Optimizing Automated Trading Systems Portfolios with Reinforcement Learning for Risk Control

Ramon de Cerqueira Silva[1*], Carlos Alberto Rodrigues[1]
[1]State University of Feira de Santana, Exact Sciences Department; Feira de Santana, Bahia, Brazil

This work proposes an innovative method for optimizing Automated Trading Systems (ATS) portfolios with advanced Deep Reinforcement Learning (DRL) techniques. The algorithms A2C, DDPG, PPO, SAC, and TD3 are assessed for their ability to learn and adapt in volatile market conditions. The main goal is to enhance ATS's risk control and operational efficiency using data from the Brazilian stock market. DRL models outperformed traditional benchmarks by offering better risk management and risk-adjusted returns. The findings demonstrate the potential of DRL algorithms in complex financial scenarios and lay the groundwork for future research on integrating machine learning in quantitative finance.
Keywords: Computational Finance. Machine Learning. Reinforcement Learning.

Reinforcement Learning (RL) has emerged as a powerful tool for tackling challenges in various areas, including real-time decision-making and stock market predictions. In RL, an agent learns to maximize rewards by interacting with the environment, making it promising for financial applications, such as automated trading [1]. ATS (Automated Trading Systems) uses algorithms to make buy-and-sell decisions based on real-time market data. However, volatile markets present risks, requiring constantly optimizing these systems [2]. The application of RL in these models offers an adaptive approach, allowing greater flexibility and efficiency in trading.

Moreover, RL stands out by eliminating the need for intermediate predictions and dynamically adapting to market changes. Recent studies demonstrate that RL outperforms traditional approaches in profitability and effectiveness, proving to be a robust technique in areas such as high-frequency trading and portfolio management [3]. Studies show that these strategies surpass traditional approaches regarding profitability and effectiveness [1,4,5].

Received on 18 September 2024; revised 22 October 2024.
Address for correspondence: Ramon de Cerqueira Silva. State University of Feira de Santana; Av. Transnordestina, S/N. Zipcode: 48.000-000. Feira de Santana, Bahia, Brazil. E-mail: ramondecerqueirasilva@gmail.com.

J Bioeng. Tech. Health 2024;7(Suppl 2):31-38

This study aims to explore the optimization of RL models in an ATS portfolio, comparing them with the traditional approach without optimization and market indices to verify the advantages and limitations of RL in risk control.

Materials and Methods

This section outlines the methods employed in the study to thoroughly evaluate and compare the performance of different trading strategies. It includes a detailed description of the dataset, the preprocessing steps, the approach to portfolio optimization using RL, the proposed environment for trading simulations, and the training process of the DRL (Deep Reinforcement Learning) agents.

Dataset

The algorithms are applied using the FinRL library, which specializes in DRL for automated stock trading [6]. For optimization, backtests are performed, which consist of simulations based on historical data of how a proposed portfolio would have behaved if it had been implemented over a past period. Based on this, backtests comprise historical strategy data and daily returns of the Brazilian Stock Index, IBOVESPA.

Each trade involves two mini contracts of the IBOVESPA index (WIN) or USDBRL (WDO), with the historical data covering 20,644 trades.

In total, 26 strategies are considered, with 18 applied to WIN contracts and 8 to WDO contracts. These strategies include both trend-following techniques and oscillators, and except for one strategy, all trades belong to the day trade category. These day trade strategies use 15 and 20-minute timeframes, allowing granular analysis and rapid execution of operations throughout the day. Using short timeframes is fundamental to capturing intraday price movements and taking advantage of profit opportunities in periods of high volatility [7].

The finance library is used to obtain data from the main Brazilian stock market index, IBOVESPA, through the Exchange Traded Fund (ETF) BOVA11, which aims to replicate the performance of the IBOVESPA, representing the other dataset to be used as a backtest of trading performance. The data is accessed from June 6, 2018, to November 11, 2019.

The diversity of metrics allows the application of technical analysis techniques and the simulation of backtests to evaluate the historical performance of the proposed strategies.

