrete time and state impulse operator as follows:

$$\mathcal{M}^{M,R}v^h(t_i,x,y,p) = \sup_{\zeta\in\mathcal{C}_{\mathcal{M},\mathcal{R}}(x,y,p)} v(t_i,\Gamma(x,y,p,\zeta)), \quad (x,y,p)\in\mathcal{Z}_l, \quad t_i\in\mathbb{T}_m. \tag{4.16}$$

Next, construct the countable state, countable action MDP to approximate the HJBQVI (3.11) through emulation of the discrete scheme (4.2) and (4.3), using

the time, state, and action discretizations defined above. Let the controlled MDP cash, shares, and price process triplet be $\{S_n\} = \{X_n, Y_n, P_n\}$ for $n \in \mathbb{T}_m = \{t_i = ih, i = 0, ..., m\}$, where $t_m = T$ is the finite liquidation horizon. WLOG rename time intervals s.t. $n \in \mathbb{T}_m := \{0, 1, ..., T\}$. Let $S_n \in \mathcal{Z}_l$ be the set of all possible states at time n, assuming the same discretization as in (4.8). Let the control space at epoch n be $\mathcal{C}_{M,R}(S_n), n = 0, 1, ..., T-1$ as defined in (4.9). Let $\{\alpha_n\}, n = 1, ..., T$ be a control process consisting of the liquidating agent's control decisions at each epoch $n = 0, 1, ..., T-1$, with $\alpha_n(x, y, p) \in \mathcal{C}_{M,R}(x, y, p)$. Let $\mathbb{P}_n(i, j, \alpha_n), i \in \mathcal{Z}_l, \alpha_n \in \mathcal{C}_{M,R}(x, y, p), j \in \mathcal{Z}_l, n = 0, 1, ..., T-1$ be the probability of transitioning from state i to state j given that action a has been made in period n, given by

$$\mathbb{P}_n(i, j, a) = \mathbb{P}(P_{n+1} = j | P_n = i, \alpha_n = a, \mathcal{F}_n) = p(i, j, a). \tag{4.17}$$

with $\mathcal{F}_n = \sigma(B_m : 0 \leq m \leq n)$, with reference to the density function implied by (3.1). The solvency constraint is automatically satisfied as all control actions $\alpha_n(x, y, p), (n, x, y, p) \in \{0, ..., T\} \times \mathcal{Z}_l$ belong to the discrete admissible policy set $\mathcal{C}_{M,R}(x, y, p)$.

An admissible policy α is a set of T functions $\alpha = \{\alpha_0(x_0, y_0, p_0), \alpha_1(x_1, y_1, p_1), ..., \alpha_{T-1}(x_{T-1}, y_{T-1}, p_{T-1}\}$, with $\alpha_n(x_n, y_n, p_n) : \mathcal{Z}_l \mapsto \mathcal{C}_{M,R}(x_n, y_n, p_n)$. So the investor using policy α in state (x, y, p) at epoch n would apply action $\alpha_n(x, y, p)$. Compare this strategy α in the discrete problem to the impulse strategy $\alpha = \{\tau_0, \zeta_0, ...\}$ - in the discrete case there are now predetermined, truncated action epochs, so keeping track of impulse instances is no longer needed.

Let $c_n(i, j, \alpha_m) \in \mathbb{R}_+$ denote the immediate reward at time n when the state is $i = (x, y, p) \in \mathcal{Z}_l$, action $\alpha \in \mathcal{C}_{M,R}(i)$ is taken, and the state transitions to state $j \in \mathcal{Z}_l$. At the end of the liquidation period at interval T, a final reward will be gained by the investor for all remaining shares of the asset in the portfolio.

Define the following MDP components for the portfolio liquidation problem:

Definition 4.1.1. *(Countable state, countable action, discrete time MDP components for the portfolio liquidation problem)*
For $(x, y, p) \in \mathcal{Z}_l$,

1. *Decision Epochs:* $n \in \mathbb{T}_m$,
2. *States:* $(x, y, p) \in \bar{\mathcal{S}}_{loc}$,
3. *Actions:* $\alpha_n(x, y, p) \in \mathcal{C}_{M,R}$ *from* (4.9), $\quad n \in \mathbb{T}_m$,

4. *Immediate Rewards:* $c_n(x,y,p,\alpha_n(x,y,p)) \in \mathbb{R}_+$, *with Assumption 3.0.2*

5. *Terminal Reward:* $C_T(x,y,p) = C_T(y,p) \in \mathbb{R}_+$, *with Assumption 3.2.1*

6. *Transition Probabilities:* $\mathbb{P}_n(i,j,a)$ *defined in* (4.17)

4.2 Convergence of the Approximating Scheme

We will show in this section that this countable state MDP can be used as an approximationg model for the same liquidation problem as the HJBQVI. This section uses a large part of [30] to show convergence analysis results, and keeps notation consistent. As a result, this section summarizes convergence results for approximating schemes to the HJBQVI described by the authors.

First , the impulse instance truncation convergence is shown. Assume the same truncation scheme (4.18), and the corresponding value function v_n, which describes the value function when the trading agent is allowed at most up to n impulse operations In the following lemma, the convergence of the value functions v_N towards the initial value function v is shown.

Lemma 4.2.1. *(Impulse truncation convergence) For all* $(t,x,y,p) \in [0,T] \times \bar{\mathcal{S}}$ *and* $l \in \mathbb{N}\backslash\{0\}$, *the truncated value function defined in* (4.19) *converges to the value function of the non-truncated HJBQVI* (3.9) *such that*

$$\lim_{l\to\infty} v_N(t,x,y,p) = v(t,x,y,p).$$

Proof. See Appendix B. □

Theorem 4.2.1. *Define* $\phi_n(t,x,y,p)$ *iteratively as a sequence of optimal stopping problems s.t.*

$$\phi_{n+1}(t,x,y,p) = \sup_{\tau\in\mathcal{T}_{t,T}} \mathbb{E}\left[\mathcal{M}\phi_n(\tau, X_\tau^{0,t,x,y,p})\right],$$
$$\phi_0(t,x,y,p) = v_0(t,x,y,p),$$

where $\mathcal{T}_{t,T}$ *is the set of all* $\mathcal{F}_s$*-stopping times in* $[t,T], t \le s \le T$. *Then*

$$\phi_n(t,x,y,p) = v_n(t,x,y,p).$$

Proof. See Theorem 3.1 of [30]. □

Given $j \in \mathbb{N}\backslash\{0\}$, introduce the following subsets of $\mathcal{A}(x,y,p)$, the set of the admissible impulse control strategies:

$$\mathcal{A}_j(x,y,p) := \{\alpha = (\tau_i, \zeta_i)_{i=0,\dots,j} \in \mathcal{A}(x,y,p)\}, \tag{4.18}$$

and the corresponding value function v_j, which describes the value function when the trading agent is allowed at most up to j impulse operations:

$$v_j^{\alpha,h,N,M,R}(t,x,y,p) = \mathbb{E}^{x,y,p}\left[\sum_{\tau_i \leq \tau_j} c(P^{\alpha}_{\tau_i^-}, \zeta_i)\right], \quad (t,x,y,p) \in [0,T] \times \bar{\mathcal{S}}.$$

Now, given the state, time, action, and decision epoch truncations of the MDP components, the MDP value function can be written as

$$v^{\alpha,h,N,M,R}(n,x,y,p) = \mathbb{E}^{x,y,p,\alpha}\left\{\sum_{n=0}^{T-1}[c_n(X_n,Y_n,P_n,\alpha_n(X_n,Y_n,P_n)] + C_T(Y_T,P_T))\right\},$$

where $\alpha = (n,\alpha_n)_{n=0,1,\dots,T} \in \mathcal{A}_T(x,y,p)$ and $(t,x,y,p) \in [0,T] \times \bar{\mathcal{S}}_{loc}$. This is a similar value function as in the continuous scenario (3.9), but in this altered framework there are only a finite number of decision epochs while having a continuum of impulse intervention times in the previous interpretation. Note that in the time discretization and trade interval truncation limit (3.9) is obtained. The goal is to find an admissible strategy α^* s.t.

$$\begin{aligned} v^{\alpha^*,h,N,M,R}(n,x,y,p) = \max_{\alpha \in \mathcal{A}_T(x,y,p)} & v^{\alpha,h,N,M,R}(n,x,y,p) \\ & (x,y,p) \in \mathcal{Z}_l, \alpha \in \mathcal{C}_{M,R}(x,y,p), \end{aligned} \tag{4.19}$$

and

$$\begin{aligned} \alpha^*(n,x,y,p) = \arg \max_{\alpha \in \mathcal{A}_T(x,y,p)} & v^{\alpha,h,N,M,R}(n,x,y,p) \\ & (x,y,p) \in \mathcal{Z}_l, \alpha \in \mathcal{C}_{M,R}(x,y,p). \end{aligned} \tag{4.20}$$

Theorem 4.2.2. *(Local Uniform Value Function Convergence)*
For all $(t,x,y,p) \in [0,T) \times \bar{\mathcal{S}}$, $l \in \mathbb{N}\backslash\{0\}$, *given discretization definitions* (4.8), (4.9), (4.10), *and* (4.1), *the value function* $v_l^{h,R,N,M}(t,x,y,p)$ *on* $\mathbb{T}_m \times \mathcal{Z}_l$ *of the discrete, epoch truncated MDP* (4.19) *which emulates the control structure of the HJBQVI discrete scheme* (4.2) *and* (4.3) *converges locally uniformly to the viscosity sub- and*

supersolution $\bar{v}$ *and* $\underline{v}$*, respectively, of the value function* $v(t,x,y,p)$ *of the HJBQVI* (3.9)*, which are equivalent s.t.* $\bar{v} = \underline{v} = v(t,x,y,p)$*, i.e.,*

$$\lim_{\substack{(t',x',y',p')\to(t,x,y,p)\\(h,M,N,R)\to(0,+\infty,+\infty,+\infty)\\(t',x',y',p')\in\mathbb{T}_m\times\mathcal{Z}_l}} v_l^{h,N,M,R}(t',x',y',p') = v(t,x,y,p). \tag{4.21}$$

Proof. See proof in Appendix B which follows the proof methodology of [30], or see also [2], [27], or [32] for similar treatments. □

It is now possible to define a countable state, countable action MDP whose value function converges to the value function of the original HJBQVI (3.11), since we have convergence in truncated sell instances from Lemma 4.2.1, and convergence in state, action, and time discretizations from Theorem 4.2.2. In the next section we use these results to determine the structure of the optimal policy (4.20) of the MDP discretization scheme (Definition 4.1.1) for the HJBQVI.

4.3 Optimal Policy Structure

The optimal liquidation total reward function and optimal liquidation strategy $v_n^{\alpha^*}(x,y,p)$ and $\alpha^* = \{\alpha_0^*, \alpha_1^*, ..., \alpha_{T-1}^*\}$, respectively are obtained from the following equations, where each decision rule $\alpha_n^*, n = 0,1,...,T-1$ are the solutions to Bellman's recursive equation, which can be rewritten and solved using a backwards induction algorithm s.t.

$$v_n(x,y,p) = \max_{\alpha_n\in\mathcal{C}_{M,R}(x,y,p)} Q_n(x,y,p,\alpha_n), \tag{4.22}$$

$$\alpha_n^*(x,y,p) = \arg\max_{\alpha_n\in\mathcal{C}_{M,R}(x,y,p)} Q_n(x,y,p,\alpha_n), \tag{4.23}$$

where $Q(\cdot,\cdot,\cdot,\cdot)$ is the state-action reward function given by

$$Q_n(x,y,p,\alpha_n) = \left\{ c(p,\alpha_n) + \sum_{j\in\mathcal{Z}_l} \left[\mathbb{P}_n(j,(x,y,p),\alpha_n) \times v_{n+1}(x+p\alpha_n, y-\alpha_n, p_j) \right] \right\}, \tag{4.24}$$

and with $v_T(x,y,p) = C_T(y,p)$.

