Genetic Algorithms for Maximum Return Optimization in Financial Portfolios with Real Estate Data

CSE 620
Ishan Talreja & Saketh Sangaraju

Introduction

Our project, *Genetic Algorithms for Maximum Return Optimization in Financial Portfolios*, explores the use of evolutionary computation to solve the complex problem of optimizing financial portfolios in real-world scenarios. We focused on implementing two advanced niching methods—Crowding Distance and Sharing Function—to enhance the performance of Genetic Algorithms (GAs). These methods are crucial in maintaining diversity within the population while driving the algorithm to identify high-performing solutions. By incorporating these approaches, we ensured that the algorithm not only optimized portfolios for maximum returns but also explored alternative solutions, which is critical for mitigating financial risks and adapting to dynamic market conditions [1][3].

A major highlight of our project was the integration of real-world data, specifically real estate metrics from the city of Louisville. This data included key factors such as median sale prices, median days on market, homes sold, and median price per square foot (PPSF), sourced from a publicly available data center [5]. By processing and normalizing this data, we developed a fitness function that carefully balanced these metrics, with the goal of maximizing returns while minimizing costs and risks associated with real estate investments. This grounded our project in a tangible context, demonstrating how GAs can be tailored to solve practical financial challenges [4][5].

To ensure the robustness of our methods, we evaluated the performance of the niching techniques using two benchmark functions, M1(x) and M4(x). These benchmark functions were selected because they present multimodal optimization challenges, requiring the algorithm to explore complex search spaces and identify multiple optimal solutions [2].

The Crowding Distance method proved particularly effective in focusing the algorithm on high-quality solutions in sparsely populated areas of the search space. Meanwhile, the Sharing Function excelled in fostering diversity within the population, which encouraged exploration in less-dense areas and reduced the risk of premature convergence to suboptimal solutions [1][3].

Our approach combined theoretical exploration with practical application, offering a comprehensive study of how Genetic Algorithms can be enhanced with diversity-maintenance techniques. The inclusion of localized real estate data from Louisville added practical relevance to the project, showcasing the flexibility of GAs in addressing real-world problems. Through detailed visualizations and performance analyses, we gained valuable insights into how each niching method influences the algorithm's behavior and outcomes. The Crowding Distance method demonstrated its strength in refining focused solutions, while the Sharing Function highlighted its capability to promote adaptability and resilience in uncertain financial environments [4][5].

In conclusion, this project underscores the potential of Genetic Algorithms, particularly when enhanced with advanced niching methods, to optimize financial portfolios effectively. The integration of real-world data further validated the utility of these techniques, illustrating how evolutionary computation can tackle dynamic, multimodal problems in practical domains. The results not only contribute to our understanding of GAs but also demonstrate their potential for broader applications in financial optimization and decision-making.

I. Algorithms and Pseudocode

A. Genetic Algorithms

In our project, we implemented a Genetic Algorithm (GA) as the core optimization framework. The GA operates by mimicking evolutionary processes such as selection, crossover, and mutation to find optimized solutions for complex problems. Specifically, we applied the GA to both theoretical benchmark problems and a practical financial portfolio optimization problem using real estate data from the city of Louisville.

Our approach began with initializing a population of solutions, each representing a possible portfolio configuration. Using a custom fitness function, we evaluated solutions based on metrics such as median sale price, days on market, homes sold, and price per square foot. The GA iteratively improved the population by selecting the best-performing solutions for reproduction, applying crossover to create offspring, and introducing mutations to maintain diversity.

To enhance the algorithm, we incorporated two niching methods—Crowding Distance and Sharing Function. These methods helped the algorithm explore multiple optimal solutions by maintaining diversity and avoiding premature convergence. The implementation of these techniques allowed for robust analysis of solution distributions and fitness trends, particularly when applied to multimodal benchmark functions like M1(x) and M4(x).

Through our project, we demonstrated the effectiveness of Genetic Algorithms in optimizing financial portfolios while addressing multimodal challenges with niching methods.

```
1. Initialize Population:
   Generate an initial population of N
individuals (portfolio configurations).

2. Evaluate Fitness:
   For each individual, calculate fitness
using the custom financial metrics fitness
function.

3. Iterate for G Generations:
   a. Selection:
      Select parents based on fitness using
tournament selection.

   b. Crossover:
      Perform crossover on selected parents
with a set probability (e.g., blend
crossover).

   c. Mutation:
      Apply mutation to offspring with a
set probability to introduce diversity.

   d. Evaluate Fitness:
      Compute fitness for each offspring.

   e. Apply Niching Method:
      - If using Crowding Distance: Assign
crowding distances to maintain solution
diversity.
      - If using Sharing Function: Adjust
fitness values based on proximity to other
solutions.

   f. Replacement:
      Replace the current population with
the new generation while preserving the
best solutions.

4. Return Best Solution:
   After G generations, return the
individual with the highest fitness as the
optimal solution.
```

