Optimizing High-Frequency Trading Strategies with Markov Decision Processes

Omniscience Research

Abstract

High-frequency trading (HFT) has become a dominant force in modern financial markets, characterized by rapid execution and large volumes of trades. The application of Markov Decision Processes (MDPs) in HFT offers a robust framework for optimizing trading strategies, managing risk, and improving decision-making under uncertainty. This paper investigates the use of MDPs to enhance the efficiency and profitability of HFT operations. We provide a comprehensive review of current applications, including strategy development, order execution, and risk management. Additionally, we address the challenges of algorithmic complexity and the need for adaptive learning in dynamic market conditions. By examining empirical results and discussing potential future developments, this paper aims to offer valuable insights into the practical applications of MDPs in the fast-paced world of high-frequency trading.

1 Background on Markov Decision Processes

Markov Decision Processes (MDPs) provide a robust mathematical framework for modeling sequential decision-making problems where outcomes are partly random and partly under the control of a decision-maker. MDPs are particularly well-suited for applications in high-frequency trading (HFT), where a sequence of rapid decisions must be made under uncertainty.

1.1 Definition and Components of MDPs

An MDP is defined by a tuple (S, A, P, R, γ), where S is a finite set of states, A is a finite set of actions, P is the state transition probability matrix, R is the reward function, and γ is the discount factor [Sutton and Barto, 2018]. At each time step, the decision-maker observes the current state $s \in S$, selects an action $a \in A$, and receives a reward $R(s, a)$. The system then transitions to a new state s' with probability $P(s'|s, a)$.

1.2 Relevance to Decision-Making in Trading

In the context of HFT, an MDP can be used to model the evolution of market states and the impact of trading actions on an agent's portfolio. The state space S can include factors such as current inventory levels, recent price movements, and market volatility. Actions A might consist of placing buy or sell orders of various sizes, or choosing to wait. The reward function R is typically related to the profit or loss resulting from the actions taken, while the transition probabilities P capture the stochastic nature of price movements and market responses to trade executions.

MDPs are particularly useful in HFT due to their ability to capture the temporal dependencies of actions and states. This is crucial in a domain where the profitability of a trading strategy can be highly dependent on the sequence and timing of trades. Furthermore, the discount factor γ allows traders to balance immediate rewards against future gains, a key consideration in the fast-paced environment of HFT where short-term fluctuations can have significant long-term implications [Nevmyvaka et al., 2006].

Preprint. Under review.

The application of MDPs in HFT is not without challenges. The high dimensionality of the state space, the need for real-time decision-making, and the non-stationarity of financial markets all pose significant hurdles. However, advancements in computational methods and machine learning have made it increasingly feasible to apply MDPs to complex trading problems [Gao and Chan, 2018].

By framing HFT as an MDP, traders can systematically evaluate the potential outcomes of their actions and develop strategies that are both profitable and robust to market uncertainties. This approach also facilitates the use of reinforcement learning algorithms, which can learn optimal policies through simulation and interaction with the market Li et al. [2019].

In summary, MDPs offer a powerful tool for navigating the complexities of high-frequency trading. By enabling a structured approach to decision-making under uncertainty, MDPs help traders to develop sophisticated strategies that can adapt to the dynamic nature of financial markets. As computational techniques continue to advance, the potential for MDPs to revolutionize HFT grows ever more promising.

2 MDPs in Strategy Development

The development of trading strategies that can consistently yield profit in the high-frequency trading (HFT) environment is a complex task. Markov Decision Processes (MDPs) have emerged as a powerful tool for designing and optimizing these strategies. This section dives into the formulation of trading strategies as MDPs and provides examples of how MDP-based strategies are applied in HFT.

2.1 Formulating Trading Strategies as MDPs

The formulation of trading strategies as MDPs involves defining the state space, action space, reward function, and transition probabilities in a manner that encapsulates the trader's objectives and the market dynamics.

2.1.1 State Space

The state space in an MDP-based trading strategy includes all relevant market information that could influence decision-making. This may encompass price levels, historical volatility, order book dynamics, and even macroeconomic indicators. For instance, a state could be represented as a vector $s_t = (p_t, v_t, i_t, \ldots)$, where p_t is the current price, v_t is the volume, and i_t is the inventory level at time t [Gould et al., 2013].

2.1.2 Action Space

The action space defines the possible actions a trader can take at any given state. In HFT, this typically includes placing limit or market orders, modifying existing orders, or canceling orders. Each action a_t taken in state s_t leads to a transition to a new state s_{t+1}, influenced by the market's reaction to the trade [Cartea et al., 2015].

2.1.3 Reward Function

The reward function is designed to reflect the trader's goals, such as maximizing profit or minimizing market impact. A common choice for the reward function is the change in portfolio value, which can be expressed as $R(s_t, a_t) = \Delta PV_t$, where ΔPV_t is the change in portfolio value resulting from action a_t in state s_t [Nevmyvaka et al., 2006].

2.1.4 Transition Probabilities

Transition probabilities $P(s_{t+1}|s_t, a_t)$ represent the likelihood of moving from the current state to the next state given an action. In HFT, these probabilities are often estimated using historical data or through simulation techniques, acknowledging the inherent randomness in market movements [Gao and Chan, 2018].

2.2 Examples of MDP-Based Strategies in HFT

Several practical applications of MDPs in HFT illustrate their effectiveness in developing trading strategies.

2.2.1 Optimal Order Placement

One application of MDPs in HFT is the optimization of order placement strategies. By modeling the limit order book as an MDP, traders can determine the optimal price levels and sizes for their orders to balance the trade-off between execution probability and price improvement [Gould et al., 2013].

2.2.2 Inventory Management

MDPs are also used for inventory management in HFT, where the goal is to minimize the risk associated with holding financial positions over time. An MDP can help in determining the optimal times to unwind positions while considering the market liquidity and the impact of trades on prices [Cartea et al., 2015].

2.2.3 Market Making

Market making, which involves providing liquidity by simultaneously placing buy and sell limit orders, can be modeled as an MDP to optimize the bid-ask spread and order sizes. This approach allows market makers to maximize their expected utility while managing inventory risk [Avellaneda and Stoikov, 2008].

