"Optimization of Trading Systems and Portfolios" by John Moody and Lizhong Wu (1997), without using mathematical notation.

1. Introduction: Objective Functions and Reinforcement Learning for Trading

The core idea of this paper is that many trading systems rely on forecasting techniques or supervised learning, which involve training the system based on historical data. These systems typically attempt to forecast future price movements and act accordingly. However, the authors argue that this approach has limitations because the main goal of trading is not just to forecast prices but to make profitable decisions based on the current market state.

Instead of focusing on prediction alone, the paper proposes the use of reinforcement learning (RL). In RL, the system learns by interacting with the market environment and adjusting its strategies based on the feedback (rewards or penalties) it receives from the market. The system learns through trial and error, continuously refining its actions to maximize its reward over time. This approach is more aligned with the trader's ultimate goal—profit maximization—since it focuses directly on optimizing the trading system's decisions rather than relying on price forecasts.

The paper argues that reinforcement learning is ideal for optimizing trading systems because it allows the agent (the trading system) to take actions based on the current state of the market, receiving immediate feedback and adjusting its future decisions accordingly. This avoids the problem of relying on inaccurate predictions or forecasts and adapts the system based on real market behavior.

2. Single Asset with Discrete Position Size

In this section, the authors delve into the optimization of trading systems that focus on a single asset (such as a stock or a commodity). The basic concept involves adjusting the position size based on the current market conditions and a set of predefined rules.

A trader can take three types of positions:

1. Long Position – buying the asset.
2. Short Position – selling the asset.
3. Neutral Position – holding no asset.

The paper assumes that the trader's position is discrete and of a constant magnitude. This means that at any given time, the position taken (long, neutral, or short) is of a fixed size. For example, the trader might decide to buy or sell exactly one unit of the asset at a time. The paper also discusses how the trader can manage risk by keeping the position size constant, which allows for easier risk control. The position is reassessed at the end of each time period to determine whether it should be adjusted.

However, there is a practical concern regarding transaction costs and market impact. These costs arise from executing trades, and they reduce the overall profitability of a trading strategy. The paper emphasizes the importance of incorporating these factors into the decision-making process. Specifically, the decision to trade must consider not just the expected reward (profit) but also the costs involved in executing the trade, as well as how the trade might affect the market price.

3. Optimizing Profit and Wealth

The paper also explores different ways to optimize the trader's profit and wealth. The two main types of objectives discussed are:

1. Additive Profits: These are used when trading involves a fixed number of shares or contracts. The idea is that profits are simply added up for each trade, and the trader's wealth increases by the accumulated profits from each successful trade.
2. Multiplicative Profits: This approach is used when the wealth of the trader is not fixed but grows over time, and profits are compounded. The total wealth increases as the profits from each trade multiply.

The paper suggests that transaction costs should be incorporated into the wealth calculations, as they influence how often the system should trade. The system needs to decide when the expected profit from a trade is sufficient to justify the cost of executing the trade. Therefore, maximizing profit must be balanced with the cost of trading.

The Sharpe ratio is highlighted as an important metric for evaluating the performance of a trading strategy. The Sharpe ratio measures the risk-adjusted return, helping the trader to understand how much return is being generated for each unit of risk taken. The authors argue that, instead of just maximizing profit, focusing on optimizing the Sharpe ratio might lead to more consistent and robust strategies. This is because optimizing for the Sharpe ratio ensures that the trader balances both return and risk.

4. Reinforcement Learning for Optimization

A significant portion of the paper focuses on how to apply reinforcement learning to optimize trading strategies. In traditional machine learning, a model might be trained on historical data to predict future price movements. However, in RL, the agent learns by interacting with the environment (the market) and receiving feedback on its actions. The agent doesn't rely on fixed historical data or forecasts; instead, it takes actions (buy, sell, or hold) and learns from the outcomes (profits or losses).

The paper emphasizes that RL models are especially useful for sequential decision-making problems, such as trading, where the outcome of each decision depends on previous actions. The RL model adjusts its decision-making over time to improve performance, and this makes it particularly well-suited for markets that are inherently dynamic and non-stationary.