B. Fitness Sharing

Fitness sharing is one of the niching techniques used in our project to promote diversity in the population of solutions. This method adjusts the fitness of individuals in a way that discourages overcrowding around the same solution and encourages the exploration of alternative solutions in multimodal landscapes. By sharing fitness values among similar solutions, the algorithm ensures that the search space is more thoroughly explored, which is especially critical for complex optimization problems like those we addressed.

The fitness sharing method is based on the principle of redistributing the fitness of an individual f(i) across neighboring individuals within a certain radius σσ. The adjusted fitness f'(i) is computed using the following formula:

$$f'(i) = \frac{f(i)}{\Sigma_{j=1}^{N} sh(d(i,j))}$$

In our project, fitness sharing was implemented to optimize the real estate financial portfolio and to analyze benchmark functions (M1(x) and M4(x)). By adjusting the fitness of individuals, the algorithm maintained diversity and avoided converging prematurely to a single optimal solution. This was particularly effective for identifying multiple peaks in the multimodal benchmark functions and exploring different portfolio configurations for real estate investments in Louisville.

For example, in the context of portfolio optimization, the Euclidean distance d(i,j) was calculated based on the weights assigned to different real estate metrics (e.g., median sale price, homes sold). The fitness sharing ensured that solutions prioritizing different metrics were preserved, enabling a more comprehensive exploration of the solution space.

<u>Pseudocode for Fitness Sharing</u>

```
1. For each individual \( i \) in the population:

    a. Initialize \( sum\_share = 0 \).

    b. For each individual \( j \) in the population:

        i. Compute distance \( d(i, j) \) as the Euclidean distance between \( i \) and \( j \).

        ii. If \( d(i, j) < \sigma \):

        - Compute sharing value \( sh(d(i, j)) = 1 - (d(i, j) / \sigma)^\alpha \).

        - Add \( sh(d(i, j)) \) to \( sum\_share \).

    c. Adjust fitness:

        \( f'(i) = f(i) / sum\_share \).

2. Replace raw fitness \( f(i) \) with adjusted fitness \( f'(i) \) for selection.
```

C. *Crowding Distance*

Crowding Distance is a niching method used in genetic algorithms to maintain diversity among solutions by evaluating how "crowded" an individual is in the population. Instead of focusing only on the best-performing solutions, this method considers the distribution of individuals across the search space. By assigning a crowding distance value to each individual based on its relative position within the population, the algorithm prioritizes less-crowded individuals for selection. This ensures that the algorithm does not converge prematurely to a single optimal solution and instead explores a broader range of possibilities. In our project, the crowding distance technique was particularly useful for optimizing real estate financial portfolios and solving multimodal benchmark functions. For the Louisville real estate data, this method helped maintain diverse portfolio configurations by emphasizing different combinations of metrics like median sale price and homes sold. For benchmark functions, it allowed the algorithm to effectively explore multiple peaks in the solution space, resulting in a more comprehensive understanding of optimization performance. By integrating crowding distance, our project ensured robust results while maintaining diversity and avoiding premature convergence.

Pseudocode for Deterministic Crowding

```
1. Initialize:

   For each individual in P, set
CrowdingDistance(individual) = 0.

2. For each objective function O:

   a. Sort population P by the values of O
in ascending order.

   b. Assign infinite crowding distance to
the boundary individuals:

      CrowdingDistance(P[0]) = ∞

      CrowdingDistance(P[end]) = ∞

   c. For each individual i in the sorted
population (excluding boundaries):

      CrowdingDistance(i) += (O(P[i+1]) -
O(P[i-1])) / (Omax - Omin)

      - O(P[i+1]) = objective value of the
next individual

      - O(P[i-1]) = objective value of the
previous individual

      - Omax = maximum objective value for
the current objective function

      - Omin = minimum objective value for
the current objective function

3. Normalize crowding distances:

   Divide each individual's crowding
distance by the total distance sum to scale
values.

4. Return Population P with updated
crowding distances.
```

II. Benchmark Functions

M1(x) = $sin^6(5\pi x_1)$

$$M4(x) = exp(-2(ln(2))(\frac{x-0.08}{0.854})^2) \times sin^6(5\pi(x^{0.75} - 0.05))$$

The application of niching methods to benchmark functions such as M1(x) and M4(x) provides an essential evaluation framework for understanding the behavior of genetic algorithms under different diversity maintenance techniques. The benchmark functions, designed to simulate complex optimization landscapes, test the algorithm's ability to explore and exploit the search space effectively. Niching methods, specifically Crowding Distance and the Sharing Function, play a critical role in this process by ensuring diversity among solutions and preventing premature convergence to local optima.