The following structural result has been adapted from [11], where the probability

of receiving an investment opportunity is given as a condition on the HJBQVI itself.

At the initial time $n = 0$, there are $y_0 = \bar{y}$ shares to liquidate over the time interval $\mathbb{T}_m = \{0, 1, ..., T\}$. At each integer epoch $n \in \mathbb{T}_m$ the trader decides how many shares to sell $\alpha_n \in [0, y_n], 0 \leq y_n \leq \bar{y}$, where y_n is the number of shares owned at epoch n. The reward for selling α_n shares is given by (5.5), and the admissible actions are given by $\mathcal{A}_T(x, y, p)$ defined in (4.18). Apply assumption 3.0.2 and $c(p, 0) = 0$. WLOG assume a constant price p and write $c(\alpha, p) = c(a, p) = c(a)$. When a time variable is used in an expression such as $v_n(x_n, y_n, p_n)$, then the subscripts are generally ignored and written as $v_n(x, y, p)$. Let $v_n(x, y, p)$ denote the maximal expected additional profit attainable when there are $T - n$ decision epochs to go and y shares of the asset left in the portfolio, when a selling opportunity is at hand.

From the HJBQVI definition (3.11), a selling opportunity is at hand when

$$\max\left\{\frac{\partial v}{\partial t} + \mathcal{L}v, \mathcal{M}v - v\right\} = \mathcal{M}v - v = 0. \tag{4.25}$$

In the discrete time scheme, let the probability that a selling opportunity is at hand be q be given by the fraction of the time between decision epochs $i, i+1 \in \mathbb{T}_m$ s.t. condition (4.25) holds. Then

$$q_n \in [0, 1], n \in \mathbb{T}_m. \tag{4.26}$$

For simplicity let $q_n = q$ for all $n \in \mathbb{T}_m$, as the numerical value of q has no impact on the structural argument.

The function $v_n(\cdot, \cdot, \cdot, \cdot)$ satisfies the optimality equation

$$\begin{aligned} &v_n(x, y, p) = \max_{0 \leq \alpha \leq y} \left\{c(\alpha, p) + \bar{v}_{n+1}(x, y - \alpha, p - m(\alpha))\right\}, \quad n \in \mathbb{T}_m, \\ &v_T(x, y, p) = C_T(y, p), \\ &v_{T+1}(x, y, p) = 0, \end{aligned} \tag{4.27}$$

where

$$\bar{v}_m(x, y, p) = \sum_{i=T-m}^{T} q(1-q)^i v_{m-1}(x, y, p).$$

Here $\bar{v}_m$ is the expected maximal addition sum of rewards when y asset shares

remain to sell at price p, there are $T-m$ decision epochs to go, and it is not yet known if a selling opportunity is available. By conditioning on whether or not a selling opportunity occurs, obtain

$$\bar{v}_m(x,y,p) = qv_m(x,y,p) + (1-q)\bar{v}_{m+1}(x,y,p).$$

Begin by showing that v inherits the concavity property of $c(\cdot,\cdot)$ in y.

Lemma 4.3.1. *(Concave Value Function)*
Given the discrete MDP (Definition 4.1.1) with immediate and terminal reward functions defined by Assumptions 3.0.2 and 3.2.1 and price impact function defined by Assumption 3.0.1, the value function $v_n(x,y,p)$ defined in (4.27) *is a nondecreasing, concave function of y.*

Proof. See Appendix B. □

Next let the process $\alpha_n(x,y,p)$ to be the value of α (or the smallest value of α if there is more than one) that maximizes the right hand side of (4.27). Then $\alpha_n(x,y,p)$ is the optimal number of shares to sell at time epoch n when there are y shares remaining in the portfolio, a selling opportunity is present, and the price of the asset is p. Theorem 4.3.1 below gives the structure of the optimal policy, but first consider the following assumption

Assumption 4.3.1. *(Probability of Incurring Market Impact) Assume that after a sell action a is made, the market impact term $m(a) : \mathbb{R}_+ \to \mathbb{R}_+$ does not occur with probability 1. Instead, assume that there is a function $f(a,p) : \mathcal{C}_{M,R} \times \mathcal{Z}_l \cap \mathcal{P} \to [0,1]$ given by*

$$f_m(a,p) = \begin{cases} 0 & \text{if } a = 0 \\ p_m(p) & \text{if } a > 0, \end{cases} \tag{4.28}$$

where $p_m(\cdot) : \mathcal{Z}_l \cap \mathcal{P} \to [0,1]$ is the market impact probability function, given that a sell action has been made.

Theorem 4.3.1. *(Optimal Policy Structure)*
Assume the MDP (Definition 4.1.1) with market impact function $m(\cdot)$ defined by Assumption 3.0.1 and with immediate and terminal reward functions defined by Assumptions 3.0.2 and 3.2.1. If the MDP has a value function defined by (4.22)

that is a nondecreasing and concave function of y and is convergent to the value function of HJBQVI (3.11) *by Lemma 4.2.1 and Theorem 4.2.2, having investment opportunity parameter q_n defined in* (4.26)*, and market impact probability function given by Assumption 4.3.1, then the optimal policy* (4.23) *of MDP (Definition 4.1.1) has the following structural properties:*
(i.) $\alpha_n^*(x,y,p)$ *is a nondecreasing function of* y*,*
(ii.) $\alpha_n^*(x,y,p)$ *is a nondecreasing function of* n*.*
(iii.) $\alpha_n^*(x,y,p)$ *is a nondecreasing function of* p*.*

Proof. See Appendix B. □

Intuitively, Theorem 4.3.1 states that the number of shares sold at a time n will increase if there are more shares owned, since the end goal is a complete liquidation by the horizon. Similarly, if the number of decision epochs that have passed increases, the investor will want to sell a larger number of shares to meet the liquidation deadline. Moreover, a higher asset price will entice the decision maker to more readily sell more shares than if the price is lower.

Lemma 4.3.1 and Theorem 4.3.1 state that as long as the instantaneous and terminal reward functions follow Assumptions 3.2.1 and 3.0.2 (are nondecreasing, concave functions of the trading action), then the optimal liquidation actions are nondecreasing in the number of shares held Y, and nondecreasing in the total time elapsed n. Lemma 4.2.1 and Theorem 4.2.2 shows that the countable state, countable action MDP from definition 4.1.1 is a valid approximation as it adheres to the impulse truncation, state, action, and time discretizations, and converges to the original HJBQVI (3.11) when discretization parameters $(h, M, N, R) \to (0, +\infty, +\infty, +\infty)$. As such, it has been shown that the HJBQVI has the same value function as a discretized MDP with a monotone optimal policy when converged.

Chapter 5

Numerical Results

5.1 Finite Difference Approximations: Ground Truth Estimates

To solve the HJBQVI problem numerically, an appropriate discretization is needed. Let the localized problem conform to the uniform grid $D_L \equiv [x_0, ..., x_L] \times [y_0, ..., y_L] \times [p_0, ..., p_L]$ where $x_0 \equiv x_{min}, x_L \equiv x_{max}, y_0 \equiv y_{min}$,etc. A similar discretization scheme is used as in [27] where

$$\partial_p^h = \frac{v(x,y,p+h,t) - v(x,y,p)}{h}$$
$$\partial_{pp}^{2,h} = \frac{v(x,y,p+h,t) + 2v(x,y,p) - v(x,y,p-h)}{h^2},$$

with $(x,y,p+h), (x,y,p)$, and $(x,y,p-h) \in \bar{\mathcal{S}}$. Consider the time discretization $\{t_i\}$ with $t_0 = 0$ and $t_N = T$ such that $i = \frac{T}{N}$. A new discrete generator function is obtained

$$\mathcal{L}^h \phi(s) = -\frac{1}{2}\sigma^2 p^2 \partial_{pp}^{2,h} \phi(s) - \mu p \partial_p^h \phi(s). \tag{5.1}$$

The state discretized HJBQVI then becomes

$$\max\left\{\frac{\partial v}{\partial t} + \mathcal{L}^h v, v - \mathcal{M}v\right\} = 0, \quad \text{in } \mathbb{R}_+. \tag{5.2}$$

Assume zero Neumann boundary conditions on the boundary such that

$$\frac{\partial v}{\partial p}(x, y, \cdot, p_{max}) = 0, \quad \forall x, y \in [x_0, ..., x_L] \times [y_0, ..., y_L]. \tag{5.3}$$

The localized, discretized problem (5.2) can be solved using the following iterative method over discrete time steps:

$$\begin{aligned} &\mathcal{L}^h v_T = 0 \quad \forall p \in [p_{min}, p_{max}], \\ &\max\{v_{n+1} - v_n + \mathcal{L} v_{n+1}, \mathcal{M} v_n - v_{n+1}\} = 0, \quad \forall n \geq 0, \end{aligned} \tag{5.4}$$

with boundary condition (5.3) and constraint

$$\max\{v_T - C_T(Y_T, P_T), \mathcal{M} v_{T-1} - v_T\} = 0, \quad \text{on } \{T\} \times \hat{\mathcal{S}}.$$

The right hand side of (5.2) is equated to zero and the value function is solved for, with results given in Fig. 7.1, with values used in the iteration process. The shape of this value function can be compared to that of the discrete MDP, shown in the Sec. 4-B, and demonstrates the convergence of the value function studied in Lemma 4.2.1 and Theorem 4.2.2.

5.2 Countable State, Action, and Time MDP

To derive the dynamic programming algorithm for the countable state, action and time finite horizon MDP [12] is used, which is the discrete time equivalent to the HJBQVI. To develop this algorithm, first the transition probabilities are defined.

Given that the system in state (x_n, y_n, p_n) at epoch n, the probability of transitioning to the state $(x_{n+1}, y_{n+1}, p_{n+1}) = (\bar{x}, \bar{y}, \bar{p})$, given that at time n the action $\alpha_n = a \in \mathcal{A}_n(x, y, p)$ was applied, is obtained by solving for the diffusion density while also considering the deterministic, permanent price impact function. Since (3.1) admits a unique solution (the coefficients of the stochastic differential equation are Lipschitz and following linear growth [33] (Sec. 5.2)), given $\mathcal{F}_n = \sigma(B_m : 0 \leq m \leq n)$ the transition probabilities are given by

$$\mathbb{P}((x_{n+1},y_{n+1},p_{n+1})=(\bar{x},\bar{y},\bar{p})|(x_n,y_n,p_n)=(x,y,p),\alpha_n=a,\mathcal{F}_n)$$

$$=\mathbb{P}(P_{n+1}-P_n=\bar{p}-p+m(a))$$
$$=\frac{1}{\sigma\sqrt{2\pi}\Delta n}\int_{\bar{p}_i^-}^{\bar{p}_i^+}\frac{1}{u}\exp\left\{-\frac{\left(\ln(u)-\ln(p-m(a))-(\mu-\frac{\sigma^2}{2})\Delta n\right)^2}{2\sigma^2\Delta n}\right\}du$$

$$\equiv\mathbb{P}_n(\bar{p},p,a)$$

where Δn is the time difference between decision epochs $n+1$ and n (if not equal to 1). Denote p_i^+ and p_i^- to be the upper and lower values of the price discretization $p\in\{p_0,p_1,...,p_T\}$ s.t. $p_i^+=p_i+1/(2T)$ and $p_i^-=p_i-1/(2T)$. Note that if no action is taken at decision epoch n and defining the market impact term $m(a)=0$ if $a=0$, the transition probability becomes

$$\mathbb{P}((x_{n+1},y_{n+1},p_{n+1})=(\bar{x},\bar{y},\bar{p})|(x_n,y_n,p_n)=(x,y,p),\alpha_n=0,\mathcal{F}_t)$$

$$=\mathbb{P}(P_{n+1}-P_n=\bar{p}-p+m(0))$$

$$=\mathbb{P}(P_{n+1}-P_n=\bar{p}-p)$$

$$=\frac{1}{\sigma\sqrt{2\pi}\Delta n}\int_{\bar{p}_i^-}^{\bar{p}_i^+}\frac{1}{u}\exp\left\{-\frac{\left(\ln(u)-\ln(p-(\mu-\frac{\sigma^2}{2})\Delta n\right)^2}{2\sigma^2\Delta n}\right\}du$$

$$\equiv\mathbb{P}_n(\bar{p},p,0)$$

So given an action a in state p, the probability of successfully transitioning to a new state $\bar{p}$ is given by $\mathbb{P}_n(i,j,a)$, where $\mathbb{P}_n(i,j,0)<\mathbb{P}_n(i,j,a)$, for all $i,j\in\mathcal{Z}_l,n\in\{0,1,...,T-1\}$ and $a_0\geq a_1\Rightarrow\mathbb{P}_n(\bar{p},p,a_0)\geq\mathbb{P}_n(\bar{p},p,a_1)$ under the natural assumption that $m(a_0)\geq m(a_1)$.