Additionally, the authors explain the concept of recurrent learning algorithms. These are used to model the system's decision-making process in a way that accounts for past trades. By using recurrent algorithms, the system can learn from past actions, adjust to market conditions, and make decisions that are based on a combination of past experiences and current market conditions.

5. Conclusions and Future Work

The paper concludes by discussing the potential for future research in the area of reinforcement learning and trading system optimization. The authors suggest that further studies could focus on refining the differential Sharpe ratio and its role in performance measurement. Additionally, they propose expanding the scope of RL-based models to accommodate more

complex market environments.
They also recommend using real-time market data to train and test RL models, which could lead to better performance in live trading scenarios. In particular, they highlight the importance of adaptive learning systems that can adjust their strategies based on evolving market conditions.

Key Contributions of the Paper

1. RL for Trading: The paper introduces the idea of using reinforcement learning to optimize trading systems, offering a solution that can adapt to dynamic market conditions.
2. Objective Functions: The authors propose multiple objective functions that can be optimized, including profit, wealth, and the Sharpe ratio, and they discuss the advantages of focusing on risk-adjusted returns.
3. Practical Implementation: The paper provides a framework for incorporating real-world factors like transaction costs and market impact into trading strategies, making the RL model more practical and applicable to real-world trading.
4. Future Research: The authors highlight the need for further research into the application of RL to portfolio management, particularly with regard to risk management and adapting the model to different asset classes.

This paper laid the groundwork for the use of reinforcement learning in algorithmic trading, influencing the development of modern trading systems that are capable of adapting to changing market conditions.

The sections of the paper you uploaded focus on portfolio optimization and economic utility in the context of trading systems.

3. Structure and Optimization of Portfolios

3.1 Portfolios: Continuous Quantities of Multiple Assets

In the context of portfolio optimization, when considering multiple assets, traders often face a decision-making process that involves continuous quantities of assets. Unlike simple single-asset models, where the focus is on taking long or short positions in a single asset, multiple asset portfolios require managing the weights of each asset within the portfolio. The paper discusses how portfolio optimization is typically modeled using mean-variance optimization (originally proposed by Harry Markowitz), which is used to balance the expected return of the portfolio with its risk.

The challenge arises because portfolio weights must be adjusted at each step, and the sum of these weights must always add up to 1, ensuring that the trader is fully invested. The weights represent how much of each asset the trader holds in the portfolio, and changes in these weights are required as asset prices fluctuate.

The authors introduce an alternative method for optimizing portfolio weights by using a softmax output. The softmax function is used here to transform raw inputs into probabilities, which in turn determine the portfolio weights. This allows the system to avoid the constraints typically involved in traditional optimization problems, such as ensuring that the weights sum to 1. Instead, the system can automatically learn the appropriate portfolio weights by adjusting based on the feedback from the market. The paper suggests that the weights of each asset can evolve continuously as the trading system adapts to changing market conditions.

3.2 Profit and Wealth for Portfolios

When multiple assets are considered, managing the portfolio requires adjusting the positions held in each asset based on their price movements. The wealth of the portfolio is calculated after accounting for the cumulative profits from each trade and the cost of rebalancing the portfolio. Each adjustment in the portfolio requires taking into account the transaction costs involved in buying or selling the assets, as well as the potential reduction in wealth due to these costs.

The paper outlines a method to calculate the profit and wealth for portfolios, where the portfolio value is updated after each period based on the current portfolio weights. As the portfolio rebalances, the system computes the changes in wealth, considering the cumulative returns from each asset. The key insight is that, when optimizing portfolios, the system must continuously adjust the portfolio to maintain desired asset allocations and maximize the overall wealth.

In essence, the process involves rebalancing the portfolio periodically to ensure that the desired portfolio weights are maintained. This rebalancing takes place after each period, during which the system evaluates the returns from each asset, adjusts the asset allocations, and computes the updated wealth of the portfolio.

