Optimal Switching via Randomized Formulation: A Numerical Approach

Overseas Research Fellowship (ORF) 2025 for Science Student
Poster Number: B05
Name: Wong Hoi Yat
University Number: 303608678
Student Major: Statistics
External Supervisor: Prof. Haoyang Cao (Johns Hopkins University)
Internal Supervisor : Prof. Marius Hofert (The University of Hong Kong)

Introduction

Motivation

- Singular control models problems involving cumulative, instantaneous actions, such as making irreversible investments or managing inventory [1, 4].
- Classically solved by PDE, system of Quasi-Variational Inequalities (QVIs) imposes significant numerical challenges.

Objective

Develop a tractable numerical framework by introducing a randomization technique

Contribution

- New HJB is highly amenable to modern deep learning solvers, turning a discrete "all-or-nothing" decision into a continuous control problem.
- While randomization has been explored for optimal stopping problems [2, 3, 6, 7, 8], its application to optimal switching is novel.
  - Increased Complexity: It involves two (or more) coupled value functions that depend on each other.
  - Economic Significance: We solve a much broader class of problems, such as the inventory management problem with both buying and selling decisions.

Model Framework

Singular Control Problem

Find policy that maximizes a value function $V(p, y)$, with cumulative purchases ξ_t^+ and sales ξ_t^-. Spot price P_t follows a Geometric Brownian Motion: $dP_t = \mu P_t dt + \sqrt{2}\sigma dW_t$

Decomposition

For each fixed inventory level z, the problem reduces to a 1D optimal switching problem [4, 5]. The agent decides between two regimes:

- Regime 0 (Buy): value function is $v_0(p; z)$
- Regime 1 (Sell): value function is $v_1(p; z)$

With b being the maximum inventory level, the original value function is reconstructed by:

$$V(p, y) = \int_0^y v_1(p, z)\, dz + \int_y^b v_0(p, z)\, dz$$

Classical Solution: The value functions v_0 and v_1 are found by solving a complex coupled QVIs [4].

References

[1] Constantinides, G. M., & Richard, S. F. (1978). Existence of Optimal Simple Policies for Discounted-Cost Inventory and Cash Management in Continuous Time. *Operations Research*, 26(4), 620-636.
[2] Dianetti, J., Ferrari, G., & Xu, R. (2024). Reinforcement Learning for Exploratory Optimal Stopping: A Singular Control Formulation. *arXiv preprint arXiv:2408.09335*.
[3] Dong, Y. (2024). Randomized Optimal Stopping Problem in Continuous Time and Reinforcement Learning Algorithm. *SIAM Journal on Control and Optimization*, 62(3), 1590-1614.
[4] Guo, X., Kaminsky, P., Tomecek, P., & Yuen, M. (2011). Optimal spot market inventory strategies in the presence of cost and price risk. *Mathematical Methods of Operations Research*, 73, 109-137.
[5] Guo, X., & Tomecek, P. (2008). Connections between Singular Control and Optimal Switching. *SIAM Journal on Control and Optimization*, 47(1), 421-443.
[6] Reppen, A. M., Soner, H. M., & Tissot-Daguette, V. (2022). Neural Optimal Stopping Boundary. *arXiv preprint arXiv:2205.04595*.
[7] Touzi, N., & Vieille, N. (2002). Continuous-Time Dynkin Games with Mixed Strategies. *SIAM Journal on Control and Optimization*, 41(4), 1073-1088.
[8] Wang, H., Zariphopoulou, T., & Zhou, X. Y. (2020). Reinforcement Learning in Continuous Time and Space: A Stochastic Control Approach. *Journal of Machine Learning Research*, 21(198), 1-34.

Randomized Problem Formulation

New Control & State Process

The agent now chooses continuous switching intensities, π_t^+ and π_t^- . **To satisfy the Markovian property for Dynamic Programming Principle (DPP), we need new state variables to summarize the entire history of the control intensities into the current state.** Thus, we add survival probabilities (haven't switched out of a given regime) y_t^+ and y_t^- , with dynamics: $dy_t^+ = -\pi_t^+ dt, dy_t^- = -\pi_t^- dt$.
We also include the entropy-based regularizer $R(\pi) \coloneqq \pi - \pi\log(\pi)$ to ensure the resulting HJB system is smooth and well-behaved [3].