Assume that the instantaneous reward function is given by $c_n(p,a) = cpa$ for a positive constant c. Similarly, the final reward function is given by $C_T(Y_T, P_T) = CY_T$ for another positive constant C. Note that these functions adhere to our assumptions 3.0.2 and 3.2.1. A discounting factor δ is used to attenuate future expected rewards, which does not affect the structural result. The discounted optimal value function is given by

$$V_n(s) = \max_{a \in \mathcal{A}(s)} Q_n(x,y,p,a) = \max_{a \in \mathcal{A}(s)} \Big\{ c(p,a) + \\ + \delta \times \sum_{p_j \in \mathcal{Z}_l} \Big[\mathbb{P}_n(p_j, p, a) \times V_{n+1}(x + pa, y - a, p_j) \Big] \Big\}.$$

Assume a uniform discretization for each state variable so that $h_x = h_y = h_p = h$. Table 1 in Appendix A provides a list of parameter values used in this simulation. Adjust Assumption 3.0.2 to the immediate reward function as follows

Assumption 5.2.1. *(Concave immediate reward function)*
Assume an immediate reward function $c_n(x,y,p,\alpha_n(x,y,p)) = c_n(p,\alpha_n(x,y,p))$ *equivalent to* (3.2)*, which is nondecreasing concave in* $\alpha(\cdot)$*.*

For example, $c(\cdot,\cdot)$ can be written as an intuitive "reward with transaction cost" function

$$c(P_n, \alpha(\cdot)) = P_n \alpha(\cdot) - K(\alpha(\cdot)), \tag{5.5}$$

where $K(a)$ is a variable transaction cost incurred when initiating a trade of quantity $a > 0$. It is clear that $c_n(\cdot,\cdot)$ is an increasing function in p_n and increasing in sell action a if $K(a) < p_n a$. This reward structure will be used to simulate the MDP in Sec. 4.

The liquidation horizon is taken as an arbitrary amount of time from the initial state, and has arbitrary units. The drift and diffusion coefficients, as well as the min/max state values are chosen to approximately emulate the values used in Test 1 of [2]. The instantaneous reward is taken as unity to match the expectation that selling a share at a given price will give you exactly that price in return. The terminal reward is chosen as less than unity to coerce the trader into pre-terminal trades. This emphasizes the optimal action structure as is shown below in Fig. 7.6

The transition probabilities for various actions (ranging from the "no trade" action to the "maximum trade" action) are represented in Fig. 7.2 and Fig. 7.3. Notice that the transition probabilities begin to cluster around a final price which is lower when larger and larger trade actions are applied. This coincides with our assumption that large trade orders have a greater negative impact on the asset price compared with smaller trades. On the other hand, a "no trade" action allows the process to follow a typical geometric brownian motion transition pattern (in this case with a positive drift term), and transitions from lower prices to higher prices are more common. These transition probabilities are now used in the discrete time MDP to find optimal actions dependent on number of trade intervals remaining and the number of shares owned by the trading agent.

Fig. 7.4 and Fig. 7.5 show the form of the value function over varying price and time taking snapshots in shares owned. As was shown in Lemma 4.3.1, the value function is monotonically decreasing in time and monotonically increasing in the number of shares owned.

Finally, the corresponding liquidation strategies are given in Fig. 7.6 and Fig. 7.7, in cross sections of constant price. The anticipated threshold nature of the optimal policy is apparent, and dependent on the price process. The shape of the policy appears exponential in both dimensions, and this fact will be exploited in the next section by using the SPSA algorithm.

5.3 Policy Search Algorithm for Unknown Price Evolution Structure

Up to this point, the state transition matrix is assumed to governed solely by known, constant drift and diffusion coefficients μ and σ. In practice, these coefficients may be difficult to estimate effectively. An on-line machine learning type algorithm is proposed to aid the agent in formulating an optimal threshold strategy under this uncertainty. This is done by approximating the parameters which describe the optimal threshold policy. Here we assume the MDP (Definition 4.1.1) has immediate and terminal reward functions defined by Assumptions 3.0.2 and 3.2.1, respectively, and has a value function (4.22) that is a nondecreasing and concave function of y and p and convergent to the value function of HJBQVI (3.11) with investment opportunity parameter q_n defined in (4.26). Under these conditions, from Theorem 4.3.1 the optimal policy $\alpha_n^*(x, y, p)$ defined in (4.23) is a nondecreasing function of p, y,

and n, and a suitable policy gradient algorithm can be used to estimate the parameters of the underlying threshold policy. The Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm is utilized, a gradient estimation methodology which allows the agent to apply the dynamic programming principles described in previous sections under this new assumed uncertainty. Detailed information about this algorithm can be found in [24], and its use in this thesis has been adapted from [34]. As opposed to other stochastic approximation methods based on finite difference-based algorithms, SPSA is beneficial because it reduces the number of loss measurements required to obtain a specified level of accuracy in the optimization procedure. When applying an optimization process to optimal action in financial markets, the difference in efficiency of the algorithm can lead directly to monetary loss for the agent and as such the quickest method is crucial. Algorithm 1 below uses the SPSA algorithm to generate a sequence of estimates $\hat{\phi}_n, n = 1, 2, ...,$ that converges to a local maximum of the optimal threshold descriptor ϕ^* with policy $\alpha_{\phi^*}(x, y, p)$ for the discrete MDP used to describe the optimal portfolio liquidation procedure in Definition 3.1.

Algorithm 1 SPSA Algorithm for Computing a Threshold Policy

Assume that the optimal liquidation policy is characterized by a threshold switching curve as described in the previous section.
Step 1: Choose initial threshold coefficients $\hat{\phi}_0$ and threshold policy $\alpha_{\hat{\phi}_0}$.
Step 2: For iterations $n = 0, 1, 2, ...,$

1. Evaluate trade cost $v_n(\hat{\phi}_n)$ computed from

$$v_T^*(s_N) = C_T(Y_T, P_T), \quad \forall s_N \in \bar{\mathcal{S}}$$

and compute the gradient estimate

$$\hat{\nabla}_\phi v_n(\hat{\phi}_n) = \frac{v_n(\hat{\phi}_n + \Delta_n \omega_n) - v_n(\hat{\phi}_n - \Delta_n \omega_n)}{2\Delta_n} \omega_n,$$
$$\omega_n = \begin{cases} -1 & \text{with probability } 0.5 \\ +1 & \text{with probability } 0.5. \end{cases}$$

Here, $\Delta_n = \Delta/(n+1)^\gamma$ denotes the gradient step size with $0.5 \leq \gamma \leq 1$ and $\Delta > 0$.

2. Update threshold coefficients $\hat{\phi}_n$ via the stochastic approximation algorithm

$$\begin{aligned} &\hat{\phi}_{n+1} = \hat{\phi}_n - \epsilon_{n+1} \hat{\nabla}_\phi v_n(\hat{\phi}_n), \\ &\epsilon_n = \epsilon/(n+1+b)^\beta, 0.5 \leq \beta \leq 1, \quad \epsilon, b > 0. \end{aligned} \tag{5.6}$$

The above SPSA algorithm picks a single random direction ω_n along which direction the derivative is evaluated at each batch n. Unlike the Kiefer-Wolfowitz finite difference algorithm to evaluate the gradient estimate $\hat{\nabla}_\phi v_n$ in (5.6), SPSA requires only 2 batch simulations, i.e., the number of evaluations is independent of dimension of parameter ϕ. Because the stochastic gradient algorithm (5.6) converges to local optima, it is necessary to try several initial conditions $\hat{\phi}_0$. The computational cost at each iteration is linear in the dimension of θ and is independent of the observation size.

For fixed θ, the samples $v_n(\mu_\theta)$ are simulated independently and have identical distribution. Thus, the proof that $\theta_n = \theta^{\hat{\phi}_n}$ generated by Algorithm 1 converges to a local optimum of $\mathbb{E}[v_n(\mu_\theta)]$ with probability one, is a straightforward application of techniques in [35](which gives general convergence methods for Markovian dependencies).

The model used to fit the threshold structure of the optimal policy is a triple exponential with maximal points at time T, y_{max}, and p_{max} (since the policy is monotone increasing in each state variable) s.t.

$$A(t,y,p) = \left\| \left(\theta_1 \mathrm{e}^{\theta_2(y-y_{max})}, \theta_3 \mathrm{e}^{\theta_4(t-T)}, \theta_5 \mathrm{e}^{\theta_6(p-p_{max})} \right) \right\|,$$

where $\|\cdot\|$ is the L^2-norm. So $\phi = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6\}$. The sub-optimal monotone policy resulting from this algorithm for the unknown parameter scenario are given in Fig. 7.8 and Fig. 7.9, and can be compared to the optimal policy simulated earlier in Fig. 7.6 and Fig. 7.7.

Chapter 6

Conclusion

A Hamilton-Jacobi-Bellman quasivariational inequality in a optimal portfolio liquidation scenario has been formulated with immediate market impact considered. This continuous time, state, and action space problem has been discretized using uniform grids, and approximated using a countable MDP that can be solved using a backward induction algorithm. The optimal decision rule and value function are explored and it is found that the value function is concave, nondecreasing in the number of shares owned, the price of the underlying asset, and the time until the liquidation horizon. Further, the optimal policy is found to be nondecreasing in the number of shares owned, the asset price, and the time elapsed. Finally, it is shown that the monotonic and threshold characteristics of the discretized decision process are maintained when converging to the HJBQVI, thus presenting a method for demonstrating optimal policy structural characteristics of an otherwise complex impulse control problem. In a situation where it is unsure how to model the asset price evolution (i.e., the model's drift and diffusion coefficients are unknown), a reinforcement learning technique called the Simultaneous Perturbation Stochastic Approximation algorithm is applied. This algorithm is able to approximate the functional structure of the threshold policy when determining the MDP solution. Therefore, a method for applying a machine learning approach to the solution structure of a Hamilton-Jacobi-Bellman Quasivariational Inequality has been presented.