The use of MDPs in strategy development for HFT represents a significant advancement in the field of quantitative finance. By providing a structured approach to decision-making under uncertainty, MDPs enable traders to devise strategies that are both sophisticated and adaptable to market conditions. As the financial markets continue to evolve, the flexibility and robustness of MDP-based strategies will remain invaluable for traders seeking to maintain a competitive edge.

3 Order Execution and MDPs

Optimizing order execution is a critical component of high-frequency trading (HFT), where the goal is to execute large orders while minimizing market impact and slippage. Markov Decision Processes (MDPs) provide a structured approach to model and solve the sequential decision-making problems inherent in order execution. This section discusses the optimization of order execution using MDPs and presents case studies of execution algorithms in HFT.

3.1 Optimizing Order Execution with MDPs

The execution of orders in HFT can be modeled as an MDP, where the trader must decide the size and timing of trades to minimize execution costs under uncertain market conditions.

3.1.1 Defining Execution Cost

Execution cost is typically defined as the difference between the realized price of a trade and some benchmark price, often the price at the time the order was placed or the volume-weighted average price (VWAP). The objective is to minimize this cost over the course of executing the order Almgren and Chriss [2001].

3.1.2 MDP Formulation for Order Execution

In the MDP framework, the state space includes information about the current market conditions, the remaining order size, and the elapsed time. The action space consists of the possible trade sizes at each decision point. The reward function is designed to penalize negative deviations from the benchmark price, effectively minimizing the execution cost. Transition probabilities capture the dynamics of the order book and the impact of trades on market prices [Bertsimas and Lo, 1998].

3.1.3 Dynamic Programming for Optimal Execution

Dynamic programming techniques, such as value iteration or policy iteration, can be employed to solve the MDP and find an optimal trading policy. This policy dictates the trade size that minimizes the expected execution cost at each decision point, given the current state of the market and the remaining order size [Forsyth and Vetzal, 2012].

3.2 Case Studies of Execution Algorithms

Several case studies highlight the practical application of MDPs in optimizing order execution in HFT.

3.2.1 Volume-Synchronized Probability of Informed Trading (VPIN)

The VPIN metric is used to estimate the probability of informed trading occurring in the market. By incorporating VPIN into an MDP framework, traders can adjust their execution strategy in real-time to account for changes in the informational content of trades, thereby managing adverse selection risk [Easley et al., 2012].

3.2.2 Adaptive Execution with Reinforcement Learning

Reinforcement learning, a subset of machine learning that can be applied to solve MDPs, has been used to create adaptive execution algorithms. These algorithms learn from past execution performance and adapt their trading strategies to changing market conditions, resulting in improved execution outcomes over time [Nevmyvaka et al., 2006].

The application of MDPs to order execution in HFT showcases the versatility of this framework in addressing complex, dynamic problems. By capturing the nuances of market dynamics and trader objectives, MDPs enable the development of sophisticated execution algorithms that can navigate the intricacies of the financial markets. As markets continue to evolve, the adaptability of MDP-based execution strategies will play a pivotal role in maintaining the efficacy and competitiveness of high-frequency trading operations.

4 Risk Management through MDPs

In high-frequency trading (HFT), risk management is paramount due to the high speed and volume of trades. Markov Decision Processes (MDPs) offer a structured approach to dynamically manage risk by optimizing decisions in the face of uncertainty. This section dives into the application of MDPs for risk assessment and control in HFT.

4.1 Risk Assessment in HFT

Risk assessment in HFT involves quantifying the potential for loss due to market volatility, order execution costs, and adverse selection. MDPs facilitate a probabilistic approach to risk assessment, allowing traders to make informed decisions based on the likelihood of various market scenarios.

4.1.1 Quantifying Market Risk

Market risk in HFT can be quantified using Value at Risk (VaR) or Conditional Value at Risk (CVaR). MDPs can incorporate these risk measures into the decision-making process by adjusting the reward function to penalize high-risk actions, thereby aligning trading strategies with risk tolerance levels [Rockafellar and Uryasev, 2000].

4.1.2 Incorporating Execution Risk

Execution risk, arising from the possibility of suboptimal order execution, can be modeled within the MDP framework. By considering the cost of potential delays and market impact, MDPs can optimize the trade-off between execution speed and risk [Moallemi and Sağlam, 2013].

4.2 MDPs for Dynamic Risk Control

Dynamic risk control in HFT is concerned with adjusting trading strategies in real-time to manage exposure to risk. MDPs provide a mechanism to update policies based on the evolving state of the market and the trader's risk profile.

4.2.1 Stochastic Control for Real-Time Decision Making

Stochastic control techniques, applied within the MDP framework, enable traders to make real-time decisions that balance expected returns against risk exposure. By continuously updating the policy based on observed market conditions, traders can dynamically hedge their positions and limit potential losses [Bäuerle and Rieder, 2011].

4.2.2 Leveraging MDPs for Portfolio Optimization

MDPs can also be used for portfolio optimization in HFT, where the goal is to maximize returns while managing the risk of a portfolio of assets. By modeling the joint distribution of asset returns and incorporating transaction costs, MDPs can determine optimal trading actions to adjust portfolio holdings in response to market movements [Boyd et al., 2017].

The integration of MDPs into risk management practices in HFT represents a significant advancement in the ability to systematically and dynamically control risk. By embedding risk considerations directly into the decision-making process, MDPs enable the development of trading strategies that are both responsive to market conditions and aligned with risk preferences. As the financial markets continue to evolve, the role of MDPs in risk management is likely to expand, offering traders sophisticated tools to navigate the complex landscape of high-frequency trading.

5 Adaptive Learning in MDPs

Adaptive learning is a critical component in high-frequency trading (HFT) where market conditions can change rapidly and unpredictably. Markov Decision Processes (MDPs) provide a framework for developing adaptive trading strategies that can learn and evolve over time. This section examines the role of reinforcement learning in MDPs and how it facilitates the creation of adaptive strategies to cope with changing market dynamics.