Profit Functions

Regime 0: $J^+(p, y^+; y^-, z) =$
$\mathbb{E}[\int_0^\infty e^{-\rho t}(G^+(P_t; z)\pi_t^+ + \lambda^+ R(\pi^+))y_t^+ dt]$
Regime 1: $J^-(p, y^-; y^+, z) =$
$\mathbb{E}[\int_0^\infty e^{-\rho t}(\tilde{h}(z)P_t + G^-(P_t; z)\pi_t^- + \lambda^- R(\pi^-)y_t^- dt]$
Where the switching payoffs are defined as
$G^+(P_t; z) = -K^1 + V^-(p, y^-; y^+, z)$
$G^-(P_t; z) = -K^0 + V^+(p, y^+; y^-, z)$

Simplification and Optimal Policy

We make the ansatz that the value function for each regime is linear in its respective survival probability.

$$V^+(p, y^+; y^-, z) = y^+ v^+(p; z)$$
$$V^-(p, y^-; y^+, z) = y^- v^-(p; z)$$

Optimal switching intensities:

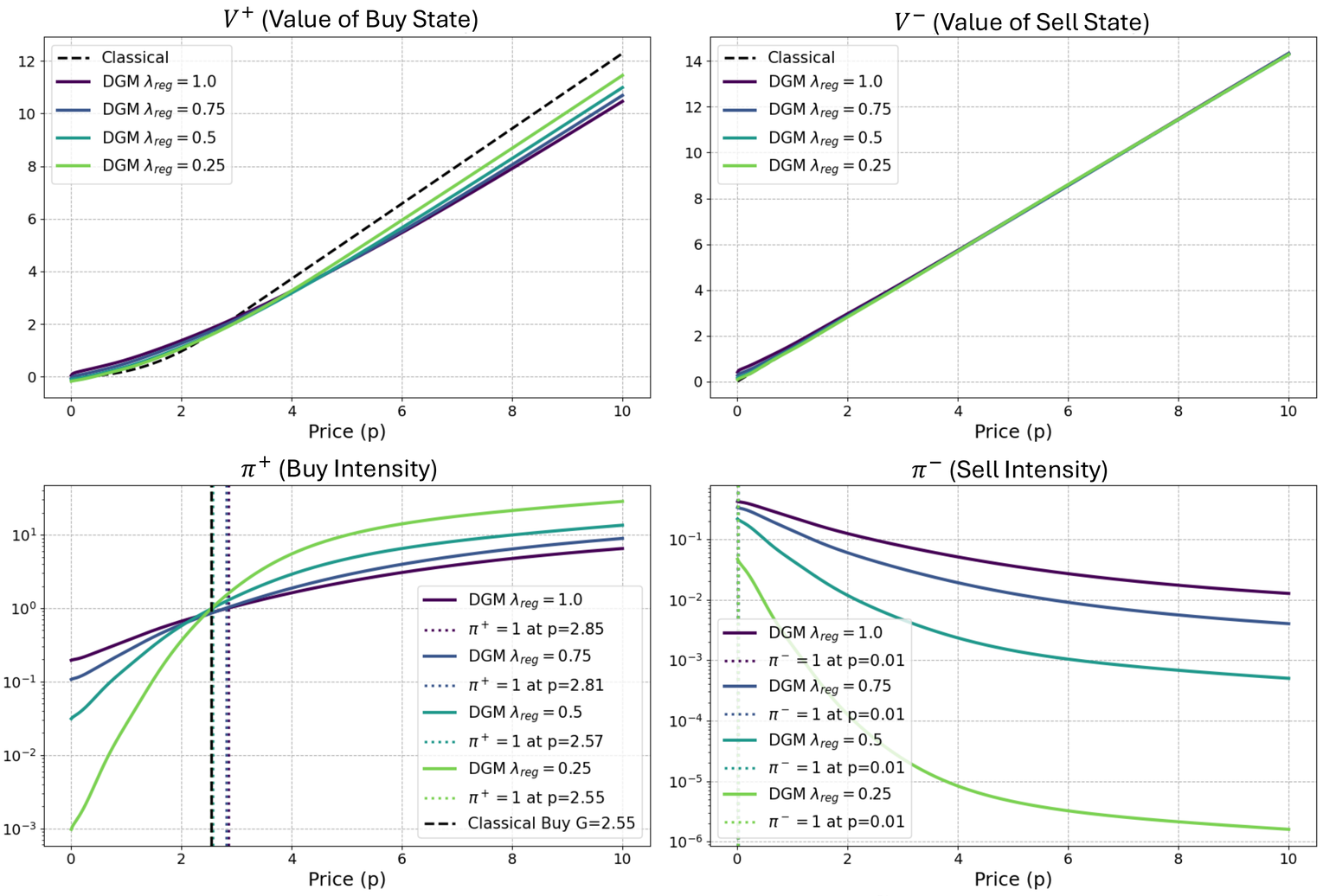
$$\bar{\pi}^+ = \exp(-\frac{v^+ - G^+}{\lambda^+})$$
$$\bar{\pi}^- = \exp(-\frac{v^- - G^-}{\lambda^-})$$