Data Preprocessing

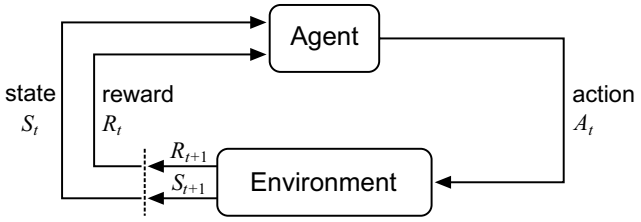
The initial historical dataset consists of detailed records of operations, including the type of operation (buy or sell), dates, entry and exit prices, results in terms of profit or loss, traded volumes, and the identification of the ATS involved. After the entire process, the generated dataset includes the date, strategy name, and daily operation profit (or loss) expressed in financial amounts. The column "data" is renamed to "date" to standardize the column names. Next, a transformation is performed via DataFrame to convert its wide format to a long format, where each row represents the profit for a specific date and ATS, indicated by the columns "close" and "tic," respectively. In addition, technical indicators are calculated to enhance analyses and assist in making trading decisions.

Portfolio Optimization using RL

One approach to solving the portfolio optimization problem is using a RL agent. In this method, the agent develops a policy by interacting directly with an environment. At each time interval, the environment provides observations that define the system's state. Based on this state, the agent decides which action to take. After the action is executed, the environment returns a reward, allowing the agent to evaluate the effectiveness of the chosen action. The goal of the RL agent is to develop a policy that maximizes the expected sum of rewards over time [8]. Figure 1 illustrates the training cycle of an RL agent interacting with the environment.

However, the RL agent needs to handle

Figure 1. Elements of RL [9].



complex state spaces to be effective in portfolio optimization tasks. A portfolio consists of multiple assets, each with its series of prices, resulting in a highly dimensional state space. Using function approximations, such as neural networks (NN), has shown notable results in various complex tasks [8]. Thus, DRL algorithms are the most suitable for this purpose.

Proposed Environment

Thus, it is necessary to design an automated trading solution for portfolio allocation. Stock trading is modeled as a Markov Decision Process (MDP), involving the observation of changes in stock prices, taking actions, and calculating rewards to adjust the agent's trading strategy [8]. All preprocessing is performed to ensure that the ATS trading data aligns with this environment, where the agent can interact and learn, considering

crucial elements such as historical stock prices and technical indicators [6].

To train a trading agent with DRL, an environment simulates real-world trading using OpenAI Gym [10]. The environment initializes by loading market data for the current day and configuring the initial state, which includes the covariance matrix and technical indicators. At each step, the agent makes allocation decisions, which are normalized and applied to calculate the weight of each ATS in the portfolio. The portfolio value is updated according to this balancing, and the reward is defined as the new value. If the episode ends, accumulated and daily reward graphs are saved, and statistics such as the sharpe ratio are calculated and displayed. The environment is then reset for a new episode.

Training the DRL Agent

The implementation of the following DRL algorithms is based on Stable Baselines. Stable Baselines is a fork of OpenAI Baselines, with significant structural refactoring and code cleanup [9]. From this library, all DRL algorithms, A2C, DDPG, PPO, SAC, and TD3, are implemented due to their widespread use in finance [6,11,12].

The selection of these algorithms for an RL agent is based on their ability to handle continuous action spaces, sample efficiency, training stability, and robust performance. Each of these algorithms brings unique characteristics that can be explored to develop effective and adaptive trading strategies, allowing agents to learn and optimize their policies efficiently and robustly [13].

This splits the ATS trading data by date, with the training period covering September 1, 2014, to June 5, 2018. For testing, the period from June 6, 2018, to November 11, 2019, is used, totaling 18,486 trades for training and 10,218 for testing, with an approximate ratio of 65% and 35%, respectively. To generate the DRL models using the FinRL library [6], the training parameters for all agents must be imported and configured. The Optuna[14] library optimizes the hyperparameters to improve the DRL agents' performance.

After training, each model is used to predict the performance of the ATS portfolio in the defined environment using the ATS dataset, specifically between June 6, 2018, and November 11, 2019. This generates two sets of results: daily returns and actions taken by the models. These results help evaluate the effectiveness of each model's strategy in terms of maximizing returns and risk management in a portfolio optimization environment.

Min-Variance

This model seeks the allocation of assets that results in the lowest possible volatility, given an expected level of return [15]. This method relies on the principles of modern portfolio theory, which promotes diversifying investments across various asset categories to minimize risk [16].