In future work it is worthwhile considering bid-ask spreads and penalizing rapid transaction requests as explored in [2]. Moreover, including an option to buy and sell shares of multiple stocks may provide new ways to minimize risk through exploiting correlations (or anti-correlations) in stock price evolutions. Other on-line

learning techniques may be explored if alternative scenarios deserve application, and multiple-investor quasivariational inequality applications with game theoretic structures may have interesting machine learning applications as well.

Chapter 7

Tables and Figures

7.1 Tables

Parameter	Description	Value
T	liquidation horizon	10
μ	drift coefficient	0.1
σ	diffusion coefficient	0.3
h	discretization parameter	6
x_{min}	minimum cash in pocket	0
x_{max}	maximum cash in pocket	1e5
y_{min}	minimum shares owned	0
y_{max}	maximum shares owned	50
p_{min}	minimum price of asset	0
p_{max}	maximum price of asset	60
c	instantaneous reward constant	1
C	terminal reward constant	0.6
Δn	time discretization size	1
δ	discounting factor on expected future value	0.9
κ	price impact constant	0.1

Table 7.1: Simulated Liquidation Problem Parameters

7.2 Figures

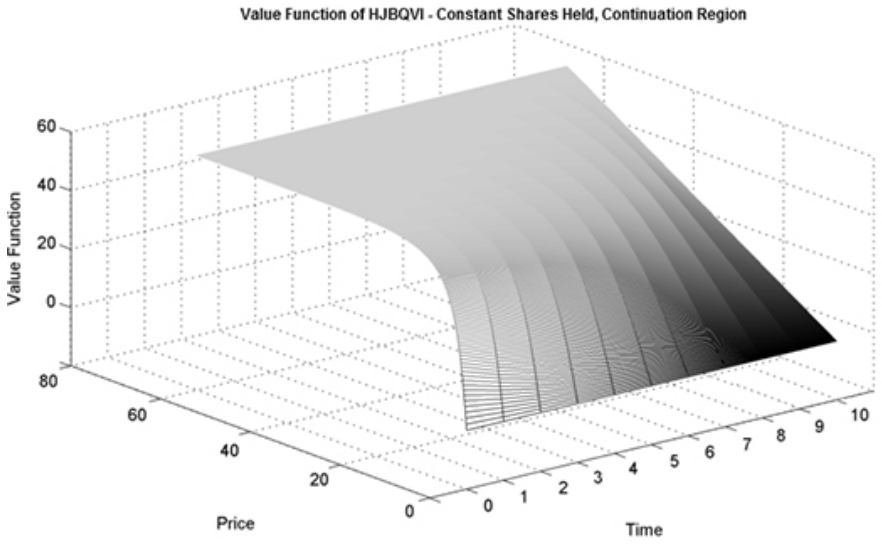

Figure 7.1: Solving for the Value Function from Generator Function (5.1), Snapshot at Constant Shares Owned

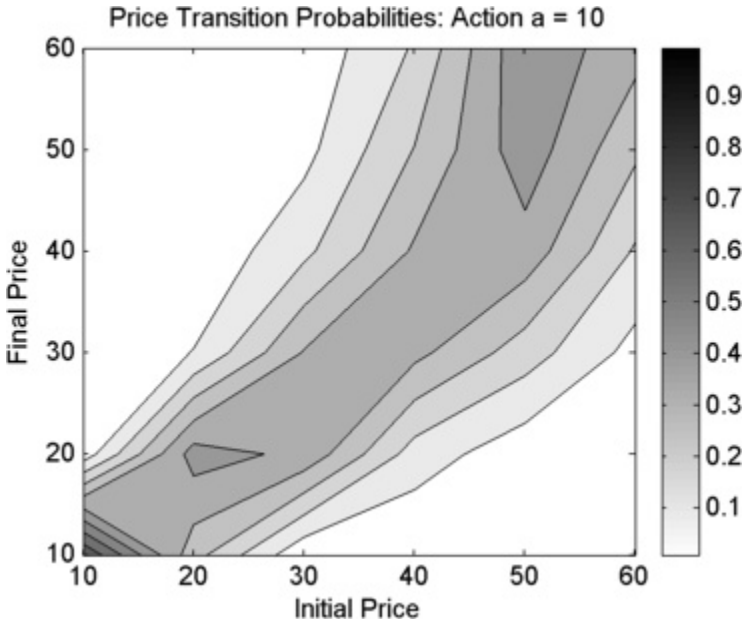

Figure 7.2: Small action taken ("Sell 10 shares")

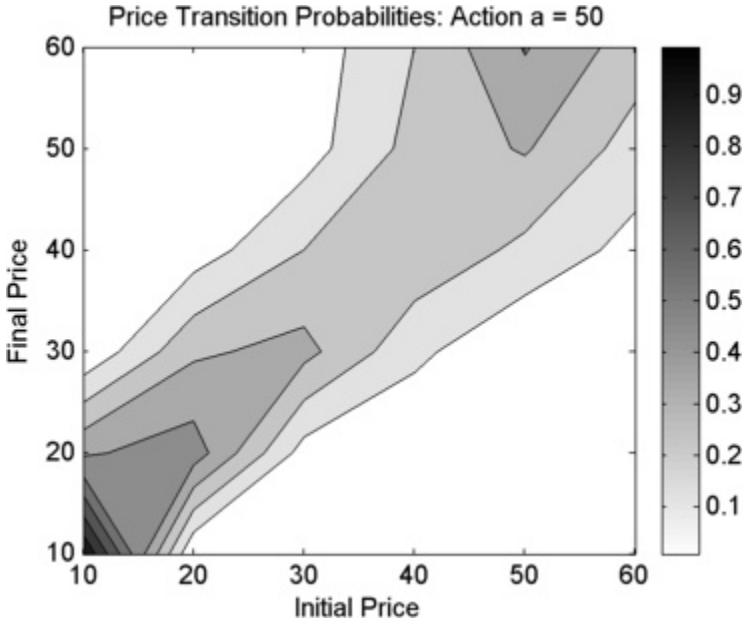

Figure 7.3: Maximum action taken ("Sell 50 shares")

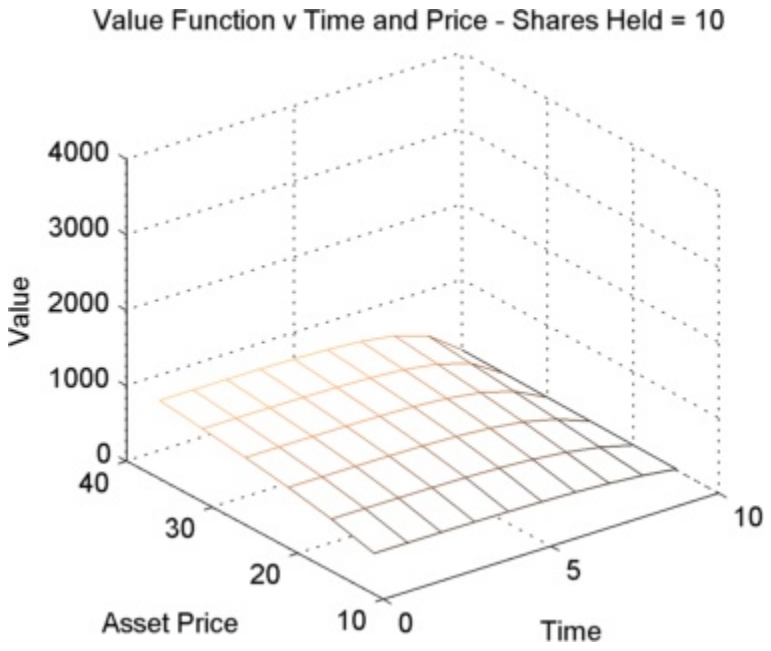

Figure 7.4: Value function over time and price, with constant shares held.

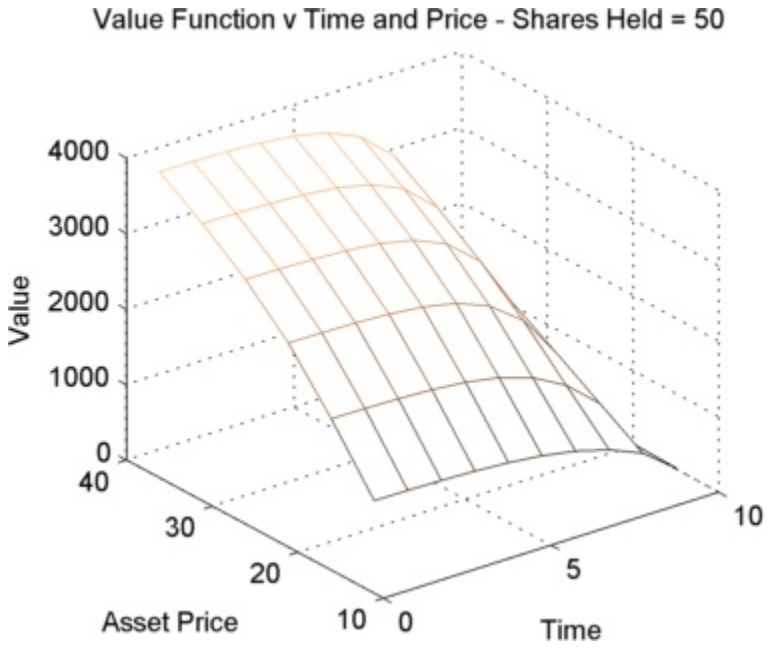

Figure 7.5: Value function over time and price, with constant shares held.

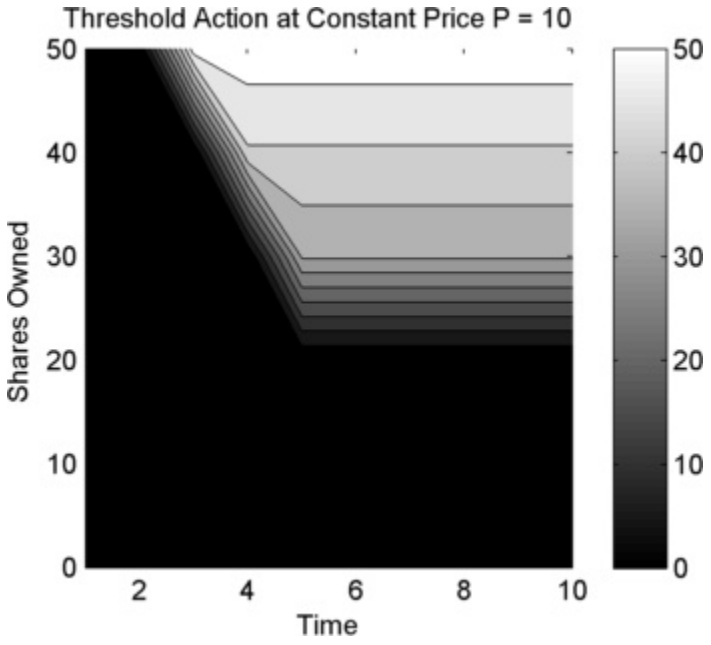

Figure 7.6: Optimal Liquidation Strategy - Constant Price $P = 10$.