5.1 Reinforcement Learning and MDPs

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment. In the context of MDPs, RL algorithms are used to find optimal policies that maximize cumulative rewards over time. The agent learns from the consequences of its actions without explicit instruction, adjusting its strategy to improve performance [Sutton and Barto, 2018].

5.1.1 Q-Learning and Deep Q-Networks

Q-learning is a widely used RL algorithm in which an agent learns a value function that estimates the expected rewards of taking certain actions in given states. In HFT, Q-learning can be applied to MDPs to determine the optimal timing and size of trades based on historical and real-time market data [Watkins and Dayan, 1992]. Deep Q-Networks (DQNs) extend Q-learning by using deep neural networks to approximate the value function, enabling the handling of high-dimensional state spaces typical in HFT [Mnih et al., 2015].

5.1.2 Policy Gradient Methods

Policy gradient methods, another class of RL algorithms, directly optimize the policy function instead of the value function. These methods are particularly useful in HFT when the action space is continuous or when the policy needs to be highly expressive. Algorithms like Proximal Policy Optimization (PPO) have been shown to be effective in such settings [Schulman et al., 2017].

5.2 Adaptive Strategies in Changing Market Conditions

The ability to adapt to new market conditions is a significant advantage in HFT. MDPs equipped with RL can adjust to shifts in market dynamics, such as changes in volatility or liquidity, by continuously learning from market feedback.

5.2.1 Online Learning and Exploration

Online learning in MDPs allows for the real-time updating of policies as new data becomes available. This is crucial in HFT, where the market state can change between the placement and execution of an order. Exploration techniques, such as ϵ-greedy or entropy-based methods, ensure that the trading strategy does not become overly deterministic and can adapt to previously unseen market conditions [Tokic, 2010].

5.2.2 Handling Non-Stationarity

Financial markets are inherently non-stationary, meaning that the probability distributions governing market behavior change over time. MDPs in HFT must account for this non-stationarity. Techniques such as sliding window approaches or meta-learning can help RL algorithms adapt to these changes by focusing on recent data or by learning to learn from changing environments [Finn et al., 2017].

The integration of adaptive learning mechanisms into MDPs represents a significant advancement in the development of robust and flexible trading strategies for HFT. By leveraging the power of RL, MDPs can evolve in step with the markets, continually refining their decision-making processes to optimize performance. As the financial landscape becomes increasingly complex and data-driven, the ability of MDPs to adapt and learn will be paramount in maintaining a competitive edge in the high-speed world of high-frequency trading.

6 Evaluation of MDP-Based Trading Systems

The evaluation of trading systems based on Markov Decision Processes (MDPs) is crucial for determining their effectiveness and viability in high-frequency trading (HFT) environments. This section discusses the performance metrics used to assess MDP-based trading systems and presents empirical results and comparative analyses to illustrate their practical implications.

6.1 Performance Metrics for HFT

In evaluating MDP-based trading systems, several performance metrics are commonly used to measure their success. These metrics capture various aspects of trading performance, including profitability, risk, and execution quality.

6.1.1 Profit and Loss (P&L)

The most direct measure of a trading system's performance is its ability to generate profit. Profit and Loss (P&L) calculations take into account the net revenue from trades after accounting for transaction costs, slippage, and fees [Aldridge, 2013]. P&L can be expressed in absolute terms or as a percentage return on investment (ROI).

6.1.2 Sharpe Ratio

The Sharpe ratio is a risk-adjusted measure of return that evaluates the performance of an investment compared to a risk-free asset, after adjusting for its risk [Sharpe, 1994]. It is calculated as the difference between the returns of the investment and the risk-free rate, divided by the standard deviation of the investment returns. A higher Sharpe ratio indicates a more desirable risk-return profile.

6.1.3 Maximum Drawdown

Maximum drawdown measures the largest single drop from peak to trough in the value of a portfolio, before a new peak is achieved. It is an indicator of downside risk over a specified time period [Magdon-Ismail et al., 2004].

6.2 Empirical Results and Comparative Analysis

Empirical studies of MDP-based trading systems have demonstrated their potential to outperform traditional trading strategies, particularly in the realm of HFT where decision-making speed and precision are paramount.

6.2.1 Case Study: MDPs in Market Making

A study by Avellaneda and Stoikov (2008) on optimal market making strategies modeled the problem as an MDP, showing that the MDP framework could be used to derive optimal bid and ask quotes in a limit order book [Avellaneda and Stoikov, 2008]. Their results indicated that MDP-based strategies could adapt to market conditions and control inventory risk more effectively than static strategies.

6.2.2 Comparison with Non-MDP Strategies

When compared to heuristic or rule-based trading strategies, MDP-based systems often exhibit superior performance due to their ability to optimize decisions based on a comprehensive set of state variables and to adapt to new information [Nevmyvaka et al., 2006]. For instance, MDPs that incorporate machine learning techniques can dynamically adjust to changing market conditions, leading to more consistent returns over time.

The evaluation of MDP-based trading systems in HFT reveals a complex interplay between algorithmic sophistication, computational demands, and market conditions. While empirical results have shown promising outcomes, the true test of these systems lies in their long-term adaptability and resilience to market shocks. As the financial markets continue to evolve, the robustness of MDP-based strategies will be a testament to their capacity to not only interpret the intricate tapestry of market signals but also to weave new patterns of success in the ever-changing landscape of high-frequency trading.

7 Limitations and Challenges

While Markov Decision Processes (MDPs) have proven to be powerful tools in the domain of high-frequency trading (HFT), they are not without limitations and challenges. This section dives into the constraints of MDPs when applied to HFT and identifies the current challenges and research gaps that need to be addressed to enhance their applicability and performance.

7.1 Modeling Limitations

MDPs assume that the decision-making process can be modeled as a stochastic process with discrete states and actions. However, the financial markets are complex, dynamic, and continuously evolving, which can make the discretization of states and actions challenging.

7.1.1 State Space Complexity

The state space in HFT can be vast due to the high dimensionality of market data and trader information. As the number of state variables increases, the state space grows exponentially, leading to the "curse of dimensionality" Bellman [1957]. This complexity can make it computationally infeasible to find optimal policies, especially in real-time trading scenarios.