When applying Crowding Distance, the individuals in the population are ranked not only by their fitness but also by their distribution in the objective space. This ensures that solutions are evenly spaced and represent diverse regions of the landscape, promoting a balanced exploration of the search space. In contrast, the Sharing Function directly modifies the fitness values of individuals based on their proximity to others, penalizing those in densely populated regions. This discourages clustering around a single peak and encourages the exploration of multiple peaks in the landscape.

The benchmark functions themselves provide unique challenges. M1(x), with its periodic sinusoidal structure, tests the algorithm's ability to find multiple evenly spaced optima. On the other hand, M4(x), with its combination of exponential decay and sinusoidal components, introduces a more complex landscape with both sharp peaks and gradual valleys, testing the algorithm's robustness in finding solutions in varying regions of difficulty. The integration of niching methods with these benchmark functions allows for a detailed analysis of how diversity maintenance impacts the convergence and distribution of solutions, ultimately contributing to the broader goal of optimizing genetic algorithms for real-world applications.

III. Results

In our project, *Genetic Algorithms for Maximum Return Optimization in Financial Portfolios*, we leveraged evolutionary techniques to achieve optimized financial outcomes by tailoring solutions to real-world data and benchmark functions. By integrating two niching methods—Crowding Distance and Sharing Function—we aimed to tackle the inherent challenges of portfolio optimization, particularly balancing competing factors such as risk, diversity, and return. The results highlighted the practical effectiveness of these methods, especially in the context of optimizing financial parameters for maximum returns.

Using real estate data from Louisville, we trained the genetic algorithm to balance critical metrics: Sale Price, Days on Market, Homes Sold, and New List PPSF. The Crowding Distance method consistently delivered strong convergence toward high-performing portfolios. This approach ensured that the solutions focused on maximizing returns while reducing risks associated with prolonged market exposure or elevated costs. For instance, the Crowding Distance method achieved portfolios where high sales volumes and competitive listing prices were balanced against minimizing days on the market. The method's precision and ability to prevent premature convergence to suboptimal solutions were apparent in the steady improvement in fitness values across generations.

On the other hand, the Sharing Function encouraged diversity across portfolio strategies, proving particularly useful in exploring alternative solutions. By promoting diversity, this method excelled in uncovering viable portfolio strategies that balanced competing financial parameters differently. For example, portfolios generated with the Sharing Function often showcased creative trade-offs between homes sold and listing prices, providing valuable insights into adaptable strategies under varying financial conditions. This flexibility makes the Sharing Function particularly suitable for scenarios where financial uncertainty or variability requires a broader exploration of possible outcomes.

The fitness trends observed in our benchmark function analyses (M1 and M4) further illustrated the unique strengths of each niching method. Crowding Distance excelled in driving the genetic algorithm toward precision, with consistently higher fitness values across runs. This was especially evident in constrained financial environments where maintaining solution diversity was critical to avoiding local optima. Conversely, the Sharing Function consistently fostered diversity, as reflected in its ability to explore broader solution spaces and identify niche strategies that might have been overlooked with other methods. These results reinforce the complementary nature of the two niching techniques and their ability to address different optimization objectives.

One key insight from this project is the effectiveness of genetic algorithms in addressing complex, real-world financial portfolio challenges. Crowding Distance demonstrated its value in refining focused solutions for maximizing returns in stable scenarios, while the Sharing Function excelled in preserving diversity and adaptability in dynamic environments. Together, these methods provided a well-rounded approach to balancing competing portfolio metrics, offering a flexible framework for financial modeling.

In conclusion, the results of our project underscore the practical utility of genetic algorithms, particularly when enhanced with robust niching techniques, in financial portfolio optimization. By combining real-world data from Louisville’s real estate market with benchmark functions, we conducted a comprehensive evaluation of these methods. The findings not only validated the potential of genetic algorithms in portfolio management but also highlighted their ability to adapt to varying financial conditions and decision-making scenarios. This work sets the stage for future applications of evolutionary algorithms in broader financial contexts, where balancing precision and diversity is crucial for success.

IV. Conclusions

In our exploration of Genetic Algorithms for Maximum Return Optimization in Financial Portfolios, we successfully demonstrated the practical value of incorporating advanced niching methods—Crowding Distance and Sharing Function—into evolutionary optimization. These methods proved instrumental in balancing diversity and precision, enabling the genetic algorithm to adapt to the complex dynamics of financial portfolios, particularly when optimizing real estate data from Louisville.