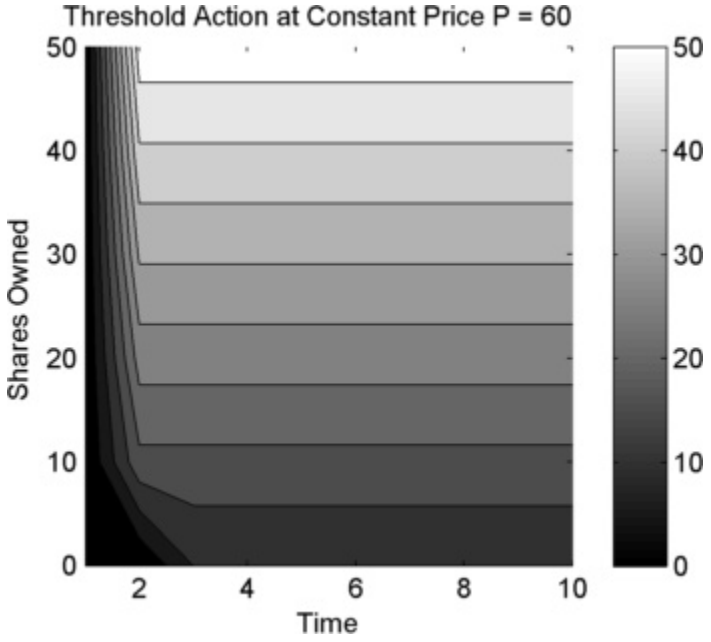

Figure 7.7: Optimal Liquidation Strategy - Constant Price $P = 60$.

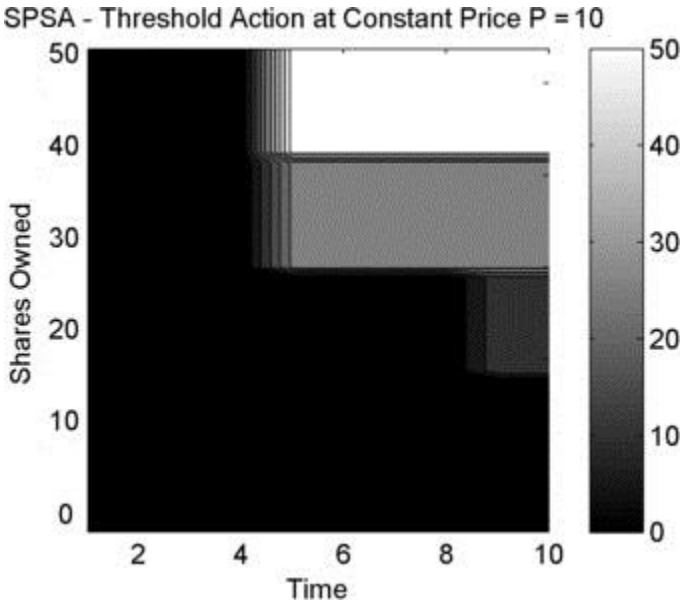

Figure 7.8: SPSA: Sub-Optimal Liquidation Strategy - Constant Price $P = 10$.

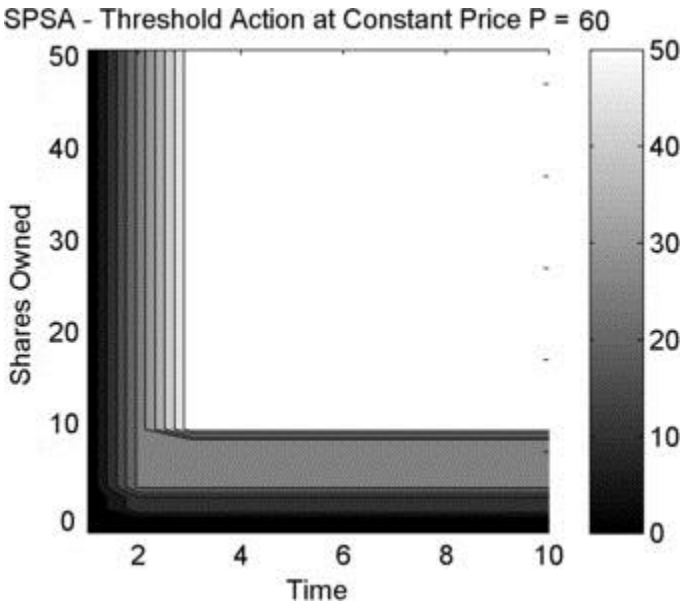

Figure 7.9: SPSA: Sub-Optimal Liquidation Strategy - Constant Price $P = 60$.

Bibliography

[1] Alain Bensoussan and Jacques Lions, *Impulse control and quasi-variational inequalities*, Gaunthier-Villars, 1984. → pages 1, 2, 8

[2] F. Guilbaud, M. Mnif, and H. Pham, "Numerical methods for an optimal order execution problem," *ArXiv e-prints*, June 2010. → pages 1, 2, 3, 4, 15, 16, 24, 30, 37, 42

[3] H. Pham, *Continuous-time Stochastic Control and Optimization with Financial Applications*, Stochastic Modelling and Applied Probability (Book 61). Springer, first edition, 2009. → pages 1, 2

[4] M. Ohnishi and M. Tsujimura, "An impulse control of a geometric brownian motion with quadratic costs," *European Journal of Operational Research*, vol. 168, no. 2, pp. 311 – 321, 2006. → pages 1, 19

[5] A. Schied and T Schöneborn, "Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets," *Finance and Stochastics*, vol. 13, no. 2, pp. 181–204, 2009. → pages 1, 2

[6] F. Black and M. Scholes, "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, vol. 81, no. 3, pp. 637–54, May-June 1973. → pages 1, 15

[7] G. Pages, H. Pham, and J. Printems, "Optimal quantization methods and applications to numerical problems in finance," in *Handbook of Computational and Numerical Methods in Finance*, SvetlozarT. Rachev, Ed., pp. 253–297. BirkhÃďuser Boston, 2004. → pages 1

[8] R. Rishel and K. Helmes, "A variational inequality sufficient condition for optimal stopping with application to an optimal stock selling problem," *SIAM Journal on Control and Optimization*, vol. 45, no. 2, pp. 580–598, 2006. → pages 1

[9] J. Getheral, "No-dynamic-arbitrage and market impact," *Quantitative Finance*, vol. 10, no. 7, pp. 749–759, sep 2010. → pages 2, 15

[10] E. Moro, J. Vicente, L. Moyano, A. Gerig, J. Farmer, G. Vaglica, F. Lillo, and R. Mantegna, "Market impact and trading profile of large trading orders in stock markets," Papers, arXiv.org, Aug. 2009. → pages 2

[11] Sheldon Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, 1991. → pages 2, 30

[12] Martin Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994. → pages 2, 5, 7, 35

[13] N. Bauerle and U. Rieder, *Markov Decision Processes with Applications to Finance*, Springer, 2011. → pages 2

[14] A. Friedman, "Optimal stopping problems in stochastic control," *SIAM Review*, vol. 21, no. 1, pp. 71–80, 1979. → pages 2, 19

[15] I. Kharroubi, J. Ma, H. Pham, and J. Zhang, "Backward sdes with constrained jumps and quasi-variational inequalities," *The Annals of Probability*, vol. 38, no. 2, pp. 794âĂŞ840, 2010. → pages 2

[16] L. Rogers and S. Singh, "The cost of illiquidity and its effects on hedging," *Mathematical Finance*, vol. 20, no. 4, pp. 597–615, 2010. → pages 2

[17] Bernt Øksendal and Agnes Sulem, *Applied Stochastic Control of Jump Diffusions*, Springer, 2007. → pages 2

[18] Nicola Bruti-Liberati, *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, Finance Discipline Group, UTS Business School, University of Technology, Sydney, 2007. → pages 2

[19] M. Davis, X. Guo, and G. Wu, "Impulse control of multidimensional jump diffusions.," *SIAM J. Control and Optimization*, vol. 48, no. 8, pp. 5276–5293, 2010. → pages 2

[20] A. Bensoussan, R. Liu, and S. Sethi, "Optimality of an (s, s) policy with compound poisson and diffusion demands: A quasi-variational inequalities approach," *SIAM Journal on Control and Optimization*, vol. 44, no. 5, pp. 1650–1676, 2005. → pages 2

[21] R. Bass, "Stochastic differential equations with jumps. probability survey," *Trans. Amer. Math. Soc*, pp. 1–19, 2004. → pages 2

[22] P. Glasserman and N. Merener, "Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities," in *Proceedings of the Royal Society: Series A*, 2001, pp. 111–127. → pages 2

[23] R. Seydel, *Impulse Control for Jump Diffusions: Viscosity Solutions of Quasi Variational Inequalities and Applications in Bank Risk Management*, 2010. → pages 2

[24] J. Spall, *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, Wiley, 2003. → pages 3, 13, 39

[25] H. Ishii M. Crandall and P. Lions, "User's guide to viscosity solutions of second order partial differential equations," *Bulletin of the American Mathematical Society*, vol. 27, no. 12, pp. 1–69, 1992. → pages 12

[26] L. Prashanth S. Bhatnagar, H. Prasad, *Stochastic Recursive Algorithms for Optimization*, Springer, first edition, 2013. → pages 13

[27] J. Chancelier, B. Øksendal, and A. Sulem, "Combined stochastic control and optimal stopping, and application to numerical approximation of combined stochastic and impulse control," *Stochastic Financial Mathematics*, pp. 149–172, 2002. → pages 15, 30, 34

[28] M. Jose Junca Pelaez, "Optimal execution strategy: Price impact and transaction cost," 2011. → pages 19

[29] V. Zakamouline, "Optimal portfolio selection with both fixed and proportional transaction costs for a crra investor with finite horizon," in *Discussion Paper 2002, Institute of Finance and Management Science. Norwegian School of Economics and Business Administration*, 2002. → pages 19

[30] M. Gaigi, V. Vath, M. Mnif, and S. Toumi, "Numerical approximation for a portfolio optimization problem under liquidity risk and costs," 2013. → pages 23, 24, 25, 28, 30, 56, 57

[31] I. Kharroubi and H. Pham, "Optimal portfolio liquidation with execution cost and risk," *SIAM Journal of Financial Mathematics*, vol. 1, pp. 897–931, 2010. → pages 24

[32] Zhuliang Chen and Peter A. Forsyth, "A numerical scheme for the impulse control formulation for pricing variable annuities with a guaranteed minimum withdrawal benefit (gmwb)," 2008. → pages 30

[33] B. Øksendal, *Stochastic Differential Equations - An Introduction with Applications*, Springer, 2003. → pages 35

[34] V. Krishnamurthy, "Bayesian sequential detection with phase-distributed change time and nonlinear penalty - a POMDP lattice programming approach.," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 7096–7124, 2011. → pages 39

[35] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Applications of mathematics. Springer, 2003. → pages 40

[36] V. Ly Vath, M. Mnif, and H. Pham, "A model of optimal portfolio selection under liquidity risk and price impact," *Finance and Stochastics*, vol. 11, no. 1, pp. 51–90, 2007. → pages 55, 60

[37] M. Ngo and V. Krishnamurthy, "Optimality of threshold policies for transmission scheduling in correlated fading channels," *Transactions on Communications*, vol. 57, no. 8, pp. 2474–2483, Aug. 2009. → pages 63

Appendix A

Proofs

A.1 Proof of Theorem 4.2.1

Proof. Given $l \in \mathbb{N}\backslash\{0\}$, from the definition of $\mathcal{A}_l(x,y,p)$ in (4.18),

$$\mathcal{A}_l(x,y,p) \subset \mathcal{A}_{l+1}(x,y,p) \subset \mathcal{A}(x,y,p).$$

The above equation implies that the corresponding value functions are monotone increasing in l

$$v_l(t,x,y,p) \leq v_{l+1}(t,x,y,p) \leq v(t,x,y,p).$$

So in the limit

$$\lim_{l\to\infty} v_l(t,x,y,p) \leq v(t,x,y,p). \tag{A.1}$$

Next, choose $\epsilon > 0$ and $\alpha = \{\tau_1, \tau_2, ...; \zeta_1, \zeta_2, ...) \in \mathcal{A}(x,y,p)$ such that, from (3.8)

$$\mathbb{E}(x,y,p)\left[\sum_{\tau_i \leq T} c(P_{\tau_i^-}, \zeta_i)\right] = v^\alpha(t,x,y,p) \geq v(t,x,y,p) - \epsilon. \tag{A.2}$$