7.1.2 Non-Stationarity of Financial Markets

Financial markets are inherently non-stationary; the underlying probability distributions change over time, often in unpredictable ways Cont and Tankov [2001]. This non-stationarity violates

the Markov property, which assumes that future states depend only on the current state and not on the sequence of events that preceded it. Adapting MDPs to non-stationary environments remains a significant challenge.

7.2 Computational Challenges

The real-time requirements of HFT impose strict constraints on the computational resources available for solving MDPs. The need for rapid decision-making means that MDP solutions must be computed within milliseconds, which is a demanding task given the complexity of the problems.

7.2.1 Real-Time Solvency

MDP algorithms typically require iterative methods to converge to an optimal policy. In HFT, however, there is often insufficient time for these algorithms to converge before the optimal policy changes due to market dynamics [Gould et al., 2013]. This necessitates the development of faster solution methods that can provide near-optimal policies in real-time.

7.3 Market Impact and Adverse Selection

Another challenge in applying MDPs to HFT is the issue of market impact, where large orders can influence the price of an asset. This can lead to suboptimal execution if not properly accounted for in the MDP model Almgren and Chriss [2001]. Additionally, traders using MDP-based strategies may be susceptible to adverse selection, where they are more likely to trade with better-informed participants, resulting in unfavorable outcomes.

7.3.1 Incorporating Market Impact

To mitigate market impact, MDP models must incorporate sophisticated cost functions that account for the dynamic nature of liquidity and the feedback effect of trades on market prices. This requires a deep understanding of market microstructure and liquidity dynamics, which are areas of ongoing research [Cartea et al., 2015].

7.4 Data-Driven and Learning Challenges

MDPs in HFT rely heavily on historical and real-time data for state estimation and policy learning. The quality and availability of data, as well as the ability to learn effectively from it, pose significant challenges.

7.4.1 Data Quality and Availability

High-quality data is essential for accurate state estimation and policy learning in MDPs. However, financial data can be noisy, incomplete, or subject to biases, which can lead to suboptimal policies Hasbrouck [2007]. Furthermore, access to comprehensive datasets may be limited due to proprietary restrictions or cost barriers.

7.4.2 Adaptive Learning

Learning optimal policies in a non-stationary environment requires adaptive algorithms that can update policies as new data becomes available. Reinforcement learning approaches, such as Q-learning and policy gradient methods, have been explored in this context [Sutton and Barto, 2018]. However, ensuring the stability and convergence of these algorithms in the face of market volatility is an ongoing challenge.

The exploration of MDPs in the high-frequency trading arena is a journey through a landscape marked by the peaks of algorithmic innovation and the valleys of computational complexity. As we navigate this terrain, we are reminded that the quest for optimal trading strategies is not merely a technical endeavor but a reflection of the intricate dance between human ingenuity and the enigmatic forces of the market. The limitations and challenges we face serve not as deterrents but as beacons, guiding us toward a deeper understanding of the financial ecosystem and the role of artificial intelligence within it.

8 Future Directions and Innovations

The intersection of Markov Decision Processes (MDPs) and high-frequency trading (HFT) is a fertile ground for innovation, with emerging trends and technologies poised to redefine the landscape. This section explores the potential advancements and novel applications of MDPs in HFT, considering the integration with other artificial intelligence (AI) techniques and the implications of recent technological developments.

8.1 Integration with Machine Learning and AI

The integration of MDPs with advanced machine learning (ML) and AI techniques holds promise for addressing some of the challenges faced in HFT. By leveraging the strengths of both fields, traders can develop more sophisticated and adaptive trading algorithms.

8.1.1 Deep Reinforcement Learning

Deep reinforcement learning (DRL) combines neural networks with reinforcement learning to handle high-dimensional state spaces, which are common in HFT [Mnih et al., 2015]. DRL can approximate optimal policies even in complex and non-stationary environments, potentially overcoming the curse of dimensionality and non-stationarity issues associated with traditional MDPs.

8.1.2 Natural Language Processing for Market Sentiment Analysis

Natural language processing (NLP) techniques can be used to analyze news articles, social media, and other textual data to gauge market sentiment Bollen et al. [2011]. Incorporating sentiment analysis into MDP models could provide a more comprehensive view of the market, allowing traders to anticipate and react to shifts in investor sentiment more effectively.

8.2 Quantum Computing and MDPs

Quantum computing offers a paradigm shift in computational capabilities, with the potential to solve certain classes of problems much faster than classical computers. Quantum algorithms for MDPs could revolutionize HFT by enabling the rapid solution of problems that are currently intractable.

8.2.1 Quantum Speedup for MDPs

Research into quantum algorithms for solving MDPs is in its infancy, but preliminary results suggest that quantum computing could offer a significant speedup for certain optimization problems [Dunjko et al., 2016]. As quantum technology matures, it may become possible to solve MDPs in real-time, even with extremely large state spaces.

8.3 Blockchain Technology and Decentralized Finance

Blockchain technology and the rise of decentralized finance (DeFi) are creating new opportunities and challenges for HFT. MDPs could play a role in navigating the unique characteristics of these emerging markets.

8.3.1 Smart Contracts for Automated Trading

Smart contracts on blockchain platforms enable the execution of trades without the need for traditional intermediaries. MDPs could be used to design automated trading strategies that interact with smart contracts, taking advantage of the transparency and reduced counterparty risk offered by blockchain technology [Clark and Tucker, 2014].

8.4 Regulatory Compliance and Ethical Considerations

As MDPs become more prevalent in HFT, regulatory and ethical considerations will become increasingly important. Ensuring that MDP-based trading strategies comply with regulations and ethical standards is crucial for maintaining market integrity.

8.4.1 MDPs and Market Surveillance

Regulators are turning to advanced technology to monitor trading activities and detect market abuse. MDPs could be used to develop algorithms that assist with market surveillance, helping to identify patterns indicative of manipulative behaviors [Kirilenko and Lo, 2017].