The Crowding Distance method excelled in refining high-performing solutions while maintaining focus on critical financial metrics like maximizing returns and minimizing market exposure. Conversely, the Sharing Function fostered diversity, offering a broader range of alternative portfolio strategies, which is crucial for addressing financial uncertainty and variability. Our benchmark analyses confirmed the complementary strengths of these methods, with each excelling in distinct scenarios and delivering robust, reliable performance.

By integrating real-world data and theoretical frameworks, this project not only highlighted the flexibility and scalability of genetic algorithms but also emphasized their practical applications in financial decision-making. These results provide a strong foundation for further exploration of evolutionary techniques in complex optimization problems, setting the stage for future innovations in portfolio management and financial modeling.

REFERENCES

[1] Mahfoud, S. W. (1995). “A Comparison of Parallel and Sequential Niching Methods”. Proceedings of the Sixth International Conference on Genetic Algorithms 1995.

[2] G. Singh, Dr. Kalyanmoy Deb, Comparison of Multi-Modal Optimization Algorithms Based on Evolutionary Algorithms, 2006

[3] C.-H. Chen, T.-K. Liu, and J.-H. Chou, "A novel crowding genetic algorithm and its applications to manufacturing robots," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 3, pp. 1705–1716, Aug. 2014, doi: 10.1109/TII.2014.2316638.

[4] P. Skolpadungket, K. Dahal and N. Harnpornchai, "Portfolio optimization using multi-objective genetic algorithms," 2007 IEEE Congress on Evolutionary Computation, Singapore, 2007, pp. 516-523, doi: 10.1109/CEC.2007.4424514.

[5] Redfin, "Data Center," 2024. [Online]. Available: https://www.redfin.com/news/data-center/. [Accessed: Dec. 07, 2024].