With P_t^α diffusing under impulse control α. Next choose l s.t.

$$\alpha_l = (\tau_1, \zeta_1, ..., \tau_{l-1}, \zeta_{l-1}, \bar{\tau}),$$

where $\tau_{l-1} < \bar{\tau} < \min\{\tau_l, T\}$. So we have the impulse control $\alpha_l \in \mathcal{A}_l(x,y,p)$ and the price process $P_t^{\alpha_l}$.
Next consider the following limit, where the inequality is given by Fatou's lemma:

$$\begin{aligned}\liminf_{l\to\infty} \mathbb{E}\left[\sum_{\tau_l \le T} c(P^{\alpha_l}_{\tau_i^-}, \zeta_i)\right] &\ge \mathbb{E}\left[\liminf_{l\to\infty} \sum_{\tau_l \le T} c(P^{\alpha_l}_{\tau_i^-}, \zeta_i)\right] \\ &= \mathbb{E}\left[\sum_{\tau_l \le T} c(P^{\alpha}_{\tau_i^-}, \zeta_i)\right] \end{aligned} \tag{A.3}$$

Combining (A.2) and (A.3) gives

$$\lim_{l\to\infty} v_l(t,x,y,p) \ge \liminf_{l\to\infty} \mathbb{E}\left[\sum_{\tau_l \le T} c(P^{\alpha_l}_{\tau_i^-}, \zeta_i)\right] \ge v(t,x,y,p) - \epsilon,$$

when combined with result (A.1), since ϵ is chosen arbitrarily, gives the desired result

$$\lim_{l\to\infty} v_l(t,x,y,p) = v(t,x,y,p).$$

□

A.2 Proof of Theorem 4.2.2

Before begining the proof outline, in [36], the authors show that the value function corresponding to (3.11) belong to a set of functions with the growth condition:

$$\mathcal{G}([0,T]\times\bar{\mathcal{S}}) = \left\{ v : [0,T]\times\bar{\mathcal{S}} \to \mathbb{R} : \sup_{[0,T]\times\bar{\mathcal{S}}} \frac{|v(t,x,y,p)|}{1+(x+\frac{p}{\lambda})} < \infty \right\},$$

where $\lambda \in [0,1)$ is a constant found in the instantaneous reward function $c(\zeta,p) = K(\zeta,p)^\gamma$, for $K > 0$. We can also assume such a reward function stucture, and assume it holds throughout the sketch as Assumption 3.0.2 is maintained.

Consider the approximating scheme to HJBQVI (3.11), with time discretization from (4.1):

$$S^h(t,x,y,p,v^h(t,x,y,p),v^h) = 0, \quad (t,x,y,p) \in [0,T]\times\bar{\mathcal{S}}, \tag{A.4}$$

where $S^h : [0,T] \times \bar{\mathcal{S}} \times \mathcal{G}([0,T]) \times \bar{\mathcal{S}}) \to \mathbb{R}$ is defined by

$$
\begin{aligned}
&S^h(t,x,y,p,z,\phi) \\
&\equiv \begin{cases} \max[z - \mathbb{E}[\phi(t+h, X_{t+h}^{0,t,x,y,p}, Y_{t+h}^{0,t,x,y,p}, P_{t+h}^{0,t,x,y,p})], z - \mathcal{M}\phi(t,x,y,p)] & \text{if } t \in [0, T-h] \\ \max[z - \mathbb{E}[\phi(T, X_T^{0,t,x,y,p}, Y_T^{0,t,x,y,p}, P_T^{0,t,x,y,p})], z - \mathcal{M}\phi(t,x,y,p)] & \text{if } t \in (T-h, T) \\ \max[r - C_T(Y_T, P_T), z - \mathcal{M}\phi(t,x,y,p)] & \text{if } t = T. \end{cases}
\end{aligned}
\tag{A.5}
$$

The notation $\Xi_{t'}^{0,t,x,y,p}$ indicates the value of a state process $\Xi_t(\omega)$ at time t' when starting at time t in state (x,y,p) with no control action applied.

Scheme (A.5) is formulated as a backward scheme for v^h as follows:

$$
\begin{aligned}
&v^h(T,x,y,p) = \max[C_T(y,p), \mathcal{M}v(T,x,y,p)], \quad \text{for } t = T, \\
&v^h(t,x,y,p) = \max[\mathbb{E}[v^h(t+h, X_T^{0,t,x,y,p}, Y_T^{0,t,x,y,p}, P_T^{0,t,x,y,p})], \mathcal{M}v^h(t,x,y,p)], \text{ for } t \in [0, T-h], \\
&v^h(t,x,y,p) = v^h(T-h,x,y,p), \quad \text{for } t \in (T-h, T).
\end{aligned}
\tag{A.6}
$$

This approximating scheme, as noted in [30], is a priori implicity due to the non local obstacle term $\mathcal{M}$. The cure is to iterate the scheme as a sequence of optimal stopping problems as follows:

$$
\begin{aligned}
&v^{h,n+1}(T,x,y,p) = \max[C_T(y,p), \mathcal{M}v(T,x,y,p)] \quad \text{for } t = T, \\
&v^{h,n+1}(t,x,y,p) = \max[\mathbb{E}[v^{h,n+1}(t+h, X_{t+h}^{0,t,x,y,p}, Y_{t+h}^{0,t,x,y,p}, P_{t+h}^{0,t,x,y,p})], \mathcal{M}v^{h,n}(t,x,y,p)], \text{ for } t \in [0, T-h], \\
&v^{h,n+1}(t,x,y,p) = v^{h,n+1}(T-h,x,y,p), \quad \text{for } t \in (T-h, T).
\end{aligned}
\tag{A.7}
$$

Next, considering the state and action discretizations (4.8) and (4.9), respectively, along with the iterated optimal stopping problem treatment of scheme (A.5) (namely, (A.7)) and the approximation of the expected value arising in the scheme given by

(4.13), [30] states that the approximating scheme (A.5) becomes

$$
\begin{aligned}
&S^{h,R,N,M}(t,x,y,p,z,\phi)\\
&\equiv\begin{cases}
\max[z-\mathcal{E}[\phi(t+h,X_{t+h}^{0,t,x,y,p},Y_{t+h}^{0,t,x,y,p},P_{t+h}^{0,t,x,y,p})],z-\mathcal{M}^{M,R}\phi(t,x,y,p)] & \text{if } t\in[0,T-h]\\
\max[z-\mathcal{E}[\phi(T,X_{T}^{0,t,x,y,p},Y_{T}^{0,t,x,y,p},P_{T}^{0,t,x,y,p})],z-\mathcal{M}^{M,R}\phi(t,x,y,p)] & \text{if } t\in(T-h,T)\\
\max[r-C_T(Y_T,P_T),z-\mathcal{M}^{M,R}\phi(t,x,y,p)] & \text{if } t=T.
\end{cases}
\end{aligned}
\tag{A.8}
$$

The iterative optimal stopping problem scheme, combined with the discrete state, action and time scheme (A.8) yield the following inductive scheme

$$
\begin{aligned}
&v^{h,n+1}(t_m,x,y,p)=\max\left[C_T(y,p),\sum_{\zeta\in\mathcal{C}_{M,R}(x,y,p)}v(T,\Gamma(x,y,p,\zeta))\right]\\
&v^{h,n+1}(t_i,x,y,p)\\
&=\max\left[\mathcal{E}^{N,R}[v^{h,n+1}(t_{i+1},X_{t_{i+1}}^{0,t,x,y,p},Y_{t_{i+1}}^{0,t,x,y,p},P_{t_{i+1}}^{0,t,x,y,p})],\sup_{\zeta\in\mathcal{C}_{M,R}(x,y,p)}v^{h,n}(t_i,\Gamma(x,y,p,\zeta))\right]
\end{aligned}
\tag{A.9}
$$

where $\mathcal{M}^{M,R}$ is given by (4.16). The development of this numerical scheme is the main focus of [30], and the proof of the convergence of the value function $v^{h,n}$ from (A.9) to the value function v which solves the HJBQVI (3.11) is accomplished when (A.9) satisfies monotonicity, stability, and consistency properties. Since this proof is explained in thourough detail in [30], only the propositions will be provided, without proof.

Proposition A.2.1. *(Monotonicity) For all* $h>0,(t,x,y,p)\in[0,T]\times\bar{\mathcal{S}},g\in\mathbb{R},$ *and* $\phi,\psi\in\mathcal{G}([0,T])$ *with* $\phi\le\psi$,

$$S^{h,R,N,M}(t,x,y,p,\phi)\ge S^{h,R,N,M}(t,x,y,p,\psi).$$

Proposition A.2.2. *(Consistency) Take* $N=\frac{1}{h^q}$ *s.t.* $q>2$.

1. *Let* $Z_{t'+h}^{0,t',x,y,p}=(X_{t'+h}^{0,t',x,y,p},Y_{t'+h}^{0,t',x,y,p},P_{t'+h}^{0,t',x,y,p})$ *and* $z=(x,y,p)$. *For all* $(t,x,y,p)\in$

$[0,T)\times\bar{\mathcal{S}}$ *and Lipschitz function* $\phi\in C^{1,2}([0,T)\times\bar{\mathcal{S}})$*:*

$$\limsup_{\substack{(h,t',z')\to(0,z)\\(M,N,R)\to+\infty}}\max\left\{\frac{\phi(t',z')-\mathcal{E}^{N,R}[\phi(t'+h,Z_{t'+h}^{0,t',x,y,p})]}{h},\Big(\phi(t',z')-\mathcal{M}^{M,R}\phi(t',z')\Big)\right\}$$
$$\leq\max\left\{(\frac{\partial\phi}{\partial t}+\mathcal{L}\phi)(t,z),\mathcal{M}\phi(t,z)-\phi(t,z)\right\}$$

and

$$\liminf_{\substack{(h,t',z')\to(0,z)\\(M,N,R)\to+\infty}}\max\left\{\frac{\phi(t',z')-\mathcal{E}^{N,R}[\phi(t'+h,Z_{t'+h}^{0,t',x,y,p})]}{h},\Big(\phi(t',z')-\mathcal{M}^{M,R}\phi(t',z')\Big)\right\}$$
$$\geq\max\left\{(\frac{\partial\phi}{\partial t}+\mathcal{L}\phi)(t,z),\mathcal{M}\phi(t,z)-\phi(t,z)\right\}$$

2. *For all* $z=(x,y,p)\in\bar{\mathcal{S}}$ *and Lipschitz function* $\phi\in C^{1,2}([0,T]\times\bar{\mathcal{S}})$

$$\limsup_{\substack{(h,t',z')\to(0,z)\\(M,N,R)\to+\infty}}\max\Big\{\phi(t',z')-C_T(y',p'),\mathcal{M}^{M,R}C_T(y',p')\Big\}$$
$$\leq\max\left\{\phi(T,z)-C_T(y,p),\mathcal{M}\phi(T,z)-\phi(T,z)\right\}$$

and

$$\liminf_{\substack{(h,t',z')\to(0,z)\\(M,N,R)\to+\infty}}\max\Big\{\phi(t',z')-C_T(y',p'),\mathcal{M}^{M,R}C_T(y',p')\Big\}$$
$$\geq\max\left\{\phi(T,z)-C_T(y,p),\mathcal{M}\phi(T,z)-\phi(T,z)\right\}.$$

