An Evaluation of Backtracking Steepest Descent and Backtracking Newton's Method for Solving Numerical Optimization Problems

Ahmad Sidani
Politecnico di Torino
S312919
ahmad.sidani@studenti.polto.it

Ali Yassine
Politecnico di Torino
S312920
ali.yassine@studenti.polto.it

Hadi Ibrahim
Politecnico di Torino
S313385
hadi.ibrahim@studenti.polto.it

***Abstract*—This report focuses on numerical optimization methods used to solve unconstrained problems. Two optimization techniques are assessed and contrasted. The results give useful insights into the optimization approaches' strengths and limitations.**

Code found in: This Github Repository

I. Introduction

Numerous fields, including engineering, finance, and data science, rely heavily on numerical optimization. Numerical optimization can be done using a variety of methods, each of which has advantages and disadvantages. This study seeks to examine the efficiency and efficacy of two numerical optimization techniques: the steepest descent method [1], and the Newton method [2].

A backtracking line search algorithm is applied to all two methods to determine the learning rate. The test problems in the study include the Rosenbrock function, the chained Rosenbrock function [3], Generalized Brown [3] and Problem 76 [5]. Insights into the numerical optimization of unconstrained problems are provided by comparing and evaluating the results of each optimization approach for each test case.

By providing a complete analysis of the optimization processes and the results of the two optimization techniques, this research seeks to assist in the development of more effective optimization algorithms.

II. Methods

To tackle the problem at hand, two optimization methods were used: steepest descent, and Newton. Backtracking line search, a standard strategy for assuring sufficient decrease of the objective function during optimization, was applied in all two techniques.

A. Steepest descent method

A gradient-based optimization process called the Steepest Descent method updates the solution by moving in the direction of the objective function's negative gradient. Because it only makes use of the gradient, it is regarded as a first-order optimization technique. Although the technique is simple to use and has few parameters that need to be modified, it can take longer for it to converge for highly nonlinear problems.

B. Newton method

The Newton method is a second-order optimization algorithm that determines the search direction by using both the gradient and the Hessian matrix of the objective function. Since the Hessian matrix provides second-order information about the objective function, the Newton approach converges significantly quicker than first-order methods such as the Steepest Descent method. Calculating the correct Hessian matrix, on the other hand, can be costly, especially for large-scale problems.

Backtracking Line Search is technique for determining the ideal step size in an optimization technique. It is applied to make sure the objective function is sufficiently reduced on each iteration. Backtracking line search's fundamental principle is to begin with an initial guess for the step size and then gradually reduce it until an enough reduction in the target function is achieved. The Armijo condition is a popular option for the stopping criterion used by the backtracking line search algorithm. According to the Armijo condition, an appropriate step size is one in which the objective function reduces by a sufficient amount in the direction of the negative gradient.

Backtracking Line Search is frequently used in various optimization algorithms, including the two techniques previously stated.

III. Test Problems

A. Rosenbrock function

The Rosenbrock function, often known as the "banana function," is a test function commonly implemented in optimization and numerical analysis. It serves as a performance test problem for algorithms for optimization. Due to the fact that the function has a number of local minima and is non-convex. The function is defined as:

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

At the position (1,1), Rosenbrock has a global minimum. Two starting points are provided: x0 = (1.2, 1.2) and x0 = (1.2,

-1.2). An optimization method can use these starting points as an initial guess to determine the function's minimum.

B. Chained Rosenbrock function

The Chained Rosenbrock function [3] is a multidimensional optimization problem that is commonly used to evaluate optimization algorithms. It is an improvement on the classic Rosenbrock function and is meant to have several local minima and small valleys, making optimization difficult. Chained Rosenbrock is defined as follows:

$$F(X) = \sum_{i=2}^{n} \left[100\left(x_{i-1}^2 - x_i\right)^2 + \left(x_{i-1} - 1\right)^2\right]$$

The best initial points $\bar{x}_i$ are defined as:

$$\bar{x}_i = \begin{cases} -1.2, & \mathrm{mod}(i,2) = 1 \\ 1.0, & \mathrm{mod}(i,2) = 0 \end{cases}$$