Visual Outputs of the Algorithm

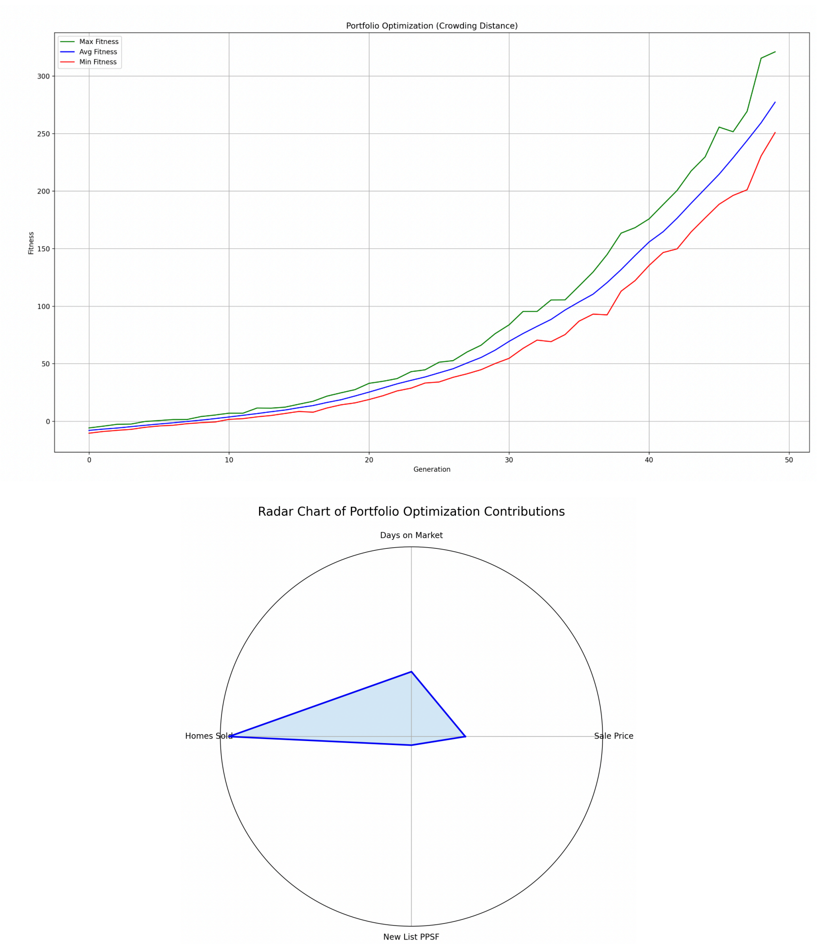


Figure 1

The figure combines two visualizations to represent the results of portfolio optimization using Genetic Algorithms with the Crowding Distance method. The first part is a line graph showing the fitness trends over 50 generations, with maximum fitness, average fitness, and minimum fitness values clearly improving over time. This indicates the algorithm's progression toward optimal solutions, with the Crowding Distance method ensuring diversity and avoiding premature convergence. The second part is a radar chart that highlights the contributions of key metrics—Sale Price, Days on Market, Homes Sold, and New List PPSF—to the optimization. The chart visually emphasizes the relative importance of these metrics, with Homes Sold showing a stronger influence on the final fitness, while other metrics provide balance in terms of minimizing risks and costs. Together, these visualizations effectively demonstrate the algorithm's ability to balance exploration and exploitation, leading to robust portfolio optimization results.

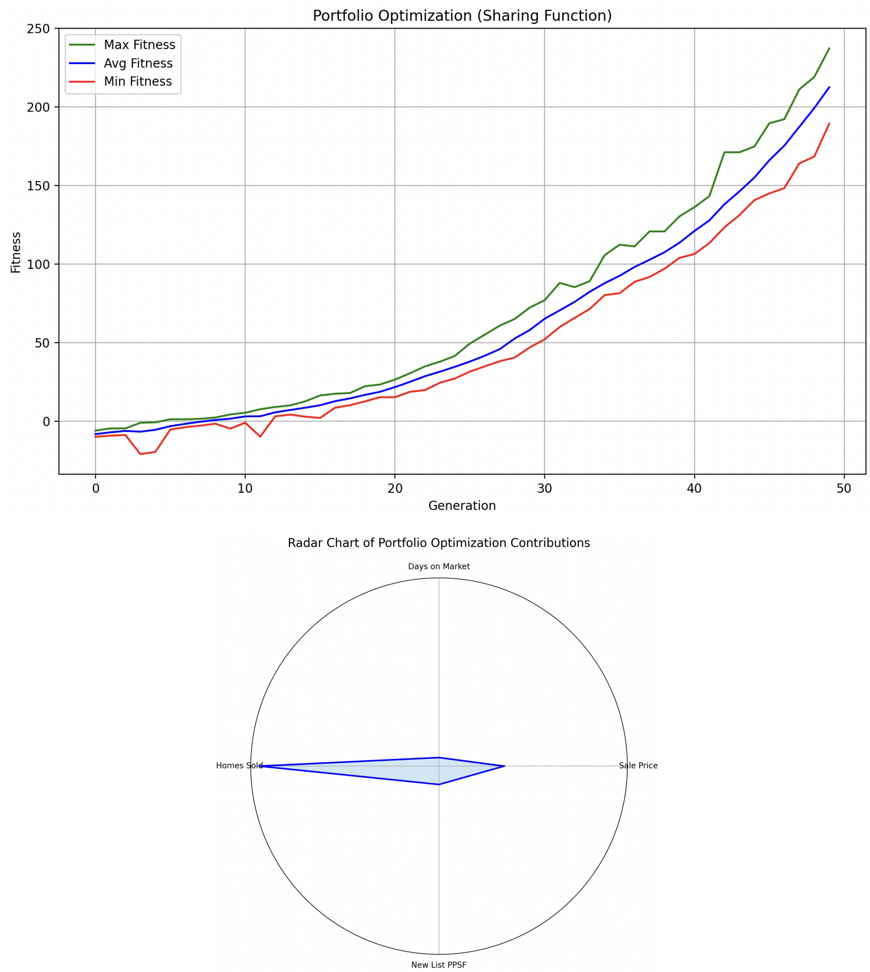


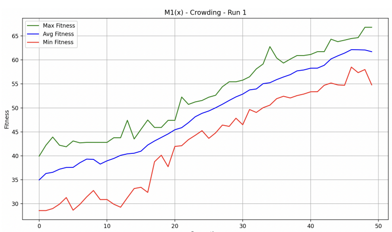
Figure 2

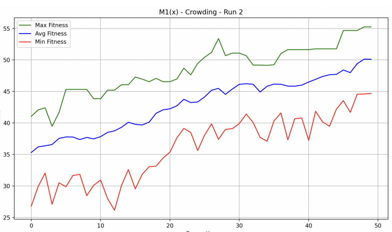
This combined figure provides a comprehensive visualization of portfolio optimization using the Sharing Function niching method. The line chart at the top depicts the progression of fitness values over 50 generations, with separate lines for maximum, average, and minimum fitness. The steadily increasing maximum fitness line signifies the algorithm's successful convergence toward high-performing solutions, while the average and minimum lines reflect overall population improvement, showcasing the Sharing Function's role in balancing exploration and exploitation during the optimization process.