Lemma A.2.1. *For all* $(t,x,y,p)\in\mathbb{T}_m\times\bar{\mathcal{S}}_{loc}$*,*

$$\lim_{n\to+\infty}v_n^{h,R,N,M}(t,x,y,p)=v^{h,R,N,M}(t,x,y,p)$$

Proposition A.2.3. *(Iteratated optimal stopping problems) Define* $\phi_n^{h,R,N,M}$ *iteratively as a sequence of optimal stopping problems*

$$\phi_{n+1}^{h,R,N,M}(t,x,y,p)=\sup_{\tau\in S_{t,T}^h}\mathcal{E}^{N,R}[\mathcal{M}^{M,R}\phi_n^{h,R,N,M}(\tau,Z_{N,R}^{0,t,x,y,p}(\tau))]\quad\forall(t,x,y,p)\in\mathbb{T}_m\times\bar{\mathcal{S}}_{loc}$$
$$\phi_0^{h,R,N,M}(t,x,y,p)=v_0^{h,R,N,M}(t,x,y,p)\quad\text{ for all }(t,x,y,p)\in\mathbb{T}_m\times\bar{\mathcal{S}}_{loc}.$$

where $S^h_{t,T}$ is the set of all $\mathcal{F}_s$-stopping times in $[t,T], t \le s \le T$, with time discretization h. Given $(t,x,y,p) \in \mathbb{T}_m \times \bar{\mathcal{S}}_{loc}$, it is shown that

$$\phi_n^{h,R,N,M}(t,x,y,p) = v_n^{h,R,N,M}(t,x,y,p).$$

Corollary A.2.1. *Given $(t,x,y,p) \in \mathbb{T}_m \times \bar{\mathcal{S}}_{loc}$,*

$$\lim_{n \to +\infty} \phi_n^{h,R,N,M}(t,x,y,p) = v^{h,R,N,M}(t,x,y,p).$$

Proposition A.2.4. *(Stability) $\forall h > 0, \exists!$ solution $v_n^{h,R,N,M}(t,x,y,p) \in \mathcal{G}([0,T] \times \bar{\mathcal{S}})$ to the approximating scheme* (A.5) *and the sequence $(v_n^{h,R,N,M})_h$ is uniformly bounded in $\mathcal{G}([0,T] \times \bar{\mathcal{S}})$, i.e., $\exists w \in \mathcal{G}([0,T] \times \bar{\mathcal{S}})$ such that $|v_n^{h,R,N,M}| \le w$ for all $h > 0$.*

Proposition A.2.5. *1. Let $\mathcal{O} \subset \bar{\mathcal{S}}$. A locally bounded function $\bar{v}$ on $[0,T) \times \bar{\mathcal{S}}$ is a viscosity subsolution (resp. $\underline{v}$ is a supersolution) of* (3.11) *in $[0,T) \times \mathcal{O}$ if for all $(\bar{t},\bar{x},\bar{y},\bar{p}) \in [0,T) \times \mathcal{O}$ and $\phi \in C^{1,2}([0,T) \times \bar{\mathcal{S}}$ s.t. $(\bar{v} - \phi)(\bar{t},\bar{x},\bar{y},\bar{p}) = 0$ (resp. $(\underline{v} - \phi)(\underline{t},\bar{x},\bar{y},\bar{p}) = 0$) and $(\bar{t},\bar{x},\bar{y},\bar{p})$ is a maximum of $\bar{v} - \phi$ (resp. a minimum of $\underline{v} * -\phi$) on $[0,T) \times \mathcal{O}$. Then*

$$\max\left\{ \frac{\partial \phi}{\partial t}(\bar{t},\bar{x},\bar{y},\bar{p}) + \mathcal{L}\phi(\bar{t},\bar{x},\bar{y},\bar{p}), \mathcal{M}\bar{v}(\bar{t},\bar{x},\bar{y},\bar{p}) - \bar{v}(\bar{t},\bar{x},\bar{y},\bar{p}) \right\} \le 0, \qquad \text{in } [0,T) \times \mathcal{O}, \tag{A.10}$$

$$\max\left\{ \frac{\partial \phi}{\partial t}(\bar{t},\bar{x},\bar{y},\bar{p}) + \mathcal{L}\phi(\bar{t},\bar{x},\bar{y},\bar{p}), \mathcal{M}\underline{v}(\bar{t},\bar{x},\bar{y},\bar{p}) - \underline{v}(\bar{t},\bar{x},\bar{y},\bar{p}) \right\} \ge 0, \qquad \text{in } [0,T) \times \mathcal{O}, \tag{A.11}$$

2. A locally bounded function u on $[0,T) \times \bar{\mathcal{S}}$ is a constrained viscosity solution of (3.11) *in $[0,T) \times \mathcal{S}$ if u is a viscosity subsolution of* (3.11) *in $[0,T) \times \bar{\mathcal{S}}$ and a viscosity supersolution of* (3.11) *in $[0,T) \times \mathcal{S}$.*

Proposition A.2.6. *(Convergence) For all $(t',x',y',p') \in \mathbb{T}_m \times \mathcal{Z}_l$ and $(t,x,y,p) \in [0,T) \times \mathcal{S}$,*

$$\lim_{\substack{(t',x',y',p') \to (t,x,y,p) \\ (h,M,N,R) \to +(0,+\infty,+\infty,+\infty)}} v^{h,R,N,M}(t',x',y',p') = v(t,x,y,p),$$

where $v^{h,R,N,M}(t',x',y',p')$ is the solution to the approximating scheme (A.6)*, and $v(t,x,y,p)$ is the solution the the HJBQVI* (3.11)

Proof. A sketch of this proof is given as follows, using the propositions determined thus far:

1. Assume $\bar{v}(t,x,y,p)$ and $\underline{v}(t,x,y,p)$ are sub- and supersolutions respectively of the value function $v(t,x,y,p)$ solving (3.11). These solutions are defined as follows:

$$\begin{aligned}\bar{v}(t,x,y,p) &= \limsup_{\substack{(t',x',y',p')\to(t,x,y,p)\\(h,M,N,R)\to+(0,+\infty,+\infty,+\infty)}} v^{h,R,N,M}(t',x',y',p')\\ \underline{v}(t,x,y,p) &= \liminf_{\substack{(t',x',y',p')\to(t,x,y,p)\\(h,M,N,R)\to+(0,+\infty,+\infty,+\infty)}} v^{h,R,N,M}(t',x',y',p').\end{aligned} \tag{A.12}$$

2. Since $\bar{v}(t,x,y,p)$ is upper semicontinuous and $\underline{v}(t,x,y,p)$ is lower semicontinuous, the comparison principle shown in [36] gives $\bar{v}(t,x,y,p) \leq \underline{v}(t,x,y,p)$ for all $(t,x,y,p) \in [0,T]\times\mathcal{S}$. However, from (A.12) it is clear that $\bar{v}(t,x,y,p \geq \underline{v}(t,x,y,p)$ for all $(t,x,y,p)\in[0,T]\times\mathcal{S}$, so it must be the case that $\bar{v}(t,x,y,p = \underline{v}(t,x,y,p) = v(t,x,y,p)$, the unique continuous solution of (3.11), and as such $v^{h,R,N,M}$ converges uniformly to v.

3. Consider $\bar{v}$. Let $(\bar{t},\bar{x},\bar{y},\bar{p})$ strict global maximum of $\bar{v}-\phi$ on $[0,T]\times\bar{\mathcal{S}}$, where $\phi\in C^{1,2}([0,T]\times\bar{\mathcal{S}}$, such that $\bar{v}(t,x,y,p)-\phi(t,x,y,p)\leq\bar{v}(\bar{t},\bar{x},\bar{y},\bar{p})-\phi(\bar{t},\bar{x},\bar{y},\bar{p}) = 0$ on $[0,T]\times\bar{\mathcal{S}}$.

4. Through the combination of the stability, monotonicity, and consistency properties shows that for the subsolution $\bar{v}$:

$$0 \leq \max\left[\frac{\partial\phi}{\partial t}(\bar{t},\bar{x},\bar{y},\bar{p}) + \mathcal{L}\phi(\bar{t},\bar{x},\bar{y},\bar{p}), \mathcal{M}\phi(\bar{t},\bar{x},\bar{y},\bar{p}) - \phi(\bar{t},\bar{x},\bar{y},\bar{p})\right].$$

   and likewise that for the supersolution $\underline{v}$:

$$0 \geq \max\left[\frac{\partial\phi}{\partial t}(\bar{t},\bar{x},\bar{y},\bar{p}) + \mathcal{L}\phi(\bar{t},\bar{x},\bar{y},\bar{p}), \mathcal{M}\phi(\bar{t},\bar{x},\bar{y},\bar{p}) - \phi(\bar{t},\bar{x},\bar{y},\bar{p})\right],$$

which, by Propostion A.2.5, shows that $\bar{v}$ and $\underline{v}$ are true sub- and supersolutions to (3.11), and when combined with the convergence of the approximating scheme $v^{h,R,N,M}$ to both sub- and supersolutions, shows that the approximating scheme does indeed converge to the solution of HJBQVI (3.11)

This completes the proof of Theorem 4.2.2 □

A.3 Proof of Lemma 4.3.1

Proof. The proof is by induction on n. Because $v_T(x,y,p) = C_T(y,p)$ is concave in y and p by Assumption 3.2.1, assume that $v_i(x,y,p)$ is concave in y and p for $i = T-1,...,n+1$.

In order to show that $v_n(\cdot,\cdot,\cdot)$ is concave in y, it is required that for $0 < \lambda < 1$,

$$v_n(x, \lambda y_1 + (1-\lambda)y_2, p) \geq \lambda v_n(x, y_1, p) + (1-\lambda) v_n(x, y_2, p).$$

Now, for some $\alpha_1 \leq y_1$ and $\alpha_2 \leq y_2$,

$$\begin{aligned} v_n(x, y_1, p) &= c(\alpha_1, p) + \bar{v}_{n+1}(x, y_1 - \alpha_1, p - m(\alpha_1)), \\ v_n(x, y_2, p) &= c(\alpha_2, p) + \bar{v}_{n+1}(x, y_2 - \alpha_2, p - m(\alpha_2)). \end{aligned}$$

However, because $\lambda\alpha_1 + (1-\lambda)\alpha_2 \leq \lambda y_1 + (1-\lambda)y_2$, it follows from (4.27) that

$$\begin{aligned} v_n(x, \lambda y_1 + (1-\lambda)y_2), p) &\geq c(\lambda\alpha_1 + (1-\lambda)\alpha_2, p - m(\lambda\alpha_1 + (1-\lambda)\alpha_2)) \\ &\quad + \bar{v}_{n+1}(x, \lambda(y_1 - \alpha_1) + (1-\lambda)(y_2 - \alpha_2), p - m(\lambda\alpha_1 + (1-\lambda)\alpha_2)) \\ &\geq \lambda c(\alpha_1, p) + (1-\lambda) c(\alpha_2, p) + \lambda \bar{v}_{n+1}(x, y_1 - \alpha_1, p - m(\alpha_1)) \\ &\quad + (1-\lambda)\bar{v}_{n+1}(x, y_2 - \alpha_2, p - m(\alpha_2)) \\ &= \lambda v_n(x, y_1, p) + (1-\lambda) v_n(x, y_2, p), \end{aligned}$$

where the second inequality follows from the concavity of $c(\cdot,\cdot)$, and the concavity of $\bar{v}_{n+1}(\cdot,\cdot,\cdot)$, the latter of which follows from the induction hypothesis. Therefore $v_n(x,y,p)$ is concave in y. □

A.4 Proof of Theorem 4.3.1

Proof. To prove (i), let $\bar{\alpha} = \alpha_n(x,y,p)$. Then, by its definition, it follows that for $\alpha < \bar{\alpha}$

$$c(\bar{\alpha},p) + \bar{v}_{n+1}(x,y-\bar{\alpha},p-m(\bar{\alpha})) > c(\alpha,p) + \bar{v}_{n+1}(x,y-\alpha,p-m(\alpha)). \tag{A.13}$$

For $\epsilon > 0$, it is shown next that $\alpha_n(x,y+\epsilon,p) \geq \bar{\alpha}$ by proving that whenever $\alpha < \bar{\alpha}$,

$$c(\bar{\alpha},p) + \bar{v}_{n+1}(x,y+\epsilon-\bar{\alpha},p-m(\bar{\alpha})) \geq c(\alpha,p) + \bar{v}_{n+1}(x,y+\epsilon-\alpha,p-m(\alpha)).$$

From (A.13) this will be proven if it can be shown that whenever $\alpha < \bar{\alpha}$,

$$\begin{aligned}\bar{v}_{n+1}(x,y-\alpha,p-m(\alpha)) - \bar{v}_{n+1}(x,y-\bar{\alpha},p-m(\bar{\alpha})) \\ \geq \bar{v}_{n+1}(x,y+\epsilon-\alpha,p-m(\alpha)) - \bar{v}_{n+1}(x,y+\epsilon-\bar{\alpha},p-m(\bar{\alpha})).\end{aligned}$$

This follows from the concavity of v_n in y, which implies the concavity of $\bar{v}_{n+1}$. Therefore (i) is proven.