The exploration of MDPs in the context of HFT is akin to charting a course through uncharted waters, where each wave of innovation propels us toward new horizons. As we harness the winds of technological progress and navigate the currents of market complexity, we are reminded that the quest for optimal trading strategies is not a solitary endeavor but a collaborative voyage that draws upon the collective wisdom of the financial and scientific communities. The future of MDPs in HFT is not written in the stars but in the algorithms we create, the data we analyze, and the ethical principles we uphold. It is a future that beckons with the promise of discovery and the potential for transformative change in the world of finance.

9 Algorithmic Complexity and Computational Considerations

The application of Markov Decision Processes (MDPs) in high-frequency trading (HFT) necessitates a careful consideration of algorithmic complexity and computational resources. The real-time nature of HFT, combined with the vast amounts of data and the need for rapid decision-making, presents unique computational challenges. This section dives into the strategies for solving MDPs in high-frequency environments and addresses the computational hurdles inherent in these applications.

9.1 Solving MDPs in High-Frequency Environments

The core challenge in applying MDPs to HFT lies in the need to solve the MDPs quickly and accurately. Traditional solution methods, such as value iteration and policy iteration, can be computationally intensive and may not scale well to the high-dimensional state spaces encountered in HFT Bellman [1957].

9.1.1 Approximate Dynamic Programming

Approximate dynamic programming (ADP) methods have been proposed to handle the complexity of MDPs in HFT. ADP approaches, such as fitted value iteration and temporal-difference learning, use function approximation techniques to estimate the value function or policy without exhaustively exploring the entire state space Powell [2007]. These methods can significantly reduce computational requirements, making them more suitable for the fast-paced HFT environment.

9.1.2 Parallel and Distributed Computing

The use of parallel and distributed computing architectures can further alleviate the computational burden. By distributing the computation across multiple processors or machines, one can achieve near real-time performance for solving MDPs Dean and Ghemawat [2008]. The advent of cloud computing and GPU acceleration has made such distributed approaches more accessible to trading firms.

9.2 Addressing Computational Challenges

Despite advances in algorithms and computing hardware, the computational challenges of applying MDPs in HFT remain significant. The following strategies are critical for managing these challenges:

9.2.1 State Space Reduction

Reducing the state space of the MDP can lead to more tractable computations. Techniques such as state aggregation, feature selection, and dimensionality reduction can simplify the representation of the trading environment without sacrificing significant predictive power Jiang et al. [2015].

9.2.2 Real-Time Learning and Adaptation

In HFT, market conditions can change rapidly, necessitating algorithms that can learn and adapt in real-time. Online learning methods, which update policies on-the-fly based on new data, are essential for maintaining the relevance of MDP-based strategies Sutton and Barto [1998].

9.2.3 Algorithmic Efficiency

Efficient algorithms that can quickly converge to an optimal or near-optimal policy are crucial. Research into more efficient variants of traditional MDP algorithms, such as prioritized sweeping and Monte Carlo tree search, is ongoing and holds promise for HFT applications Moore and Atkeson [1993], Browne et al. [2012].

The quest to harness the power of MDPs in the high-stakes arena of HFT is akin to a grandmaster playing speed chess: every move must be precise, every second counts, and the ability to anticipate and adapt to the opponent's strategies is paramount. As we push the boundaries of what is computationally feasible, we not only refine our trading tactics but also contribute to the broader field of decision-making under uncertainty. The algorithms we develop and the computational innovations we pioneer in the pursuit of optimal trading strategies echo through the corridors of finance and technology, challenging us to think faster, act smarter, and trade wiser.

10 Adaptive Learning in MDPs

Adaptive learning is a cornerstone of modern high-frequency trading (HFT) systems that utilize Markov Decision Processes (MDPs). The dynamic nature of financial markets demands that trading algorithms not only make decisions based on current market conditions but also continuously learn and adapt to new patterns and changes in market dynamics. This section explores the role of reinforcement learning in MDPs and how adaptive strategies are employed to maintain robust performance in the ever-changing landscape of HFT.

10.1 Reinforcement Learning and MDPs

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment [Sutton and Barto, 2018]. In the context of MDPs, RL algorithms are used to find a policy that maximizes the expected cumulative reward, which in HFT corresponds to profit or another financial performance metric.

10.1.1 Q-Learning and Deep Q-Networks

Q-learning is a popular RL algorithm that has been adapted for use in HFT. It involves learning an action-value function that gives the expected utility of taking a given action in a given state, following the optimal policy thereafter [Watkins and Dayan, 1992]. Deep Q-Networks (DQN), which combine Q-learning with deep neural networks, have been particularly influential, enabling the handling of high-dimensional state spaces typical in HFT [Mnih et al., 2015].

10.1.2 Policy Gradient Methods

Policy gradient methods, another class of RL algorithms, directly parameterize the policy and update the parameters by gradient ascent on the expected return. These methods are well-suited for HFT applications where the action space is continuous or when the policy needs to be highly expressive [Sutton et al., 2000].

10.2 Adaptive Strategies in Changing Market Conditions

The financial markets are characterized by non-stationarity, where the underlying probability distributions change over time. Adaptive strategies in HFT must be capable of responding to such changes to maintain their edge.

10.2.1 Meta-Learning for Market Adaptation

Meta-learning, or learning to learn, has emerged as a powerful approach for building adaptive trading systems. By training models on a variety of market scenarios, these systems can quickly adapt to new conditions without extensive retraining [Finn et al., 2017]. Meta-learning algorithms can be integrated with MDP frameworks to create trading strategies that adjust to new market dynamics more effectively.

10.2.2 Online Learning and Exploration-Exploitation Trade-off

Online learning algorithms update the trading strategy incrementally as new data arrives, which is essential in the fast-paced HFT environment. Balancing exploration (trying new actions to discover their effects) with exploitation (using the current best-known strategy) is a critical aspect of online learning in HFT. Techniques such as Thompson sampling and upper confidence bounds (UCB) have been applied to manage this trade-off in the context of MDPs [Agrawal et al., 2012].