In this work, the Min-Variance model is implemented using the PyPortfolioOpt library [17], which offers robust tools for financial portfolio optimization based on financial theories.

Using Min-Variance as a benchmark is crucial because it establishes a performance reference in terms of minimum risk for a given level of return. This allows for evaluating how practical other portfolio optimization approaches are compared to a well-established model. A new approach that achieves superior performance to Min-Variance in metrics such as the Sharpe ratio can be considered more efficient.

Results and Discussion

Three distinct experiments are conducted to analyze the results. In the first experiment, the evaluation focuses exclusively on WIN contracts. In the second experiment, the analysis is dedicated to WDO contracts. Finally, in the third experiment, the dataset is evaluated considering both WIN and WDO contracts. Each experiment aims to explore the effectiveness of the applied strategies in different trading contexts.

In each experiment, the performance of each DRL strategy is assessed using the performance

metrics mentioned earlier. Comparisons are also made with a baseline, represented by the non-optimized portfolio. The choice of the DRL strategy is based on the highest sharpe ratio, as this index evaluates the relationship between accumulated return and the volatility of returns, providing a risk-adjusted measure.

Subsequently, the selected DRL strategy is compared with established benchmarks to contextualize the results achieved. In the experiment, the benchmarks used are IBOVESPA and minimum variance in all comparisons.

Performance Evaluation of DRL Strategies

As mentioned, the five DRL algorithms are trained to find the best parameter configuration using the hyperparameter optimization technique with 50 trials. The tables present the agents with the best configuration obtained for each experiment, showing their cumulative results from June 6, 2018, to November 11, 2019.

Performance of Strategies in the WIN Contract

In Table 1, the annual returns ranged from 17.2% to 19.1%, indicating robust performance in a relatively stable period. DDPG stood out with the highest annual return of 19.1%, while TD3 presented the lowest, with 17.2%. The annual volatility of these strategies is consistently low, ranging between 4.1% and 4.8%, suggesting considerable stability in daily operations, with PPO being the method with the lowest volatility.

At the same time, the baseline maintained a volatility close to the group's average, at 4.2%. Regarding the sharpe ratio, which measures the relationship between return and risk, all strategies presented values above 3.5. The baseline had an excellent performance, with a sharp ratio equal to 3.9, just slightly below SAC, which obtained the highest value of 4.0. These values indicate excellent risk-adjusted efficiency. The maximum drawdown, which indicates the most significant drop in portfolio value before a new high, remained below 1.4% for all strategies and the baseline. PPO showed the lowest maximum drawdown of only 1.3%. This demonstrates the notable resilience of the DRL models against potential market drops. Among the analyzed DRL models, SAC is selected as the most efficient due to its highest Sharpe ratio, which demonstrates its superiority in risk control.

Performance of Strategies in the WIN Contract

Observing Table 2, the annual returns varied significantly among the strategies, with DDPG presenting the highest return of 14.6% and PPO the lowest, at 11%. The annual volatilities of these strategies also showed variations, ranging from 6.7% to 7.6%, with SAC presenting the highest volatility and the baseline matching the lowest observed volatility at 6.7%. This suggests that the baseline managed to maintain stability comparable to that of the more complex strategies. Regarding the sharpe ratio, which measures the risk-adjusted return, the values ranged between 1.6 and 1.9. DDPG led with the highest sharpe ratio, indicating superior efficiency in managing risk relative to the returns obtained. The baseline, with a sharpe ratio of 1.6, offered reasonable efficiency, surpassing PPO and TD3. The analysis of the maximum drawdown

Table 1. Performance comparison between DRL Strategies - WIN Contract.

Metrics	A2C	DDPG	PPO	SAC	TD3	Baseline
Annual Return	17.4%	19.1%	17.7%	18.7%	17.2%	17.7%
Annual Volatility	4.6%	4.8%	4.1%	4.3%	4.3%	4.2%
Sharpe Ratio	3.5	3.7	3.9	4.0	3.7	3.9
Max. Drawdown	1.4%	1.4%	1.3%	1.3%	1.3%	1.3%

Table 2. Performance comparison between DRL Strategies - WDO Contract.