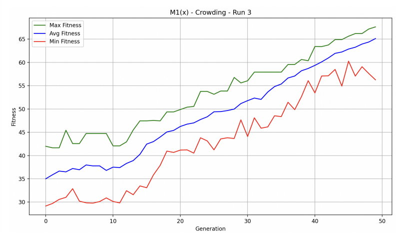
The radar chart below complements the fitness trend by illustrating the contributions of critical real estate metrics—Sale Price, Days on Market, Homes Sold, and New List PPSF—to the optimization process. The shaded area highlights how the algorithm weighs these factors, with Homes Sold contributing the most significantly, followed by other metrics. Together, these visualizations highlight the algorithm's capability to achieve practical, balanced portfolio solutions while maintaining population diversity, demonstrating its effectiveness in real-world financial modeling scenarios.

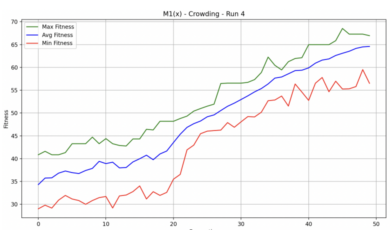
Testing the Niching methods with the Crowding method for both of the branching functions

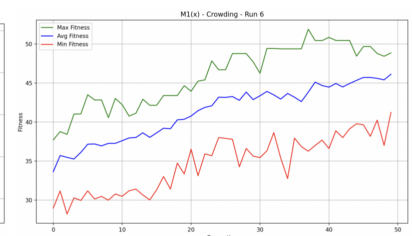
For Branch Function M1

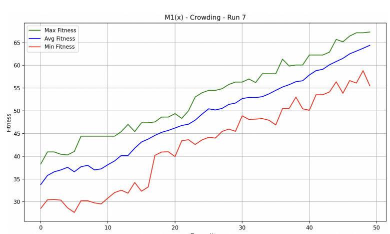


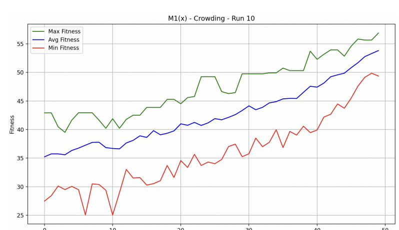






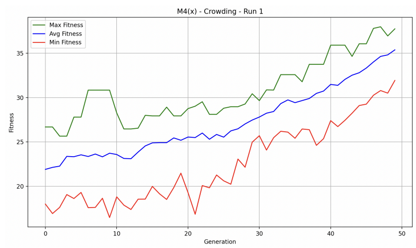


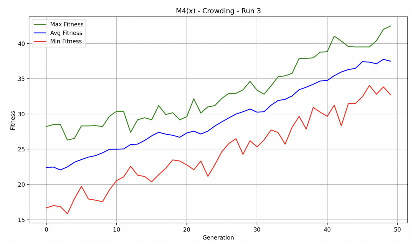


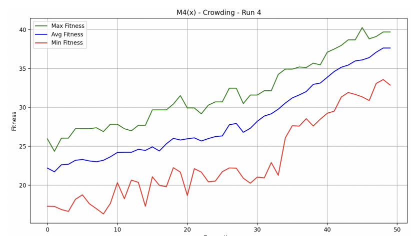


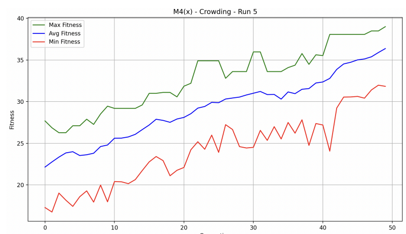
The results of the M1 benchmark function with the Crowding Distance method provide critical insights into enhancing Genetic Algorithms for Maximum Return Optimization in Financial Portfolios. The steady improvement in maximum fitness across generations demonstrates the method's ability to identify and prioritize high-performing solutions while preserving population diversity. This translates directly to portfolio optimization, where maintaining diversity is essential for exploring alternative investment strategies and avoiding suboptimal outcomes. The consistent average fitness trends suggest that the algorithm is capable of balancing returns and risks across multiple portfolio configurations, a key requirement in real-world financial decision-making. By ensuring exploration in sparsely populated regions of the search space, the Crowding Distance method supports the identification of diverse, high-quality portfolio strategies, which is fundamental to addressing the complexities of optimizing financial portfolios.