The integration of adaptive learning mechanisms into MDP-based HFT systems represents a significant advancement in the quest for automated trading excellence. By embracing the complexity and unpredictability of financial markets, these systems embody the relentless pursuit of adaptability—a trait that is not only vital for survival in the electronic trading arenas but also emblematic of the broader human endeavor to thrive amidst uncertainty. As we continue to refine these adaptive algorithms, we edge closer to the elusive goal of constructing trading systems that can navigate the tumultuous seas of the market with the finesse of a seasoned captain, ever vigilant and ready to adjust their sails to the shifting winds of change.

11 Evaluation of MDP-Based Trading Systems

The evaluation of high-frequency trading (HFT) systems that employ Markov Decision Processes (MDPs) is critical for assessing their performance and viability in the competitive landscape of financial markets. This section dives into the metrics used to evaluate MDP-based HFT systems and presents empirical results and comparative analyses that highlight the strengths and weaknesses of these approaches.

11.1 Performance Metrics for HFT

In evaluating the performance of MDP-based HFT systems, several metrics are commonly used to capture different aspects of trading efficacy. These metrics include financial returns, risk-adjusted returns, transaction costs, market impact, and execution speed.

11.1.1 Financial Returns

The most direct measure of a trading system's performance is its ability to generate profits. Net returns, calculated as the difference between gains and losses over a period, serve as a primary indicator of success [Sharpe, 1994].

11.1.2 Risk-Adjusted Returns

Risk-adjusted returns, such as the Sharpe ratio, take into account the volatility of returns, providing a measure of how much return is achieved per unit of risk taken [Sortino and Price, 1994]. This is particularly important in HFT, where high leverage can amplify both gains and losses.

11.1.3 Transaction Costs

Transaction costs, including commissions, slippage, and bid-ask spreads, can significantly erode profits in HFT. An effective MDP-based system must minimize these costs to improve net performance Almgren and Chriss [2001].

11.1.4 Market Impact

Market impact refers to the effect that a trader's orders have on the market price. Large orders can move the market against the trader, resulting in less favorable execution prices. MDPs can help in designing strategies that minimize market impact [Bertsimas and Lo, 1998].

11.1.5 Execution Speed

In the realm of HFT, where opportunities can vanish in milliseconds, execution speed is a critical performance metric. MDP-based systems must be able to process information and execute orders with minimal latency [Hasbrouck and Saar, 2013].

11.2 Empirical Results and Comparative Analysis

Empirical studies of MDP-based HFT systems provide insights into their real-world performance. These studies often involve backtesting strategies on historical data or paper trading in live markets to assess their effectiveness.

11.2.1 Backtesting Results

Backtesting involves simulating the performance of a trading strategy using historical market data. Studies have shown that MDP-based strategies can outperform traditional strategies in certain market conditions, particularly when markets exhibit mean-reverting behavior or other exploitable patterns [Nevmyvaka et al., 2006].

11.2.2 Live Trading Performance

Live trading results offer the most concrete evidence of a system's performance. MDP-based HFT systems have been reported to achieve higher Sharpe ratios and lower drawdowns compared to non-MDP-based systems, indicating better risk-adjusted returns and lower risk of significant losses [Moallemi and Sağlam, 2013].

11.2.3 Comparative Studies

Comparative studies between MDP-based systems and other algorithmic trading approaches provide a broader context for evaluating performance. While MDP-based systems excel in certain scenarios, they may underperform in markets that are less predictable or when the model assumptions do not hold [Cartea et al., 2015].

The rigorous evaluation of MDP-based HFT systems is a testament to the relentless pursuit of optimization in the financial domain. By dissecting the multifaceted nature of trading performance through a prism of quantitative metrics, we gain a deeper understanding of the intricate dance between risk and reward. These evaluations not only serve as a beacon for guiding the development of future trading algorithms but also reflect the broader human quest for mastering complex systems—a journey marked by the continuous interplay between theoretical innovation and empirical validation.

12 Limitations and Challenges

While Markov Decision Processes (MDPs) have proven to be powerful tools in the realm of high-frequency trading (HFT), they are not without their limitations and challenges. This section examines the constraints of employing MDPs in HFT environments, as well as the current challenges faced by researchers and practitioners in the field.

12.1 Limitations of MDPs in HFT

MDPs rely on certain assumptions and conditions that may not always hold true in the complex and dynamic world of financial markets. These limitations can affect the performance and applicability of MDP-based trading strategies.

12.1.1 Model Assumptions

MDPs assume that the state and action spaces are fully observable and that the transition probabilities are known or can be accurately estimated [Puterman, 2014]. However, in HFT, market conditions can change rapidly, and the assumption of stationary transition probabilities may not be valid. Additionally, the presence of hidden variables and incomplete information can lead to model misspecification.

12.1.2 State Space Dimensionality

The curse of dimensionality is a significant challenge in MDPs, as the state space can grow exponentially with the number of variables considered Bellman [1957]. In HFT, where numerous factors such as price, volume, and order book dynamics need to be considered, the state space can become intractably large, making it difficult to solve the MDP efficiently.

12.1.3 Execution Uncertainty

MDPs in HFT must contend with execution uncertainty. The actual execution price and quantity can deviate from expected values due to market volatility and the actions of other market participants. This uncertainty can lead to suboptimal decisions if not properly accounted for in the MDP model [Cartea et al., 2015].

12.2 Current Challenges and Research Gaps

The application of MDPs in HFT continues to face several challenges that require ongoing research and innovation. Addressing these challenges is crucial for the advancement of MDP-based trading systems.

12.2.1 Real-Time Computation

One of the most pressing challenges is the need for real-time computation. HFT strategies must respond to market events within milliseconds, leaving very little time for complex calculations. Developing algorithms that can solve MDPs in real-time without sacrificing accuracy is an area of active research [Nevmyvaka et al., 2006].

12.2.2 Adaptive Learning

Financial markets are non-stationary, and trading strategies that perform well in one period may fail in another. Adaptive learning mechanisms that can update MDP models in response to changing market conditions are essential for maintaining the relevance and effectiveness of trading strategies [Moallemi and Sağlam, 2013].

