MONOTONE OPTIMAL POLICIES IN PORTFOLIO LIQUIDATION PROBLEMS

Daniel Crawford, Vikram Krishnamurthy

The University of British Columbia - Department of Electrical and Computer Engineering

ABSTRACT

This work considers the problem of optimal liquidation of a single risky asset portfolio as a denumerable Markov Decision Processes (MDP) control problem. The model is defined over discrete time, state, and action sets, and the optimal liquidation strategy is the solution to Bellman's equation. It is shown that the optimal strategy is monotone in the number of shares owned , the time remaining to liquidation, and the price of the underlying asset. This structural result can be exploited to estimate the optimal policy via the simultaneous perturbation stochastic approximation (SPSA) algorithm. Therefore, the optimal policy can be estimated without knowledge of the parameters of the model.

Index Terms— optimal portfolio liquidation, market impact, Bellman's equation, monotone policy, concave rewards, supermodularity

1. INTRODUCTION

Optimal portfolio liquidation is an optimal control problem where the interested agent wishes to completely rid herself of a position in a risky asset. 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. 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. When finding the optimal policy of an optimal control problem, it is useful to determine if there exists any predetermined structure of the policy with respect to the system's state variables. In this work, an optimal policy is determined for an optimal liquidation problem and shown to be non-decreasing in the number of risky stock shares owned, time accumulated since the liquidation period has begun, and the price of the asset. This structured result allows the implementation of reinforcement learning techniques to approximate the optimal policy when model state variables are not fully known (see [1]). The portfolio liquidation problem has been studies in great detail in [2] and [3], but never with a structured policy MDP approach with possible machine learning applications.

2. THE 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, [4],[2],[5], over the time interval $t \in [0, T]$, where $T > 0$ is a terminal portfolio liquidation horizon. Consider an impulse control strategy $\alpha = \{\tau_n, \zeta_n\}_{n=0,1,...}$, with each τ_n an $\mathcal{F}_t$-stopping time and ζ_n an $\mathcal{F}_t$-measurable random variable, where $\mathcal{F}_t = \sigma(B_s : 0 \leq s \leq t, \tau_n \leq t)$ for a 1-dimensional Brownian Motion, B_t. Each τ_n indicates a trading impulse instant, and each ζ_n indicates a trade amount. Assume $\zeta_n > 0$ 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.

$$P_t = \int_0^t \mu P_t dt + \int_0^t \sigma P_t dB_t - \sum_{\tau_i \leq T} m(\zeta_i), \qquad (1)$$

where $p > 0$ is a given constant and $P_0 = p$ is a deterministic constant. The market impact function $m(\cdot)$ is a bounded function described in [6]. Assume the following about the price impact function

Assumption 2.1. *(Market Impact)*

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

Well-definedness of (1) follows from the coefficients of (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

$$c(X_{\tau_n^-}, Y_{\tau_n^-}, P_{\tau_n^-}, \zeta_n) = c(P_{\tau_n^-}, \zeta_n). \qquad (2)$$

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

Assumption 2.2. *(Instantaneous reward function.)* $\inf_{|\zeta|\in\mathbb{R}_+} c(\cdot,\zeta) = L > 0$, *and* $c(\cdot,\cdot)$ *is an increasing function in the asset price* P_t *and an increasing concave function in the sell action* ζ.

Assume that the terminal reward function follows $\forall Y_T, P_T \in \mathbb{R}_+, \delta_y, \delta_p \geq 0$:

Assumption 2.3. *(Terminal reward function).*

$$C_T(Y_T+\delta_y, P_T) \geq C_T(Y_T, P_T), \quad C_T(0,P_T)=0, \quad (3)$$
$$C_T(Y_T, P_T+\delta_p) \geq C_T(Y_T, P_T), \quad C_T(Y_T,0)=0.$$

Call the shares process Y_t taking values in $\mathbb{R}_+ \cup \{0\}$. Given the impulse control $\{\tau_n, \zeta_n\}_{n=0,1,...}$, the dynamics of Y_t is given by

$$Y_s = Y_{\tau_n} \text{ for } \tau_n \leq s \leq \tau_{n+1}, Y_{\tau_{n+1}} = Y_{\tau_n} - \zeta_{n+1}, \quad (4)$$

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.