Metrics	A2C	DDPG	PPO	SAC	TD3	Baseline
Annual Return	12.2%	14.6%	11.0%	14.4%	11.3%	11.2%
Annual Volatility	7.1%	7.2%	6.7%	7.6%	6.9%	6.7%
Sharpe Ratio	1.7	1.9	1.6	1.8	1.6	1.6
Max. Drawdown	3.7%	4.1%v	5.0%	4.7%	5.1%	3.7%

shows a maximum loss in portfolio value, ranging between 3.7% and 5.1%, with PPO and TD3 exhibiting the highest drawdown.

DDPG is selected as the most efficient among the evaluated strategies due to its superior sharpe ratio, which indicates an excellent capacity for risk management.

Performance of Strategies in Combined WIN and WDO Contracts

Table 3 shows that the annual returns of the strategies ranged from 13.5% to 17.6%. SAC achieved the highest annual return, while A2C presented the lowest annual return. The baseline obtained a return of 15.7%, surpassing A2C and positioning itself competitively among the other DRL strategies. Regarding annual volatility, the values ranged between 3.8% and 4.4%, with SAC again presenting the lowest value, where lower volatility implies lower risk.

The sharpe ratio of the strategies varied from 2.9 to 4.2, indicating the effectiveness of SAC, which recorded the highest value, in maximizing return per unit of risk. The baseline achieved a ratio of 3.7, showing robust performance, just slightly below PPO and TD3. In terms of maximum drawdown, all strategies, including the baseline, demonstrated significant resilience with a maximum drawdown between 1.2% and 1.4%, indicating their ability to significantly minimize potential losses during the evaluated period.

Comparison with Benchmarks

This section presents the results of applying the DRL methods compared to their respective benchmarks involving the WIN, WDO, and combined contracts.

Comparison in the WIN Contract Experiment

As presented in Table 4, SAC recorded an annual return of 18.7%, positioning itself below IBOVESPA, which had a return of 28%, and Min-Variance, with 22.8%. The annual volatility is considerably lower for SAC and Min-Variance than IBOVESPA, which presented a high volatility of 21.2%. The sharpe ratio is superior for Min-Variance, achieving a value of 5.2, indicative of exceptionally efficient risk management. SAC also showed efficiency with a ratio of 4.0, while IBOVESPA

Table 3. Performance comparison between DRL Strategies - WIN and WDO Contract.

Metrics	A2C	DDPG	PPO	SAC	TD3	Baseline
Annual Return	13.5%	15.5%	16.0%	17.6%	16.9%	15.7%
Annual Volatility	4.3%	4.4%	4.0%	3.8%	4.3%	4.0%
Sharpe Ratio	2.9	3.3	3.7	4.2	3.7	3.7
Max. Drawdown	1.2%	1.4%	1.2%	1.2%	1.3%	1.3%

Table 4. The Best Agent: IBOVESPA and Min-Variance - WIN Contract.

06/06/2018 to 11/11/2019	SAC	IBOV	Min-Variance
Annual Return	18.7%	28%	22.8%
Annual Volatility	4.3%	21.2%	3.9%
Sharpe Ratio	4.0	1.3	5.2
Max. Drawdown	1.3%	11.4%	1.1%

had the lowest value of 1.3, reflecting higher relative risk to the returns generated. The maximum drawdown is considerably lower for SAC and Min-Variance than IBOVESPA, which experienced a significant maximum drawdown of 11.4%. Regarding annual return, expressed as a percentage, IBOVESPA led with 28%, followed by Min-Variance with 22.8% and SAC with 18.7%. As shown in Table 4, these results highlight the differences in investment strategies, where IBOVESPA provides higher total returns but with considerably higher risks.