Final coupled non-linear ODEs:

$$(\mathcal{L}_p - \rho)v^+ + \lambda^+ \exp(-\frac{v^+ - G^+}{\lambda^+}) = 0$$
$$(\mathcal{L}_p - \rho)v^- + \tilde{h}(z)p + \lambda^- \exp(-\frac{v^- - G^-}{\lambda^-}) = 0$$

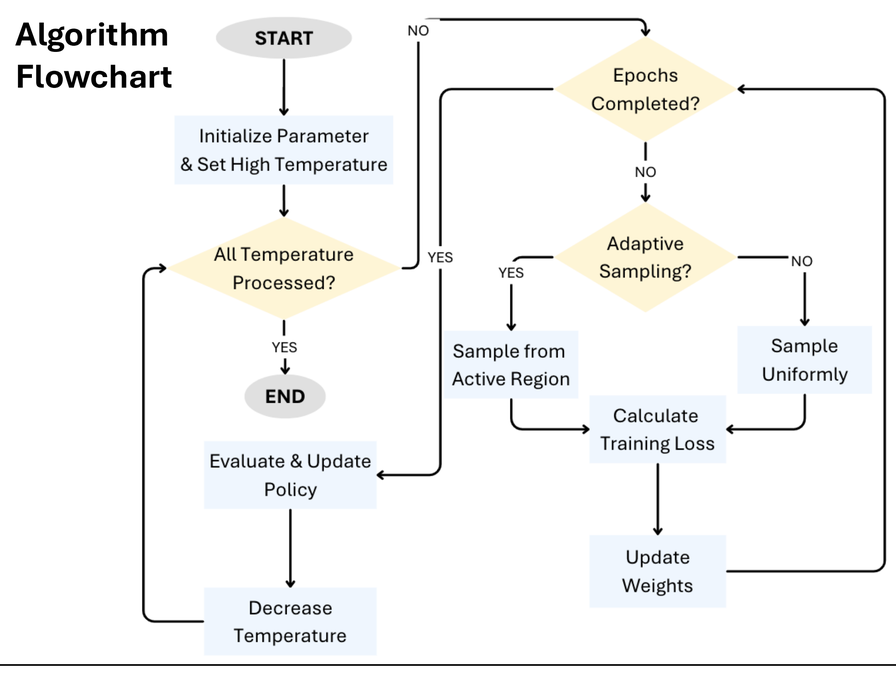
where $\mathcal{L}_p V(p; \dots) \coloneqq \mu p \delta_p V + \sigma^2 p^2 \delta_{pp} V$. K^0 and K^1 are switching cost from regime 0 and 1.

Numerical Experiment



Value Function and Policy Convergence for $\tilde{h}(z)$=1.0

Algorithm Flowchart



Numerical Method

Vanishing temperature approach solves a sequence of HJB systems with decreasing λ to test for convergence to the classical solution.

Performance

The method proved effective and robust across all three economic scenarios.

- Training loss confirms HJB solution found.
- All value functions converged to classical solutions as λ→0.
- All policy boundaries converged to classical boundary as λ→0.

Conclusion

- **We develop an entropy-based randomization to establish a bridge between the analytically difficult QVI formulation and modern deep learning solvers.**
- **The numerical experiments confirm this is a robust method that approximates the optimal policy across various economic conditions.**
- Next Steps: 1. Provide a formal mathematical proof for the observed convergence. 2. Test on higher-dimensional problems and different economic scenarios.