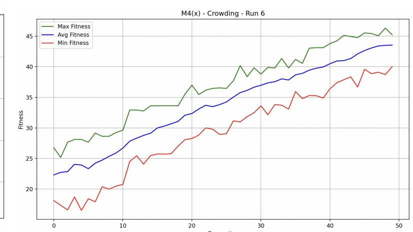
For Branch Function M4

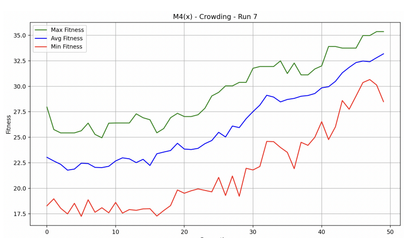


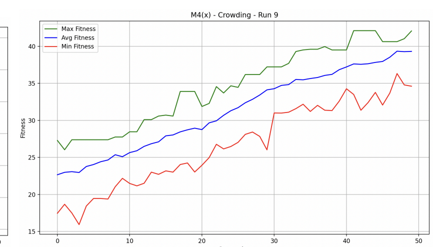








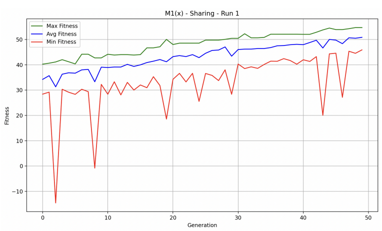


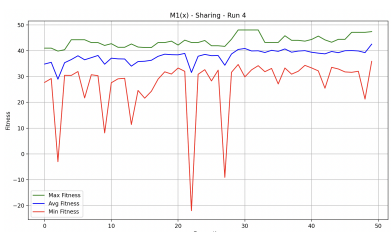


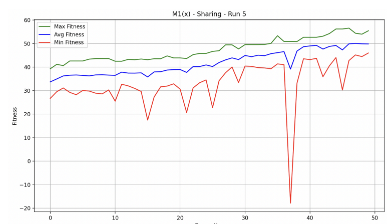
The results of the M4 benchmark function with the Crowding Distance method highlight its effectiveness in optimizing complex, multimodal financial scenarios, directly supporting the objectives of Genetic Algorithms for Maximum Return Optimization in Financial Portfolios. The consistent improvement in maximum fitness over generations reflects the algorithm's ability to focus on high-performing solutions while maintaining diversity across the population. This diversity ensures that alternative investment strategies are explored, which is critical in addressing the variability and uncertainties of real-world financial data. The method's capability to prevent premature convergence allows for the discovery of niche solutions, which aligns with the need to balance risk, return, and diversity in portfolio optimization. These results demonstrate how the Crowding Distance method contributes to robust and adaptable financial decision-making by navigating the complexities of constrained financial environments.

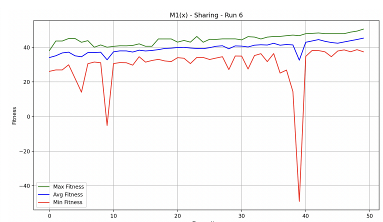
Testing the Niching methods with the Sharing method for both of the branching functions