C. Generalized Brown function

Generalized Brown function [3] is a test function that is frequently used in the discipline of numerical optimization to assess and compare optimization techniques. This function can be used to model actual optimization issues because it is multi-dimensional, non-linear, and non-convex. The Generalized Brown function is defined as:

$$F(x) = \sum_{j=1}^{k} (x_{i-1}^2)^{(x_i^2+1)} + (x_i^2)^{(x_{i-1}^2+1)}$$

$$\text{i} = 2j, k = n/2$$

The starting point for the optimization is given by:

$$\bar{x}_i = \begin{cases} -1, & \mathrm{mod}(i,2) = 1 \\ 1, & \mathrm{mod}(i,2) = 0 \end{cases}$$

The function is a useful tool for testing optimization methods since it contains several global minima and is well known for its capacity to replicate the behavior of a wide variety of optimization problems.

D. Problem 76

This function [4], which is a sum of squared differences, is frequently used in numerical analysis and optimization. It is defined as:

$$F(x) = \frac{1}{2} \sum_{k=1}^{n} [f_k(x)]^2$$

where

$$f_k(x) = \begin{cases} x_k - \frac{x_{k+1}^2}{10} & 1 \le k < n \\ x_k - \frac{x_1^2}{10} & k = n \end{cases}$$

The starting point for the optimization is given by:

$$\bar{x}_l = 2, \text{for all } l \ge 1$$

The minimum of the function may be found iteratively by using $\bar{x}_l$ as an initial guess in an optimization method.

IV. Results

A. Rosenbrock function

In this study, two evaluations were carried out with initial points of (1.2,1.2) and (-1.2,1.0), respectively.

1) Results obtained using the steepest descent method: The initial evaluation took place over 1000 iterations, lasted 0.56 seconds, and gradient norm 30.98. This analysis produced an optimal point with a coordinates (1.08, 1.24). The second review was place over 1000 iterations, took around 0.6 seconds, and gradient norm 43.89. This analysis produced an optimal point with a coordinates (-0.98, 1.08).

2) Results obtained using the Newton method: The initial evaluation took 8 iterations, lasted 0.001 seconds, and gradient norm equal to 0. This analysis produced an optimal point with a coordinates (1, 1). The second evaluation took it lasted for 21 iterations, lasted 0.04 seconds, and gradient norm equal to 0. This analysis produced an optimal point with a coordinates (1, 1).

3) Comparison of the results obtained using each optimization technique: We compare the elapsed time, number of iterations, and gradient norm for the initial starting point to compare the effectiveness of Steepest Descent and Newton's method in optimizing the Rosenbrock function. To offer insights, we give the findings in Table I.

	Iterations	Time	Gradient Norm
Steepest Descent	1000	0.56s	30.98
Newton	8	0.001s	0

TABLE I
COMPARISON OF TWO METHODS ON THE ROSENBROCK FUNCTION USING THE FIRST STARTING POINT

Figure 1 shows the convergence of the steepest descent and the Newton method for the Rosenbrock function on the initial point towards the optimal solution and highlights the performance differences between them.

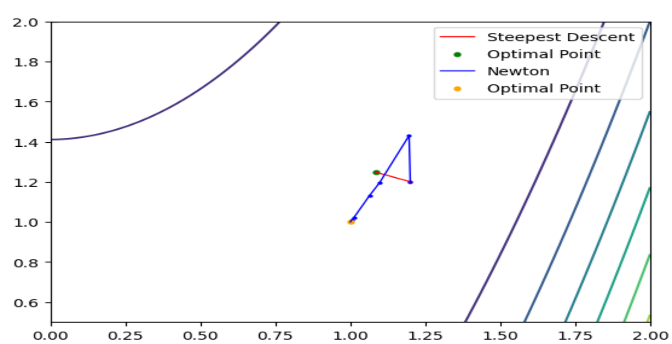

Fig. 1. Behavior of both methods for the Rosenbrock problem on the first starting point

By examining the same metrics for the second starting point indicated in the subsection, we also assess how well Steepest Descent and Newton's approach perform in terms of optimizing the Rosenbrock function. We give the results in Table II for a clear comparison.

	Iterations	Time	Gradient Norm
Steepest Descent	1000	0.6	43.89
Newton	24	0.004	0

TABLE II
COMPARISON OF TWO METHODS ON THE ROSENBROCK FUNCTION USING THE SECOND STARTING POINT

Figure 2 shows the convergence of the steepest descent and the Newton method for the Rosenbrock function on the second point towards the optimal solution and highlights the performance differences between them.