4. Optimizing Economic Utility

Optimizing Profit and Wealth

The paper explains that the optimization of trading systems and portfolios often assumes that investors are risk-insensitive, meaning they are focused solely on maximizing their profits or wealth. However, in reality, most investors are risk-averse, meaning they are more sensitive to losses than to gains. As a result, they may be willing to give up some potential gains to avoid taking on too much risk.

To model this risk aversion, the paper discusses the concept of utility functions, which capture the trader's or investor's risk preferences. A utility function maps a trader's wealth to a value that represents their satisfaction or utility, with higher values indicating greater satisfaction. The utility function can capture varying degrees of risk sensitivity, allowing the system to adjust the trader's behavior based on their risk tolerance.

One of the most commonly used utility functions is the logarithmic utility function, which is often associated with constant relative risk aversion. The function indicates that as wealth increases, the investor becomes less sensitive to risk, meaning they are willing to take on more risk as they accumulate more wealth. The paper also introduces gamma utility functions as a more general class of utility functions, where the parameter gamma (
γ
\gamma
) controls the degree of risk aversion. For example, when
γ
=

$\gamma = 1$, the utility function becomes logarithmic, indicating neutral risk aversion. If $\gamma < 1$, the investor is more risk-averse, and if $\gamma > 1$, the investor is more risk-seeking.

Risk Aversion and Economic Utility

The paper also introduces the concept of risk-adjusted utility, which accounts for the trade-off between risk and reward. For example, when the trader takes on more risk, the expected utility of the trade may increase, but the variance of the returns also increases. The paper provides a framework to model this risk-return trade-off by using utility functions, which allow the trading system to adjust based on the trader's risk tolerance.

This concept of utility functions is critical for creating economic models of trading, as it allows the system to optimize for both profitability and risk simultaneously. The system can adjust its trading strategy to achieve a balance between maximizing wealth and minimizing risk, in accordance with the trader's preferences.

The sensitivity to risk can also be adjusted on a per-trade basis. For instance, some trades might be deemed more risky than others, and the system can adjust the trading behavior accordingly. The utility function allows for dynamic adaptation to changing market conditions, ensuring that the trader's actions align with their long-term goals and risk tolerance.

Contributions and Insights

1. Portfolio Optimization: The paper presents methods for optimizing portfolios of multiple assets, incorporating transaction costs and market fluctuations into the decision-making process. By using softmax outputs, the system can dynamically adjust portfolio weights without relying on traditional constrained optimization techniques.
2. Utility Functions: The authors provide a detailed discussion of utility functions and their role in modeling risk aversion. These functions allow the system to balance profit maximization with risk management, ensuring that the trader's preferences are taken into account.
3. Economic Utility Models: By introducing the concept of economic utility, the paper highlights the importance of risk-adjusted returns and provides a framework for optimizing wealth in a way that aligns with the trader's risk preferences.

In conclusion, the paper outlines strategies for optimizing trading systems and portfolios by considering both profit and risk. By integrating reinforcement learning with utility functions, the proposed system can adapt to market conditions, account for risk, and optimize the trader's long-term wealth.

5. The Sharpe Ratio and the Differential Sharpe Ratio

5.1 Optimizing the Sharpe Ratio

The Sharpe ratio is a commonly used measure for evaluating the risk-adjusted return of an investment or trading strategy. It was introduced by William Sharpe in 1966 and is used to understand the return of an investment compared to its risk (volatility). The ratio is defined as the ratio of the average return of a trading strategy to the standard deviation of those returns. In simpler terms, it shows how much return an investor can expect per unit of risk taken.

The objective here is to maximize the Sharpe ratio by adjusting the trading system parameters. The trading system's returns over time are denoted by

represents time periods. The Sharpe ratio is defined as the average return divided by the standard deviation of returns, which gives us a measure of how much return the system produces for each unit of risk.