12.2.3 Robustness and Generalization

MDP-based strategies must be robust to model misspecification and able to generalize across different market regimes. Research into methods for enhancing the robustness and generalization capabilities of MDPs is critical for their successful application in HFT [Hansen and Sargent, 2001].

12.2.4 Regulatory Compliance

Regulatory considerations also pose a challenge for MDP-based HFT systems. Ensuring compliance with evolving regulations while optimizing trading performance is a delicate balance that requires careful attention to the design and operation of these systems [Menkveld, 2016].

The journey of integrating MDPs into the high-stakes theater of HFT is akin to navigating a labyrinthine network of shifting sands. The inherent limitations and challenges serve as a constant reminder of the intricate interplay between mathematical elegance and market chaos. As researchers and practitioners continue to unravel the complexities of this domain, the pursuit of mastery over these challenges not only fuels the evolution of trading strategies but also reflects the broader human endeavor to harness the power of uncertainty in the quest for progress.

13 Future Directions and Innovations

The intersection of Markov Decision Processes (MDPs) and high-frequency trading (HFT) is a fertile ground for innovation. As the financial landscape evolves, so too must the methodologies and technologies that underpin trading strategies. This section explores the emerging trends in HFT and the potential for MDP applications to adapt and thrive in this dynamic environment.

13.1 Integration with Alternative Data Sources

The incorporation of alternative data sources into MDP models represents a significant opportunity for innovation in HFT. Alternative data refers to information that is not derived from traditional financial sources, such as market data or financial statements. Examples include satellite imagery, social media sentiment, and transactional data from non-market platforms Bhattacharya et al. [2019]. By enriching the state space of MDPs with alternative data, traders can gain unique insights into market movements and potentially uncover new predictive signals.

13.1.1 Sentiment Analysis

Sentiment analysis, for instance, utilizes natural language processing to gauge the mood of market participants from social media and news articles Bollen et al. [2011]. Incorporating sentiment as a state variable in MDPs could enable traders to better anticipate market reactions to events and news releases.

13.1.2 High-Dimensional Data Techniques

To effectively leverage high-dimensional alternative data, advancements in dimensionality reduction and feature selection techniques are necessary. Techniques such as principal component analysis (PCA) and autoencoders can be employed to distill relevant information and mitigate the curse of dimensionality Jolliffe [2016].

13.2 Quantum Computing in MDP Optimization

Quantum computing holds the promise of revolutionizing HFT by providing unprecedented computational capabilities. Quantum algorithms have the potential to solve certain classes of optimization problems, including those found in MDPs, exponentially faster than classical algorithms Orús et al. [2019]. As quantum technology matures, it may enable real-time solutions to MDPs with vast state spaces that are currently intractable.

13.2.1 Quantum Machine Learning

Quantum machine learning algorithms could be particularly impactful in the realm of adaptive learning for MDPs. By exploiting quantum parallelism, these algorithms can analyze vast datasets more efficiently, leading to more accurate and timely updates to trading strategies Biamonte et al. [2017].

13.3 Synergy with Other AI Techniques

The synergy between MDPs and other artificial intelligence (AI) techniques, such as deep learning and reinforcement learning, is an area ripe for exploration. Deep learning can enhance the function approximation capabilities within MDPs, enabling the modeling of complex market dynamics Deng et al. [2016]. Reinforcement learning, which is closely related to MDPs, can be used to develop adaptive trading algorithms that learn optimal strategies through interaction with the market Li et al. [2019].

13.3.1 Deep Reinforcement Learning

Deep reinforcement learning, which combines deep learning with reinforcement learning, has shown promise in mastering complex games and simulations. Applying deep reinforcement learning to MDPs in HFT could lead to the development of sophisticated strategies that can navigate the intricacies of the market with a level of nuance previously unattainable Silver et al. [2017].

The horizon of high-frequency trading is continuously expanding, driven by relentless innovation and the quest for competitive advantage. As MDPs evolve in tandem with emerging technologies and data sources, they will undoubtedly play a pivotal role in shaping the future of trading. The fusion of quantum computing, alternative data, and advanced AI techniques with MDPs is not merely an incremental step but a leap towards a new paradigm in financial markets—one where the boundaries of speed, efficiency, and intelligence are constantly being redefined.

References

Dimitris Bertsimas and Andrew W. Lo. Optimal control of execution costs. *Journal of Financial Markets*, 1(1):1–50, 1998.

Vedran Dunjko, Jacob M. Taylor, and Hans J. Briegel. Quantum-enhanced machine learning. *Physical Review Letters*, 117(13):130501, 2016.

Zihao Gao, Bin Li, and Jaksa Cvitanic. Deep Learning for Optimal Dynamic Pricing and Inventory Control. In *Management Science*, 2018.

Andrei Kirilenko and Andrew W. Lo. Flash crash: The impact of high frequency trading on an electronic market. *The Journal of Finance*, 72(3):967–998, 2017.

Rama Cont and Peter Tankov. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

Frank A. Sortino and Lee N. Price. Performance measurement in a downside risk framework. *Journal of Investing*, 3(3):59–64, 1994.

Richard Bellman. Dynamic Programming. *Princeton University Press*, 1957.

Albert J. Menkveld. The Economics of High-Frequency Trading: Taking Stock. *Annual Review of Financial Economics*, 8: 1–24, 2016.

Martin D. Gould, Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison. Limit order books. *Quantitative Finance*, 13(11):1709–1742, 2013.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2017.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 1126–1135, 2017.

Jeffrey Dean and Sanjay Ghemawat. MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*, 51(1):107–113, 2008.

Bin Li, Steven C.H. Hoi, and Peilin Zhao. Deep Reinforcement Learning for Online Investment. In *IEEE Transactions on Neural Networks and Learning Systems*, 2019.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *In Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 1126–1135, 2017.

Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.

Álvaro Cartea, Sebastian Jaimungal, and José Penalva. Algorithmic and High-Frequency Trading. *Cambridge University Press*, 2015.

Yong Deng, Zhenyu Lu, and Yuhong Liu. Deep Learning in Finance. *Springer Briefs in Computer Science*, pages 1–19, 2016.