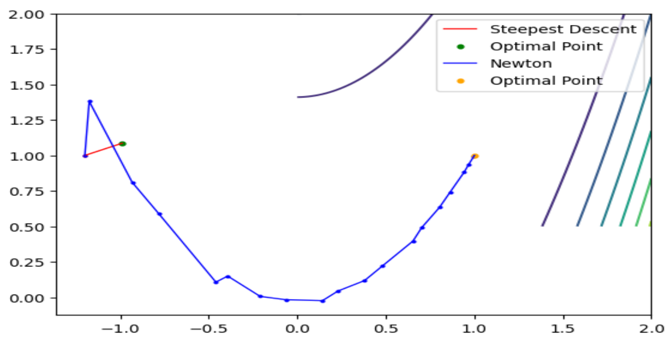

Fig. 2. Behavior of both methods for the Rosenbrock function on the second starting point

B. Chained Rosenbrock function

In this analysis, two evaluations were conducted, the first using initial points provided by the function as $\bar{x}_i$, and the second using random points between 0 and 1. Both utilized a dimension of 10^3.

1) Results obtained using the Steepest Descent method: The initial evaluation took place over 1000 iterations, lasted around 54 seconds, and gradient norm 12536. The second evaluation took 1000 iterations, lasted around 55 seconds, and gradient norm equal to 1957.

2) Results obtained using the Newton method: The initial evaluation took place over 1000 iterations, lasted around 97 seconds, and gradient norm 7152. The second evaluation took 1000, lasted around 55 seconds, and gradient norm equal to 1918.

3) Comparison of the results obtained using each optimization technique: We compare the elapsed time, number of iterations, and gradient norm for the initial starting point to compare the effectiveness of Steepest Descent and Newton's method in optimizing the Chained Rosenbrock function. To offer insights, we give the findings in Table III.

	Iterations	Time	Gradient Norm
Steepest Descent	1000	54s	12536
Newton	1000	97s	7152

TABLE III
COMPARISON OF TWO METHODS ON THE CHAINED ROSENBROCK FUNCTION USING THE FIRST STARTING POINT

Figure 3 shows the convergence of the steepest descent and the Newton method for the Chained Rosenbrock function on the first point towards the optimal solution and highlights the performance differences between them.

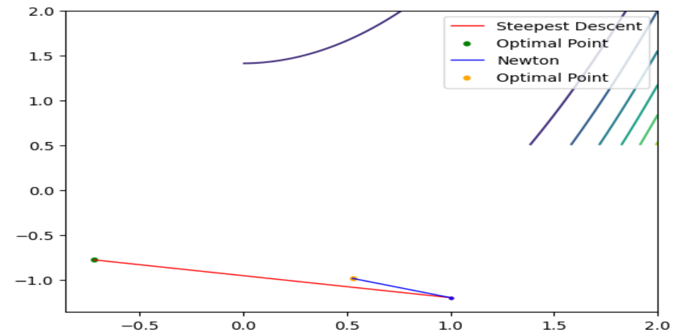

Fig. 3. Behavior of both methods for the Chained Rosenbrock function on first starting point

By examining the same metrics for the second starting point indicated in the subsection, we also assess how well Steepest Descent and Newton's approach perform in terms of optimizing the Chained Rosenbrock function. We give the results in Table IV for a clear comparison.

	Iterations	Time	Gradient Norm
Steepest Descent	1000	55s	1957
Newton	1000	55s	1918

TABLE IV
COMPARISON OF TWO METHODS ON THE CHAINED ROSENBROCK FUNCTION USING THE SECOND STARTING POINT

Figure 4 shows the convergence of the steepest descent and the Newton method for the Chained Rosenbrock function on the second point towards the optimal solution and highlights the performance differences between them.