The system can be trained to maximize this ratio by evaluating the returns over several periods and adjusting the parameters to achieve higher returns for the same level of risk. In other words, the goal is to maximize the risk-adjusted return of the system. The Sharpe ratio is calculated using the first and second moments (average and variance) of the returns distribution. The use of these moments allows the system to take into account both the mean return and the volatility (how spread out the returns are) when determining the performance of the trading system.

To perform the optimization, the system needs to compute the Sharpe ratio continuously and adjust the trading parameters to maximize the ratio, typically using gradient descent or other optimization methods.

5.2 Running and Moving Sharpe Ratios

For online learning—which involves continually adapting the system to new data—the moving Sharpe ratio is employed. This method calculates the Sharpe ratio over a moving window of time, continuously updating as new data comes in. The purpose of this is to ensure that the trading system can adapt to changing market conditions, which is critical for long-term performance.

To achieve this, a recursive estimate of the first and second moments (mean and variance) of the returns is used. This method updates the Sharpe ratio incrementally, allowing the system to adjust its parameters without waiting for the entire dataset to be processed at once. The exponential moving average method is used to give more weight to recent returns while still considering older ones, ensuring the Sharpe ratio reflects more current market conditions.

5.3 Differential Sharpe Ratios for On-Line Optimization

To further enhance the performance of the system during online optimization, the paper introduces the differential Sharpe ratio. This approach focuses on how much the Sharpe ratio will change if the parameters of the trading system are adjusted. By considering the differential change in the Sharpe ratio, the system can fine-tune its parameters more effectively, improving real-time decision-making and responsiveness to market shifts.
This method is particularly useful because it allows the trading system to adjust its learning process based on the most recent information and continue optimizing the Sharpe ratio in real-time. This approach contrasts with traditional optimization methods that require large amounts of historical data to perform optimally, making the differential Sharpe ratio more adaptable to live market conditions.
Key Insights
1. Maximizing the Sharpe Ratio: The paper emphasizes maximizing the Sharpe ratio as the key objective of the trading system. This helps ensure that the system achieves high returns without taking on excessive risk.
2. Online Learning with Moving Sharpe Ratios: The use of moving Sharpe ratios ensures that the system can adapt to changing market conditions in real time. By continuously updating the ratio using recursive calculations, the system is able to stay responsive to new data without the need for retraining with large datasets.
3. Differential Sharpe Ratios: The introduction of differential Sharpe ratios enables the system to optimize its parameters more efficiently, adjusting the trading strategy incrementally based on the most recent performance metrics.
This approach to optimizing the Sharpe ratio and introducing differential measures is critical for creating a self-adjusting, real-time trading system capable of performing well in dynamic markets. The focus on risk-adjusted returns ensures that the system not only seeks profit but does so with an understanding of the risks involved, leading to more sustainable long-term profitability.

