COS435 / ECE433: Introduction to RL Due: March 24, 2024

Project Proposal: Optimal Option Hedging and Pricing

Authors. Arif Ansari (aa4433@princeton.edu), Christos Avgerinos Tegopoulos (ct3125@princeton.edu), Jeremy Jun-Ping Bird (jb9895@princeton.edu),

Project type: 2. Applying RL to a new problem

Introduction: We aim to apply RL to the well-established financial problem of option pricing. In perfect, frictionless markets, options are priced through a replicating portfolio of a risky asset and a risk-free bond. This renders the problems of optimal hedging and pricing synonymous.

The Black-Scholes-Merton's (BSM) model analytically solves the problem of optimal option hedging and pricing for European Options. Yet options trading remains a multi-billion dollar business where traders employ discretion is pricing and risk management. Clealy, this highlights the BSM model's shortcomings.

Pitfalls of BSM:

- **Non-Continuous Rehedging:** The assumption of continuous rebalancing is costly and impractical. Thus, the replicating portfolio will carry a certain degree of risk.
- **Transaction Costs:** Traders face transaction costs (e.g. spread, market impact) each time they rebalance. However, these costs are entirely neglected within the classical BSM model.
- **Stochastic Volatility:** Real markets have time-varying volatility, with strong autocorrelation. The BSM model however assumes log-normal price dynamics.

RL Approach: The optimal hedging problem requires the agent to balance the tradeoff between transaction costs and perfectly hedging, which is an even tougher problem to solve for analytically under stochastic volatility assumptions. This suggests applying RL techniques to learn the optimal policy. Formally we have:

- **Actions** a_t**:** The amount of the underlying hedged at each timestep t[2][7]
- **States** S_t**:** The price of the underlying (or some time transformed variable X_t which normalizes for drift in the underlying dynamics)[2][7]
- **Dynamics:** The underlying follows geometric Brownian motion with stochastic volatility (SABR model)[1]
- **Reward Function** R_t**:** Here our reward at each period $0 < t < T$ is equal to our change in wealth $\Delta w_t = S_t(a_{t-1} - a_t) - \kappa|S_t(a_{t-1} - a_t)|$ minus the variance of Δw_t scaled by some risk-aversion parameter λ[3]
$$R_t = \Delta w_t - \lambda \mathbb{V}[\Delta w_t] = \Delta w_t - \lambda(\Delta w_t - \mathbb{E}[\Delta w_t])^2$$
Note that at $t = 0$ the initial change in wealth $\Delta w_0 = -S_0 a_0 - \kappa|S_0 a_0|$ as we must buy the initial replicating portfolio and at matrity we have $\Delta w_T = S_T a_T - \kappa|S_T a_T| - G(S_T)$ where $G(S_T)$ is the option payoff.

This gives final Bellman Equation
$$V^\pi(S_t) = \mathbb{E}_t^\pi[R_t(S_t, a_t, S_{t+1}) + \gamma V^\pi(S_{t+1})]$$

Why Not Bandit?: Due to our transaction costs, how much we choose to hedge at period t influences the transaction costs we pay when rebalancing at $t + 1$. For example, if we accumulate a large position of stock in a single period, at some point in the future, we will have to unwind this position (given our terminality condition), resulting in larger transaction costs. Hence, treating this as a RL problem encourages small, incremental hedges and no wild swings in position size.

What We Hope to Achieve: Our aim is for our RL agent to achieve the following:

- **BSM Outperformance:** As established above, we expect delta hedging according to the BSM model to be suboptimal in a more realistic market setting. Subsequently, we hope for our RL-learned hedging strategy to outperform BSM hedging (i.e. achieve superior risk adjusted PnL).
- **Model Independence:** Our RL approach does not require an explicit knowledge of market dynamics, the "greeks" or a known model for transaction costs as it is model-free and aims to learn optimal hedging based solely on observed rewards and state transitions. This is a more appropriate reflection of reality where it is impossible to perfectly model these factors.

Final Methodology:

1. Simulate N monte-carlo paths for the underlying according to our predefined model dynamics
2. Using the RL problem setup described above, apply:
   - **Tabular Q-Learning** *(State-spaced discretized using lattice approximation of stock prices. Action-space discretized into 0.01 increments of quantity held.)*
   - **DDPG** *From the Lillicrap, 2019 paper*[4]
   - **PPO** *From the Schulman, 2017 paper* [5]
   - **GRPO:** *From DeepSeek 2024 paper* [6]

   Compare each algorithm's performance, data efficiency and rate of convergence.
3. Repeat the above under stochastic volatility dynamics. Investigate the potency of the RL policy in the new environment.

We're specifically interested in looking into **GRPO** as it doesn't need to learn the critic function which can be difficult due to stochasticity of transaction costs. By using only observed Monte Carlo methods it may be better at dealing with the constraints of the problem. Although it may be potentially unstable at points with high variance we are hoping to see if running a numerous amount of trials will help normalize these results.

References

[1] Jay Cao, Jacky Chen, John Hull, and Zissis Poulos. Deep hedging of derivatives using reinforcement learning. *The Journal of Financial Data Science*, 3(1):10–27, December 2020.

[2] Igor Halperin. Qlbs: Q-learner in the black-scholes(-merton) worlds, 2019.

[3] Petter N. Kolm and Gordon Ritter. Dynamic replication and hedging: A reinforcement learning approach. *The Journal of Financial Data Science*, 2019.

[4] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2019.

[5] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017.

[6] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024.

[7] Zoran Stoiljkovic. Applying reinforcement learning to option pricing and hedging, 2023.