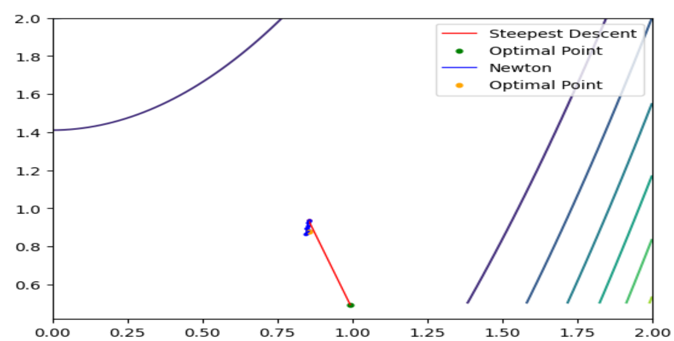

Fig. 4. Behavior of both methods for the Chained Rosenbrock function on second starting point

C. Generalized Brown function

In this analysis, two evaluations were conducted, the first using initial points provided by the function as $\bar{x}_i$, and the second using random points between 0 and 1. Both utilized a dimension of 10^3.

1) Results obtained using the steepest descent method: The initial evaluation took place over 0 iterations, lasted around 0 seconds, and gradient norm 0. The second evaluation took 1000 iterations, lasted around 34 seconds, and gradient norm equal to 15.

2) Results obtained using the Newton method: The initial evaluation took place over 0 iterations, lasted around 0 seconds, and gradient norm 0. The second evaluation took 1000 iterations, lasted around 79 seconds, and gradient norm equal to 74.

3) Comparison of the results obtained using each optimization technique: We compare the elapsed time, number of iterations, and gradient norm for the initial starting point to compare the effectiveness of Steepest Descent and Newton's method in optimizing the Rosenbrock function. To offer insights, we give the findings in Table V.

	Iterations	**Time**	**Gradient Norm**
Steepest Descent	0	0s	0
Newton	0	0s	0

TABLE V
COMPARISON OF TWO METHODS ON THE GENERALIZED BROWN FUNCTION USING THE FIRST STARTING POINT.

Figure 5 shows the convergence of the steepest descent and the Newton method for the Generalized Brown function on the first point towards the optimal solution and highlights the performance differences between them.

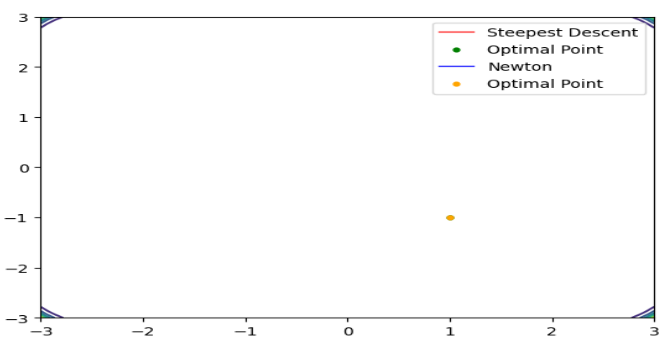

Fig. 5. Behavior of both methods for the Generalized Brown function on first starting point

By examining the same metrics for the second starting point indicated in the subsection, we also assess how well Steepest Descent and Newton's approach perform in terms of optimizing the Generalized Brown function. We give the results in Table VI for a clear comparison.

	Iterations	**Time**	**Gradient Norm**
Steepest Descent	1000	34s	15
Newton	1000	79s	74

TABLE VI
COMPARISON OF TWO METHODS ON THE GENERALIZED BROWN FUNCTION USING THE SECOND STARTING POINT.

Figure 6 shows the convergence of the steepest descent and the Newton method for the Generalized Brown function on the second point towards the optimal solution and highlights the performance differences between them.

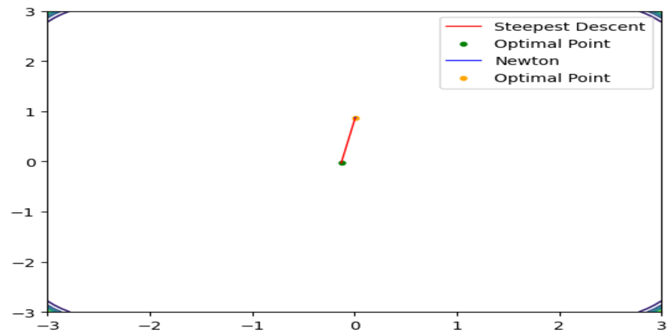

Fig. 6. Behavior of both methods for the Generalized Brown function on second starting point

D. Problem 76

In this analysis, two evaluations were conducted, the first using initial points provided by the function as $\bar{x}_i$, and the second using random points between 0 and 1. Both utilized a dimension of 10^3.