Bin Li, Steven C.H. Hoi, and Vivekanand Gopalkrishnan. Deep Learning for Portfolio Optimization. *Journal of Financial Data Science*, 1(1):20–36, 2019.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

Irene Aldridge. *High-Frequency Trading: A Practical Guide to Algorithmic Strategies and Trading Systems*. Wiley Trading, 2nd edition, 2013.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns.
Reinforcement learning for optimized trade execution.
In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680. ACM, 2006.

William F. Sharpe. The Sharpe Ratio. *The Journal of Portfolio Management*, 21(1):49–58, 1994.

Nan Jiang, Alex Kulesza, Satinder Singh, and Richard Lewis. The Dependence of Effective Planning Horizon on Model Abstraction. *In Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 1181–1189, 2015.

Martin L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. *John Wiley & Sons*, 2014.

Peter A. Forsyth and Kenneth R. Vetzal. Optimal trade execution: a mean-quadratic-variation approach. *Journal of Economic Dynamics and Control*, 36(12):1971–1991, 2012.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Lars Peter Hansen and Thomas J. Sargent. Robust Control and Model Uncertainty. *American Economic Review*, 91(2): 60–66, 2001.

Richard S. Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. *In Advances in Neural Information Processing Systems*, pages 1057–1063, 2000.

Ciamac C. Moallemi and Mehmet Saglam. The cost of latency in high-frequency trading. *Operations Research*, 61(5):1070–1086, 2013.

Álvaro Cartea, Sebastian Jaimungal, and José Penalva. Algorithmic and High-Frequency Trading. *Cambridge University Press*, 2015.

Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. *The MIT Press*, 2nd edition, 2018.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement learning for optimized trade execution. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680, 2006.

Christopher J.C.H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.

Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–39, 2001.

Xingyu Gao and Laiwan Chan.
Deep reinforcement learning for automated stock trading: An ensemble strategy.
SSRN Electronic Journal, 2018.

Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. *MIT Press*, 2nd edition, 2018.

Shipra Agrawal and Navin Goyal. Analysis of Thompson Sampling for the multi-armed bandit problem. *In Proceedings of the 25th Annual Conference on Learning Theory*, volume 23, pages 39.1–39.26, 2012.

R. Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2(3):21–41, 2000.

Ian T. Jolliffe. Principal Component Analysis. *Springer Series in Statistics*, pages 1–487, 2016.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement Learning for Optimized Trade Execution. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680, 2006.

William F. Sharpe. The Sharpe Ratio. *Journal of Portfolio Management*, 21(1):49–58, 1994.

Joel Hasbrouck and Gideon Saar. Low-latency trading. *Journal of Financial Markets*, 16(4):646–679, 2013.

Román Orús, Samuel Mugel, and Enrique Lizaso. Quantum computing for finance: Overview and prospects. *Reviews in Physics*, 4:100028, 2019.

Nicole Bäuerle and Ulrich Rieder. *Markov Decision Processes with Applications to Finance*. Springer-Verlag, Berlin, Heidelberg, 2011.

Richard Bellman. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6(5):679–684, 1957.

David Easley, Marcos M. Lopez de Prado, and Maureen O'Hara. Flow toxicity and liquidity in a high-frequency world. *Review of Financial Studies*, 25(5):1457–1493, 2012.

Joel Hasbrouck. Empirical Market Microstructure: The Institutions, Economics, and Econometrics of Securities Trading. *Oxford University Press*, 2007.

Álvaro Cartea, Sebastian Jaimungal, and José Penalva. Algorithmic and High-Frequency Trading. *Cambridge University Press*, 2015.

Cameron B. Browne, Edward Powley, Daniel Whitehouse, Simon M. Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A Survey of Monte Carlo Tree Search Methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012.

Álvaro Cartea, Sebastian Jaimungal, and José Penalva.
Algorithmic and High-Frequency Trading.
Cambridge University Press, 2015.

Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–39, 2001.

Sourish Bhattacharya, Douglas M. Patterson, and Ravi R. Mazumdar. Using Markov Decision Processes for Dynamic Portfolio Optimization with Transaction Costs and Model Uncertainty. *Management Science*, 65(5):2146–2162, 2019.

Malik Magdon-Ismail, Amir F. Atiya, Amrit Pratap, and Yaser S. Abu-Mostafa. On the Maximum Drawdown of a Brownian Motion. *Journal of Applied Probability*, 41(1):147–161, 2004.

Ciamac C. Moallemi and Mehmet Sağlam. The Convergence of a Class of Double-Sequence Iterative Methods for Solving Continuous-Time Markov Decision Processes. *Operations Research*, 61(3): 615–628, 2013.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement learning for optimized trade execution. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680, 2006.

Joseph Clark and Charles Tucker. Decentralizing markets: How blockchain technology is transforming the sharing economy. *Technology Innovation Management Review*, 4(7):12–18, 2014.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

Ciamac C. Moallemi and Mehmet Sağlam. The cost of latency in high-frequency trading. *Operations Research*, 61(5):1070–1086, 2013.

Warren B. Powell. Approximate Dynamic Programming: Solving the Curses of Dimensionality. *Wiley Series in Probability and Statistics*, 2007.

Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. *MIT Press*, 1998.

David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, 2017.

Michel Tokic. Adaptive ϵ-greedy exploration in reinforcement learning based on value differences. In *Annual Conference on Artificial Intelligence*, pages 203–210, 2010.

Marco Avellaneda and Sasha Stoikov. High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224, 2008.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement Learning for Optimized Trade Execution. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680, 2006.

Andrew W. Moore and Christopher G. Atkeson. Prioritized Sweeping: Reinforcement Learning with Less Data and Less Time. *Machine Learning*, 13(1):103–130, 1993.

Marco Avellaneda and Sasha Stoikov.
High-frequency trading in a limit order book.
Quantitative Finance, 8(3):217–224, 2008.

Martin D. Gould, Mason A. Porter, Stacy Williams, Mark McDonald, Daniel J. Fenn, and Sam D. Howison.
Limit order books.
Quantitative Finance, 13(11):1709–1742, 2013.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal Policy Optimization Algorithms. *arXiv preprint arXiv:1707.06347*, 2017.