5. The Sharpe Ratio and the Differential Sharpe Ratio (Continued)
5.1 Optimizing the Sharpe Ratio (Continued)
The Sharpe ratio is a measure of how much return an investor can expect to earn per unit of risk. The paper focuses on optimizing the Sharpe ratio, which is important for understanding the performance of a trading strategy that accounts for both profits and risk. As mentioned, the Sharpe ratio is the ratio of the average return to the standard deviation of returns. Optimizing this ratio is key for improving the risk-adjusted return of the trading system.
To optimize the Sharpe ratio, the paper suggests expanding the equation for the Sharpe ratio into a differential form that considers how small changes in the trading system parameters (such as weights or positions) influence the ratio. By calculating the first-order change in the Sharpe ratio, the system can adjust its parameters to move in the direction that maximizes the risk-adjusted return.
The influence of risk and return on the differential Sharpe ratio is outlined in the paper. The system penalizes large values of the returns that exceed the moving average, thereby minimizing risk and making the system more conservative when market fluctuations are large. The Differential Sharpe ratio provides a more responsive metric, enabling the system to adjust based on past performance and adapt more effectively to changes in market conditions.
5.2 Running and Moving Sharpe Ratios
For online learning, where the system is updated in real-time, the paper introduces the concept of a running Sharpe ratio. This method calculates the Sharpe ratio over a moving window of time, using recursive estimates of the first and second moments of returns. By updating the Sharpe ratio as new data arrives, the system can adapt dynamically to market changes without having to recalculate everything from scratch.
An exponential moving average is used to smooth the Sharpe ratio over time, giving more weight to recent observations while still considering older data. This adjustment helps the system become more sensitive to recent market behavior, allowing for real-time optimization.
5.3 Differential Sharpe Ratios for On-Line Optimization
In addition to optimizing the Sharpe ratio, the paper introduces differential Sharpe ratios for online optimization. The key difference here is that the system doesn’t just compute the Sharpe ratio at each time step but also considers how small changes in the system parameters (such as position sizes or weights) will affect the Sharpe ratio moving forward.
This allows the system to adjust dynamically based on real-time performance, making it more responsive and adaptive to changing market conditions. By optimizing the differential Sharpe ratio, the system can fine-tune its parameters in an ongoing manner, ensuring that it performs optimally throughout the trading process.
6. Empirical Results
6.1 Data
The authors test their methods by generating a log price series using an artificial market model. The data is generated using an autoregressive trend process, where the price of the asset is influenced by a combination of past values and random noise. The paper uses this synthetic data to simulate market conditions and test the performance of the proposed trading strategies.
The artificial price series is modeled using two parameters, where one of the parameters represents the autoregressive trend, and the other introduces random noise into the system. This noisy price series simulates real market conditions, where prices tend to follow trends but are subject to random fluctuations.
6.2 Simulated Trading Results
The paper presents the results of a simulated trading system designed to optimize the differential Sharpe ratio in an artificial market. The trading system is initialized randomly, and it learns through real-time recurrent learning. The goal of the system is to maximize the differential Sharpe ratio while adapting to the changing market environment.
The trading system's performance is evaluated based on several metrics, including cumulative profits, trading signals, and the Sharpe ratio. The system’s performance over time is shown in graphs, demonstrating how it evolves as it learns from market

data. In particular, the system starts off performing poorly, as it has no initial knowledge of the market. However, after some learning, it begins to adapt, and its performance improves significantly.

Comparison of Trading Systems

The paper compares three types of trading systems:

1. Max.SR (Maximize Sharpe Ratio): This system aims to maximize the differential Sharpe ratio and is trained to adjust its parameters accordingly.
2. Max.Profit: This system focuses solely on maximizing profits, without explicitly considering risk.
3. Min.MSE (Minimize Mean Squared Error): This system tries to minimize the forecasting error by adjusting its parameters to reduce the discrepancy between predicted and actual returns.

The results show that the Max.SR and Max.Profit systems significantly outperform the Min.MSE system. The Max.SR system is particularly effective at maximizing risk-adjusted returns over time, and it outperforms the Max.Profit system slightly in terms of mean returns. Both Max.SR and Max.Profit demonstrate better consistency and stability than Min.MSE, particularly when tested over multiple trials.

Key Insights

1. Optimizing Risk-Adjusted Return: The differential Sharpe ratio provides a more dynamic way to optimize trading strategies. By focusing on maximizing risk-adjusted returns, the system can adjust in real-time to market changes and improve its long-term performance.
2. Empirical Validation: The use of simulated trading results shows that the system trained to maximize the differential Sharpe ratio performs better than traditional systems that optimize solely for profits or forecasting accuracy.
3. Comparison of Trading Systems: The empirical results highlight the importance of risk-adjusted performance metrics like the Sharpe ratio. Systems that focus on maximizing this ratio consistently outperform systems that optimize for profit alone, suggesting that risk management is a key factor in long-term trading success.

In conclusion, the paper demonstrates that by optimizing the differential Sharpe ratio through reinforcement learning, a trading system can adapt to changing market conditions, reduce risk, and maximize long-term returns. The findings suggest that risk-adjusted performance metrics, such as the Sharpe ratio, should be central to the design of trading strategies.