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

$$X_t = X_{\tau_k}, \tau_k \leq t < \tau_{k+1}, \quad n \geq 0. \quad (5)$$

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 $\Delta X_t \equiv X_t - X_{t^-} = -\Delta Y_t$. As such, we clearly ignore illiquidity effects, frequency trading penalties, and bid-ask spreads in this formulation. So for this impulse control we have

$$X_{\tau_n} = X_{\tau_n^-} + \zeta_n P_{\tau_n} = X_{\tau_{n-1}} + \zeta_n P_{\tau_n}, \quad n \geq 1. \quad (6)$$

Consider a liquidation function $L(x,y,p) = x + yp - K(y)$, 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 $t \in [0,T]$. Define the solvency region

$$\bar{\mathcal{S}} = \{(x,y,p) \in \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+ : L(x,y,p) > 0\}. \quad (7)$$

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

Definition 2.1. *(Conditions for an Admissible Impulse Control) The couple process* $\{\zeta_k, \tau_k\}_{k\geq 1}$ *is admissible if the following properties are satisfied:* $\{X_s, Y_s, P_s\}$ *follow state laws* (1), (4), (5), *and* (6) *and remain in* (7) *for all liquidation impulses, for all time* $t \leq s \leq T$, *the state of the system remains in the solvency region* $\bar{\mathcal{S}}$ *for each control pair* $(\tau_i, \zeta_i), 0 \leq \tau_i \leq T$, *and* $0 < \tau_i < \tau_{i+1}, i \geq 1$,

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

Problem 2.1. *(Optimal Liquidation Problem)*

Associate with the above controlled price process (1) *an objective function*

$$v^\alpha(t,x,y,p) = \mathbb{E}^{x,y,p}[\sum_{\tau_i} c(P^\alpha_{\tau_i^-}, \zeta_i)]. \quad (8)$$

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

$$v^*(t,x,y,p) = \max_{\alpha \in \mathcal{A}(x,y,p)} v^\alpha(t,x,y,p) \quad (9)$$

2.1. Discretization Scheme

This section adapts a discrete derivation and convergence analysis from [7]. Consider a time discretization to Problem (2.1) as follows. Let the time step be

$$h = T/m, m \in \mathbb{N}\backslash\{0\} \quad (10)$$

and let $\mathbb{T}_m = \{t_i = ih, i = 0, ..., m\}$ be the uniform gird over the interval $[0,T]$. Let $\Gamma(x,y,p,\zeta) = (x + \zeta p, y - \zeta, p - m(\zeta)) \in \bar{\mathcal{S}}$ and $\mathcal{C}(x,y,p) = \{e \in \mathbb{R}_+ : (\Gamma(x,y,p,e) \in \bar{\mathcal{S}})\}$. For state discretization, define a finite, localized subset of the admissible state space as a uniform grid, denoted by

$$\bar{\mathcal{S}}_{loc} = \bar{\mathcal{S}} \cap \mathcal{X} \times \mathcal{Y} \times \mathcal{P}, \quad (11)$$

where $\mathcal{X} = [x_{min}, x_{max}]$, $\mathcal{Y} = [y_{min}, y_{max}]$, and $\mathcal{P} = [p_{min}, p_{max}]$. Here, $\mathcal{X}$ has increments of size $(x_{max} - x_{min})/N, N \in \mathbb{N}\backslash\{0\}$, and similar for $\mathcal{Y}$, $\mathcal{P}$. Define the regular grid

$$\mathcal{Z}_l = \{(x,y,p) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{P} : (x,y,p) \in \bar{\mathcal{S}}_{loc}\}. \quad (12)$$

Define a similar grid for the admissible controls

$$\mathcal{C}_{M,R}(x,y,p) = \{\zeta_i = \zeta_{min} + \frac{i}{M}(\zeta_{max} - \zeta_{min}) : 0 \leq i \leq M,$$
$$\Gamma(x,y,p,\zeta_i) \in \bar{\mathcal{S}}_{loc}\}, \quad (13)$$

where $\zeta_{min} < \zeta_{max} \in \mathbb{R}_+$ and $M \in \mathbb{N}\backslash\{0\}$ are fixed constants, and $R := \min(|x_{min}|, |x_{max}|, |y_{min}|, |y_{min}|, |p_{max}|)$.