Comparison in the WDO Contract Experiment

According to the data in Table 5, DDPG presented an annual return of 14.6%, positioning itself among the lowest return values when compared with IBOVESPA, which had a significant return of 28%, and Min-Variance, with returns of 11.3%. The annual volatility of DDPG is 7.2%, demonstrating more excellent stability compared to IBOVESPA and comparable to Min-Variance with 7.3%. Regarding sharpe ratio, DDPG achieved 1.9, superior to IBOVESPA with 1.3, but inferior to Min-Variance with 1.5. The maximum drawdown analysis revealed that DDPG had a drawdown of 4.1%, significantly lower than IBOVESPA and comparable to Min-Variance with 4.5%. This highlights the ability of DDPG and Min-Variance to limit potential losses more effectively than the more volatile market indices. The analysis of the results in Table 5 evidences the efficiency of DDPG, even when compared with IBOVESPA, highlighting its ability to capitalize on market opportunities compared to traditional benchmarks and minimum variance.

Comparison in the Combined WIN and WDO Contracts Experiment

Table 6 shows that SAC achieved an annual return of 17.6%, lower than IBOVESPA, which registered 28%, and slightly below Min-Variance with 21.1%. Despite the lower annual return, SAC demonstrated an extremely low annual volatility of 3.8%, equivalent to Min-Variance and much below IBOVESPA's 21.2%. This low volatility indicates more excellent stability of SAC and Min-Variance compared to the more volatile IBOVESPA. The sharpe ratio of SAC is 4.2, reflecting high efficiency in adjusting return for the risk taken, although Min-Variance presented a still higher ratio of 5.

Table 5. The Best Agent: IBOVESPA and Min-Variance - WDO Contract.

06/06/2018 to 11/11/2019	SAC	IBOV	Min-Variance
Annual Return	14.6%	28%	11.3%
Annual Volatility	7.2%	21.2%	7.3%
Sharpe Ratio	1.9	1.3	1.5
Max. Drawdown	4.1%	11.4%	4.5%

On the other hand, IBOVESPA, with a sharpe ratio of 1.3, showed lower efficiency under the same metric. The maximum drawdown, which measures the most significant drop in portfolio value before a new high, is only 1.2% for SAC and 1.1% for Min-Variance, significantly lower than IBOVESPA's 11.4%. This result emphasizes the robustness of SAC and Min-Variance in terms of risk management and loss limitation.

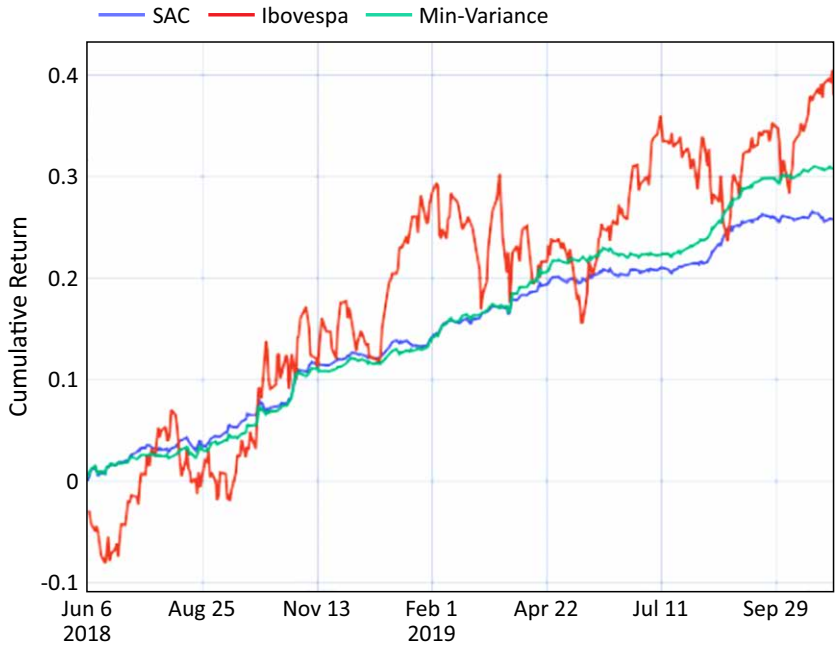
According to Table 6, regarding the final accumulated portfolio value, SAC increased 26.6%, compared to 41.8% of IBOVESPA and 31.1% of Min-Variance. Although IBOVESPA offered a higher total return, it came with considerably higher risks. Notably, the IBOVESPA index shows a significant recovery from mid-2019, surpassing the other strategies in the last quarter of the observed period, highlighting its capacity for recovery after downturns. The graph in Figure 2 shows that the SAC strategy achieved a significant sharpe ratio with low volatility, as desired.