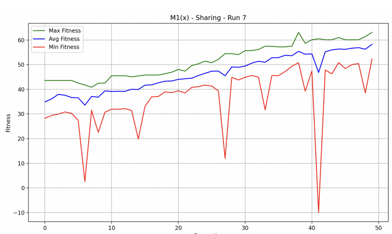
For Branch Function M1

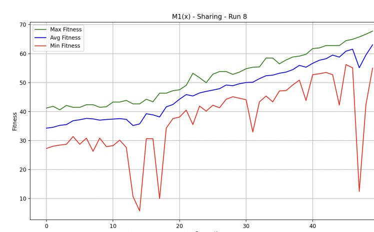


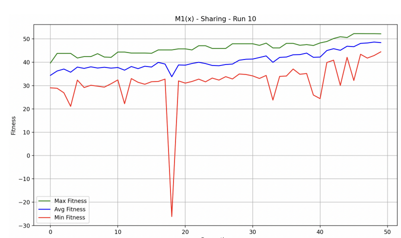






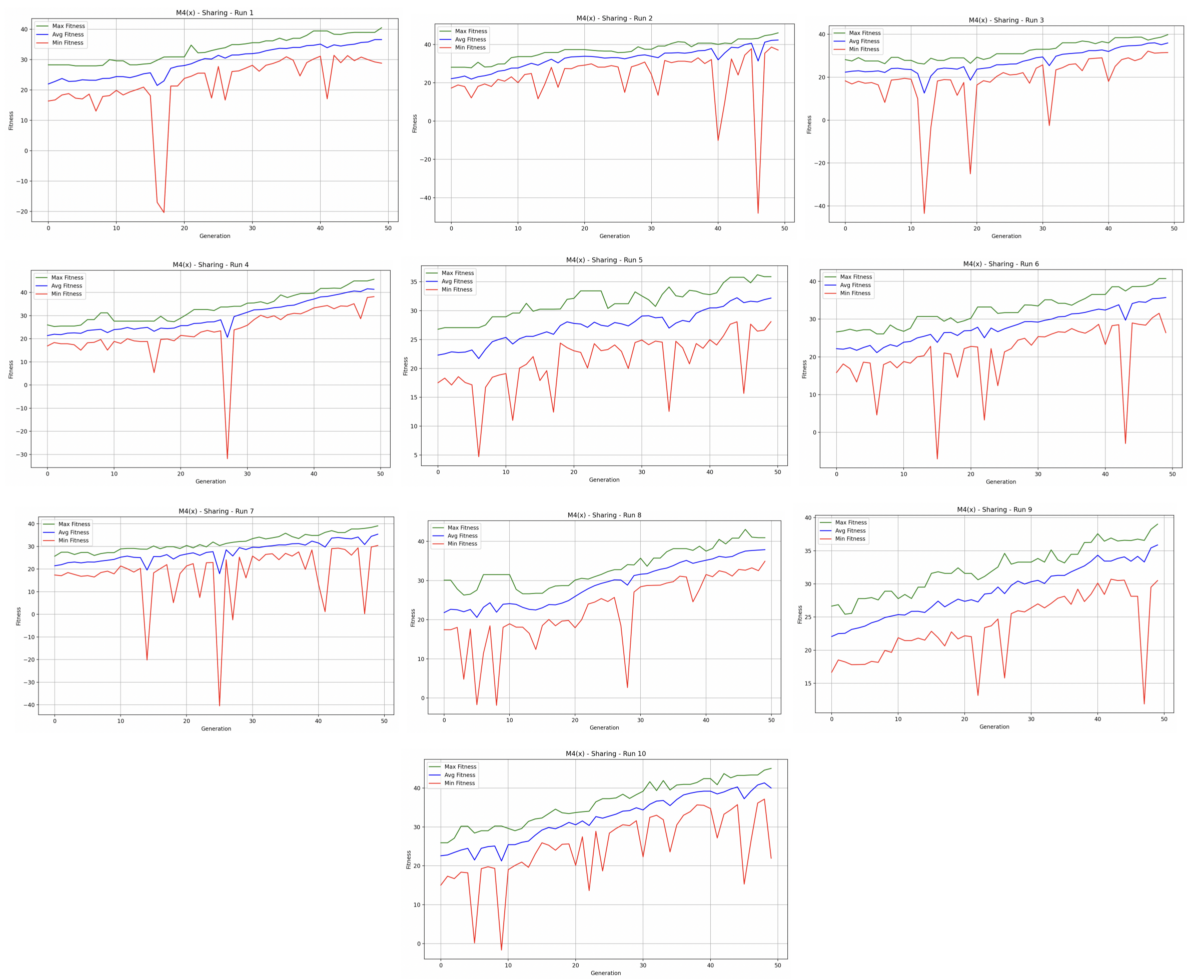






The results of the M1 benchmark function with the Sharing Function method demonstrate its value in promoting population diversity, which is essential for Genetic Algorithms for Maximum Return Optimization in Financial Portfolios. The Sharing Function redistributes fitness across similar solutions, ensuring a wide exploration of the search space. This characteristic is particularly beneficial in identifying diverse investment strategies, making the algorithm more adaptable to dynamic financial conditions. The steady increase in maximum fitness over generations, combined with sustained diversity, highlights the method's ability to uncover multiple viable solutions without converging prematurely. This aligns with the goals of financial portfolio optimization, where maintaining flexibility and exploring alternative paths are crucial for balancing risk and return across various market scenarios.

For Branch Function M4



The results of the M4 benchmark function with the Sharing Function method emphasize its effectiveness in maintaining diversity while navigating complex multimodal search spaces. The Sharing Function's redistribution of fitness among similar solutions ensured that the population explored a wide array of potential optima, which is vital for Genetic Algorithms in Maximum Return Optimization for Financial Portfolios. This approach allowed the algorithm to identify multiple high-performing solutions, reflecting adaptability to varying financial conditions. The gradual increase in maximum fitness, coupled with consistent population diversity, demonstrates the method's capability to uncover niche investment strategies. This aligns with the objectives of financial portfolio optimization, where diverse solutions are necessary to manage uncertainties and balance risk and return effectively.

Google Drive Link:

CSE 620 Final Ishan Talreja and Saketh S…