3. DENUMERABLE MDP

Construct the countable state, countable action MDP to approximate Problem (2.1) through emulation of the discrete scheme in Sec. 2.1.

Define the following MDP components for the portfolio liquidation problem:

Definition 3.1. *(Denumerable MDP)*

For $(x,y,p) \in \mathcal{Z}_l$, *Decision Epochs:* $n \in \mathbb{T}_m$, *States:* $(x,y,p) \in \bar{\mathcal{S}}_{loc}$, *Actions:* $\alpha_n(x,y,p) \in \mathcal{C}_{M,R}$ *from* (13), $n \in \mathbb{T}_m$, *Immediate Rewards:* $c_n(x,y,p,\alpha_n(x,y,p)) \in \mathbb{R}_+$, *with Assumption 2.2 Terminal Reward:* $C_T(x,y,p) = C_T(y,p) \in \mathbb{R}_+$, *with Assumption 2.3 Transition Probabilities:* $\mathbb{P}_n(i,j,a)$ *defined as an appropriate discretization of* (1)

3.1. 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.

$$v_n(x,y,p) = \max_{\alpha_n \in \mathcal{C}_{M,R}(x,y,p)} Q_n(x,y,p,\alpha_n), \quad (14)$$

$$\alpha_n^*(x,y,p) = \arg \max_{\alpha_n \in \mathcal{C}_{M,R}(x,y,p)} Q_n(x,y,p,\alpha_n), \quad (15)$$

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

$$Q_n(x,y,p,\alpha_n) = \{c(p,\alpha_n) + \sum_{j \in \mathcal{Z}_l} [\mathbb{P}_n(j,(x,y,p),\alpha_n) \times v_{n+1}(x + p\alpha_n, y - \alpha_n, p_j)]\}, \quad (16)$$

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

The following structural result has been adapted from [8] and [9]. From [2] and [7], we know that the optimal control adheres to a Hamilton-Jacobi-Bellman quasivariational inequality, and a selling opportunity is at hand when

$$\max\{\tfrac{\partial v}{\partial t} + \mathcal{L}v, \mathcal{M}v - v\} = \mathcal{M}v - v = 0. \quad (17)$$

Where $\mathcal{L}$ and $\mathcal{M}$ are the infinitesimal generator function for price process (1) and the impulse control intervention operator for the optimal impulse control Problem 2.1. See pg. 151 and pg. 159 of [5] for more discussion of operators $\mathcal{L}$ and $\mathcal{M}$, respectively. 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 (17) holds. Then

$$q_n \in [0,1], n \in \mathbb{T}_m. \quad (18)$$

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, for $n \in \mathbb{T}_m$,

$$v_n(x,y,p) = \max_{0 \le \alpha \le y} \{c(\alpha,p) + \bar{v}_{n+1}(x, y-\alpha, p - m(\alpha))\},$$
$$v_T(x,y,p) = C_T(y,p), \quad v_{T+1}(x,y,p) = 0, \quad (19)$$

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

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

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

Before continuing, apply the following assumption to the market impact term defined in (1)

Assumption 3.1. *(Probability of Incurring Market Impact) Assume that after a sell action a is made, the market impact term $m(a)$ does not occur w.p.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*

$$f_m(a,p) = \begin{cases} 0 & \text{if } a = 0 \\ p_m(p) & \text{if } a > 0, \end{cases} \quad (20)$$

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 3.1 below gives the structure of the optimal policy.

Theorem 3.1. *(Optimal Policy Structure)*
Assume the MDP (Definition 3.1) with market impact function $m(\cdot)$ defined by Assumption 2.1 and with immediate and terminal reward functions defined by Assumptions 2.2 and 2.3. If the MDP has a value function defined by (14) *that is a nondecreasing and concave function of y, having investment opportunity parameter q_n defined in* (18)*, and market impact probability function given by Assumption 3.1, then the optimal policy* (15) *of MDP (Definition 3.1) 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. □

4. NUMERICAL RESULTS

Table 1 below provides a list of parameter values used in this simulation. The transition probabilities for various actions (ranging from the "no trade" action to the "maximum trade" action) are represented in Fig. 1a and Fig. 1b. Fig. 1c and Fig. 1d show the form of the value function over varying price and time taking snapshots in shares owned. As expected, 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. 1e and Fig. 1f, 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 can be exploited in order to use the SPSA algorithm.