The results illustrate the diversity of performance between different investment strategies, especially in volatile contexts. RL-based strategies show the potential to outperform traditional benchmarks like IBOVESPA in certain periods, although there are significant variations among them in terms of return and risk. Min-Variance, while offering the lowest volatility, also provides the lowest returns, confirming its suitability for investors who prioritize capital preservation over growth.

Table 6. The Best Agent: IBOVESPA and Min-Variance - WDO Contract.

06/06/2018 to 11/11/2019	SAC	IBOV	Min-Variance
Annual Return	17.6%	28%	21.1%
Annual Volatility	3.8%	21.2%	3.8%
Sharpe Ratio	4.2	1.3	5.0
Max. Drawdown	1.2%	11.4%	1.1%
Acc. Portfolio Value	26.6%	41.8%	31.1%

Figure 2. Cumulative results of SAC, Min-Variance, and IBOVESPA.



Conclusion

This study presented an innovative approach to optimizing ATS system portfolios with DRL algorithms, with a specific focus on risk control in highly volatile market environments. The DRL techniques, particularly the DDPG and SAC algorithms, demonstrated a notable ability to learn and adapt trading strategies in real-time, optimizing returns while efficiently managing the associated risks and outperforming the baseline in several aspects.

The results indicate that DRL can significantly surpass traditional trading methods, such as heuristic-based or even other quantitative models that do not incorporate continuous learning and adaptation. The ability to process and react to market conditions in real time, learning from past interactions without explicit predictions, makes DRL-based systems promising tools for modernizing financial trading practices.

This work demonstrates the effectiveness of DRL models in reducing risks and optimizing portfolio performance and points to the potential of applying these techniques in other financial areas, indicating a promising field. Future research could explore integrating RL techniques with other data types, such as macroeconomic signals or sentiment analysis, to develop even more robust and adaptive systems. Therefore, applying advanced RL techniques, such as DRL in finance, represents a promising and innovative direction with substantial implications for the theory and practice of investment management and market operations.

References

1. Chekhlov A, Uryasev S, Zabarankin M. Drawdown measure in portfolio optimization. International Journal of Theoretical and Applied Finance 2005;8(01):13–58.
2. Treleaven P, Galas M, Lalchand V. Algorithmic trading review. Communications of the ACM 2013;56(11):76–85.
3. Buşoniu L, De Bruin T, Tolić D, Kober J, Palunko I. Reinforcement learning for control: Performance, stability, and deep approximators. Annual Reviews in Control 2018;46:8–28.
4. Framework O. Review of business and economics studies. Studies 2013;1(1).
5. Parker K, Fry R. More than half of US households have some investment in the stock market. 2020.
6. Liu X-Y, Yang H, Gao J, Wang CD. Finrl: Deep reinforcement learning framework to automate trading in quantitative finance. In Proceedings of the second ACM international Conference on AI in Finance 2021:1–9.
7. Day Trade Review. Best time frame for day trading - when and how to trade, 2023. [Online]. Available at:
8. Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press, 2018.
9. Sutton RS, McAllester D, Singh S, Mansour Y. Policy gradient methods for reinforcement learning with function approximation. Advances in Neural Information Processing Systems 1999;12.
10. Liu X-Y, Yang H, Chen Q, Zhang R, Yang L, Xiao B, Wang CD. Finrl: A deep reinforcement learning library for automated stock trading in quantitative finance. arXiv preprint arXiv 2020:2011.09607.
11. Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. arXiv preprint arXiv 2018:1801.01290.
12. Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods. In International Conference on Machine Learning 2018:1587–1596.
13. Buehler H, Gonon L, Teichmann J, Wood B. Deep hedging. Quantitative Finance 2019;19(8):1271–1291.
14. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2019.
15. Yang H, Liu X-Y, Wu Q. A practical machine learning approach for dynamic stock recommendation, In 2018 17th IEEE international conference on trust, security and privacy in computing and communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE) 2018:1693–1697.
16. Markowitz H. Portfolio Selection. The Journal of Finance 1952;7(1):77-91.
17. Martin RA. Pyportfolioopt: Portfolio optimization in Python. Journal of Open Source Software 2021;6(61):3066.