To prove that $\alpha_n(x,y,p) \geq \alpha_{n-1}(x,y,p)$, let $\bar{\alpha} = \alpha_n(x,y,p)$. If $\bar{a} = y$, then the result is immediate; so suppose that $\bar{\alpha} < y$. Fix $\epsilon < y - \bar{\alpha}$ and let $\alpha^* = \alpha_n(x,y-\bar{a}-\epsilon,p-m(\bar{\alpha}))$. Note that by part (i) of Thm. 4.3.1, it follows that $\alpha^* \leq \bar{\alpha}$. Now,

$$\begin{aligned}&c(\bar{\alpha}+\epsilon,p) + v_n(x,y-\bar{\alpha}-\epsilon,p-m(\bar{\alpha}+\epsilon)) \\ &\quad = c(\bar{\alpha}+\epsilon,p) + c(\alpha^*,p) \\ &\qquad + \bar{v}_{n+1}(x,y-\alpha-\epsilon-\alpha^*,p-m(\alpha+\epsilon)-m(\alpha^*)) \\ &\quad \leq c(\bar{\alpha},p) + c(\alpha^*+\epsilon,p) + \bar{v}_{n+1}(x,y-\bar{\alpha}-\epsilon-\alpha^*,p-m(\bar{\alpha}+\alpha^*+\epsilon)) \\ &\quad \leq c(\bar{\alpha},p) + v_n(x,y-\bar{\alpha},p-m(\bar{\alpha})),\end{aligned} \tag{A.14}$$

where the first inequality follows from the concavity of c in y and p and the fact

that $\alpha^* \leq \bar{\alpha}$. By using the relationship

$$\bar{v}_n(x,y,p) = qv_n(x,y,p) + (1-q)\bar{v}_{n+1}(x,y,p),$$

it follows that

$$c(\bar{\alpha}+\epsilon,p) + \bar{v}_n(x,y-\bar{\alpha}-\epsilon,p-m(\bar{\alpha}+\epsilon)) - c(\bar{\alpha},p) - \bar{v}_n(x,y-\bar{\alpha},p-m(\bar{\alpha}))$$

$$= q\Big[c(\bar{\alpha}+\epsilon,p) + v_n(x,y-\bar{\alpha}-\epsilon,m(\bar{\alpha}+\epsilon)) - c(\bar{\alpha},p) - v_n(x,y-\bar{\alpha},p-m(\bar{\alpha}))\Big]$$

$$+ (1-q)\Big[c(\bar{\alpha}+\epsilon,p) + \bar{v}_{n+1}(x,y-\bar{\alpha}-\epsilon,m(\bar{\alpha}+\epsilon)) - c(\bar{\alpha},p) - \bar{v}_{n+1}(x,y-\bar{\alpha},p-m(\bar{\alpha}))\Big].$$

The first term in the square brackets is non-positive by (A.14), and the second is non-positive because $\bar{a} = a_n(Y)$. Therefore

$$c(\bar{\alpha}+\epsilon,p) + \bar{v}_n(x,y-\bar{\alpha}-\epsilon,p-m(\bar{\alpha}+\epsilon)) \leq c(\bar{\alpha},p) + \bar{v}_n(x,y-\bar{\alpha},p-m(\bar{\alpha})),$$

implying that

$$\alpha_{n-1}(x,y,p) \leq \bar{\alpha} = \alpha_n(x,y,p).$$

The proof of (iii.) does not utilize the sequential allocation methodology used above in the proofs of (i.) and (ii.), since the price state variable p differs from shares y and time index n in that it is stochastic in nature. Therefore we make an argument involving the supermodularity adapted from [37] of the state-action reward function given in (4.24) in variables (a,p).

Definition A.4.1. *(Supermodularity)*
A function $F(x,y): X \times Y \to \mathbb{R}$ is supermodular *in $F(x_1,y_1)+F(x_2,y_2) \geq F(x_1,y_2)+F(x_2,y_1)$ for all $x_1,x_2 \in X$ and $y_1,y_2 \in Y$, such that $x_1 > x_2, y_1 > y_2$. Similarily, if the inequality is reversed, the function $F(\cdot,\cdot)$ is called submodular.*

The methodology to the proof is given in two steps:

1. *Monotonicity* of the optimal reward function $v_n(\cdot,\cdot,\cdot)$ in p

2. *Supermodularity*: show that the state-action reward function $Q_n(x,y,p,a)$ is

supermodular in (a,p) using mathematical induction. The monotonicity and therefore the nondecreasing structure of the optimal liquidation policy $\alpha_n^*(\cdot,\cdot,\cdot)$ given in (4.23) follows.

Step 1 is shown in the following lemma.

Lemma A.4.1. *The optimal cost to go function $v_n(x,y,p)$ defined by (4.22) is increasing in the price of the asset p.*

Proof. It is shown that $v_n(x,y,p)$ is monotone in p by induction. $V_T(x,y,p)$ is increasing in p since $C(\cdot,\cdot)$ is increasing from Assumption 3.2.1. From the definition of $v_n(x,y,p)$ given by (4.22) and (4.24), the monotonicity of $v_n(x,y,p)$ in p follows immediately by induction. □

The next theorem outlines Step 2.

Theorem A.4.1. *If the terminal reward function $C(\cdot,\cdot)$ is an increasing function in p and is integer convex such that*

$$C(y,p+2)-C(y,p+1)\geq C(y,p+1)-C(y,p)\quad \forall y,p\geq 0, \tag{A.15}$$

then the state-action reward fucntion $Q_n(x,y,p,a)$ is supermodular in (a,p), i.e.

$$Q_n(x,y,p,a)-Q_n(x,y,p,0)\leq Q_n(x,y,p+1,a)-Q_n(x,y,p+1,0),\quad \forall a\leq 0. \tag{A.16}$$

and as a result, the optimal liquidation policy $\alpha_n^*(x,y,p)$ is nondecreasing with respect to the asset price p.

Proof. First, rewrite the state-action reward function $Q_n(x,y,p,a)$ given by (4.24) when considering (4.28) such that

$$\begin{aligned}
Q_n(x,y,p,a) &= c(p,a)+\sum_{p'\in\mathcal{Z}_l\cap\mathcal{P}}\mathbb{P}(p'|p)\\
&\times\Big[f_m(a,p)v_{n+1}(x+pa,y-a,p'-m(a))+(1-f_m(a,p))v_{n+1}(x+pa,y-a,p')\Big]\\
&= c(p,a)+\sum_{p'\in\mathcal{Z}_l\cap\mathcal{P}}\mathbb{P}(p'|p)\times(v_{n+1}(x+pa,y-a,p'-m(a))-v_{n+1}(x+pa,y-a,p'))\\
&+\sum_{p'\in\mathcal{Z}_l\cap\mathcal{P}}\mathbb{P}(p'|p)v_{n+1}(x+pa,y-a,p').
\end{aligned} \tag{A.17}$$

From (A.17) we can see that $Q_n(x,y,p,a)$ is supermodular in (a,p) iff $v_n(x,y,p)$ has increasing differences in p. The following lemma completes the proof of (iii.)

Lemma A.4.2. *If the terminal reward function $C(\cdot,\cdot)$ is an increasing function and satisfies* (A.15) *then the optimal value function $v_n(x,y,p)$ given by* (4.22) *has increasing differences in the price state state:*

$$v_n(x,y,p+2)-v_n(x,y,p+1) \geq v_n(x,y,p+1)-v_n(x,y,p), \tag{A.18}$$

for all $x,y,p \in \bar{\mathcal{S}}_{loc}$. The result is that the state-action reward function $Q_n(x,y,p,a)$ satisfies (A.16)*, and as such is supermodular in (a,p).*

□

Proof of (A.18) by mathematical induction: First notice that (A.18) holds for $n=T$ due to (A.15). Assume (A.18) holds for $n=k$. We prove that (A.18) holds for $n=k-1$. Let $v_{k-1}(x,y,p+2)=Q_{k-1}(x,y,p+2,a_2), v_{k-1}(x,y,p+1)=Q_{k-1}(x,y,p+1,a_1), v_{k-1}(x,y,p)=Q_{k-1}(x,y,p,a_0)$ for some $a_2,a_1,a_0 \in \mathcal{C}_{M,R}$ which are optimal controls at time $k-1$ for states $(x,y,p+2),(x,y,p+1)$, and (x,y,p) respectively. We must show that

$$Q_{k-1}(x,y,p+2,a_2)-Q_{k-1}(x,y,p+1,a_1)-Q_{k-1}(x,y,p+1,a_1)+Q_{k-1}(x,y,p,a_0) \leq 0$$

$$\Updownarrow$$

$$\underbrace{Q_{k-1}(x,y,p+2,a_2)-Q_{k-1}(x,y,p+1,a_2)}_{\text{A}}\underbrace{+Q_{k+1}(x,y,p+1,a_2)-Q_{k-1}(x,y,p+1,a_1)}_{\leq 0 \text{ by optimality.}}$$

$$\underbrace{-Q_{k-1}(x,y,p+1,a_1)+Q_{k-1}(x,y,p+1,a_0)}_{\leq 0 \text{ by optimality.}}\underbrace{-(Q_{k+1}(x,y,p+1,a_0)-Q_{k-1}(x,y,p,a_0)}_{\text{B}} \leq 0.$$

It follows from the induction hypothesis that

$$\begin{aligned} A &= \sum_{p' \in \mathcal{Z}_l \cap \mathcal{P}} \mathbb{P}(p'|p)\Big[f_m(a,p)(v_k(x,y,p'+1)-v_k(x,y,p')) \\ &\qquad + (1-f_m(a,p))(v_k(x,y,p'+2)-v_k(x,y,p'+1))\Big] \\ &\leq \sum_{p' \in \mathcal{Z}_l \cap \mathcal{P}} \mathbb{P}(p'|p)(v_k(x,y,p'+1)-v_k(x,y,p')). \end{aligned}$$

Similarly, $B \geq \sum_{p' \in \mathcal{Z}_l \cap \mathcal{P}} \mathbb{P}(p'|p)(v_k(x,y,p+1)-v_k(x,y,p))$, so $A-B \leq 0$. Therefore

$v_n(x,y,p)$ satisfies (A.18), which implies that $Q_n(x,y,p,a)$ is supermodular in (a,p) due to (A.17). As such, part (iii.) of theorem 4.3.1 has been shown. □