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
κ	price impact constant	0.1

Table 1. Simulated Liquidation Problem Parameters

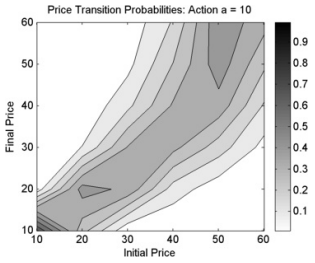

Fig. 1a, Sell 10 shares

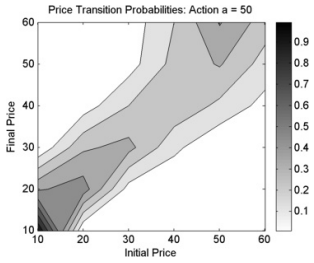

Fig. 1b, Sell 50 shares

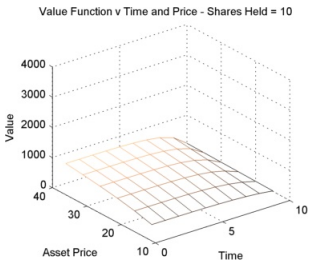

Fig. 1c, Value function

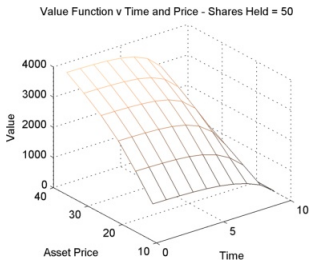

Fig. 1d, Value function

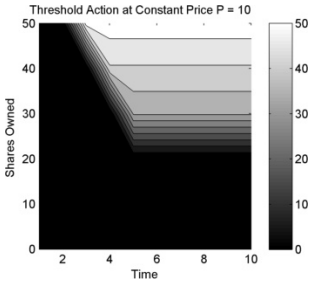

Fig. 1e, Optimal Strategy

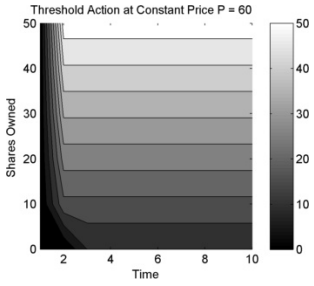

Fig. 1f, Optimal Strategy

5. CONCLUSION

An optimal portfolio liquidation problem has been defined in a denumerable MDP context. The elements of the MDP have been defined in such a way that there exists a monotone optimal policy. As such, this work can be used to apply reinforcement learning techniques (e.g. the Simultaneous Perturbation Stochastic Approximation algorithm) to the otherwise complex impulse control problem defined in Sec. 2.

Appendix

Proof of Theorem 3.1

Proof. See [8]: sequential allocation problem for the proof of (i) and (ii). For (iii.), notice that 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 of the state-action reward function given in (16) in variables (a, p).

Definition A.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 $x_1 > x_2, y_1 > y_2$. Similarly, 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 nondecreasing structure of the optimal liquidation policy $\alpha_n^*(\cdot, \cdot, \cdot)$ given in (15) follows. Step 1 is shown in the following lemma.

Lemma A.1. *The optimal cost to go function $v_n(x, y, p)$ defined by* (14) *is increasing in the price of the asset p.*

See proof in [9] □

The next theorem outlines Step 2.

Theorem A.1. *If the terminal reward function $C(\cdot, \cdot)$ is an increasing function in p and is integer convex then the state-action reward function $Q_n(x, y, p, a)$ is supermodular in (a, p), i.e.*

$$Q_n(x, y, p, a) - Q_n(x, y, p, 0) \leq Q_n(x, y, p+1, a) - Q_n(x, y, p+1, 0), \quad \forall a \leq 0. \tag{21}$$

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

This proof is a simple adaptation of the analogous proof in [9]. □

A. REFERENCES

[1] 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.

[2] F. Guilbaud, M. Mnif, and H. Pham, "Numerical methods for an optimal order execution problem," *ArXiv e-prints*, June 2010.

[3] 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.

[4] 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.

[5] 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.

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

[7] M. Gaigi, V. Vath, M. Mnif, and S. Toumi, "Numerical approximation for a portfolio optimization problem under liquidity risk and costs," 2013.

[8] Sheldon Ross, *Introduction to Stochastic Dynamic Programming*, Academic Press, 1991.

[9] 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.