1) Results obtained using the steepest descent method: The initial evaluation took place over 1000 iterations, lasted around 31 seconds, and gradient norm 85. The second evaluation took 1000 iterations, lasted around 32 seconds, and gradient norm equal to 34

2) Results obtained using the Newton method: The initial evaluation took place over 5 iterations, lasted around 0.68 seconds, and gradient norm 0. The second evaluation took 4 iterations, lasted around 0.59 seconds, and gradient norm equal to 0

3) Comparison of the results obtained using each optimization technique: We compare the elapsed time, number of iterations, and gradient norm for the initial starting point to compare the effectiveness of Steepest Descent and Newton's method in optimizing the Problem 76 function. To offer insights, we give the findings in Table VII.

	Iterations	**Time**	**Gradient Norm**
Steepest Descent	1000	31s	85
Newton	5	0.68s	0

TABLE VII
COMPARISON OF TWO METHODS ON THE PROBLEM 76 FUNCTION USING THE FIRST STARTING POINT.

Figure 7 shows the convergence of the steepest descent and the Newton method for the Problem 76 function on the first point towards the optimal solution and highlights the performance differences between them.

By examining the same metrics for the second starting point indicated in the subsection, we also assess how well Steepest Descent and Newton's approach perform in terms of optimizing the Problem 76 function. We give the results in Table VIII for a clear comparison.

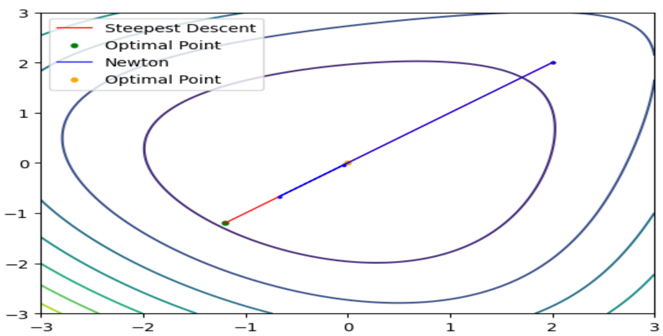

Fig. 7. Behavior of both methods for the Problem 76 function on first starting point

	Iterations	Time	Gradient Norm
Steepest Descent	1000	32s	34
Newton	4	0.59s	0

TABLE VIII
COMPARISON OF TWO METHODS ON THE PROBLEM 76 FUNCTION USING THE SECOND STARTING POINT.

Figure 8 shows the convergence of the steepest descent and the Newton method for the Problem 76 function on the second point towards the optimal solution and highlights the performance differences between them.

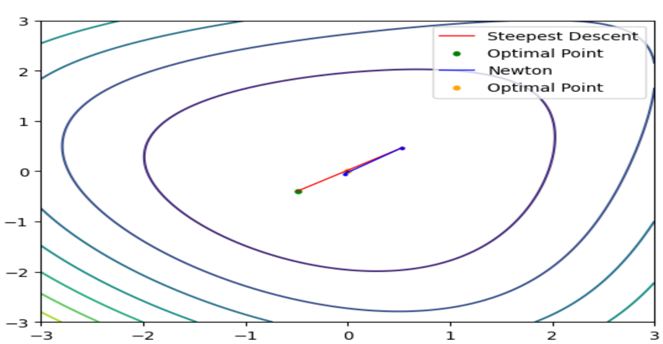

Fig. 8. Behavior of both methods for the Problem 76 function on second starting point

V. Discussion

The aim of this work was to compare and contrast two numerical optimization methods for addressing unconstrained problems, the Steepest Descent Method and the Newton Method with backtracking line search algorithm. The report focused on four test problems: the Rosenbrock function, the Chained Rosenbrock function, the Generalized Brown function, and Problem 76.

The efficacy and efficiency of each optimization strategy differed depending on the test problem, according to the results. In general, the Newton method performed the best in terms of iterations and computing time. Although the steepest descent approach was slow, it yielded acceptable results in certain cases.

The parameter tuning of the backtracking line search algorithm utilized in the steepest descent method is one probable area for future advancements. This can help in further improving the method's performance. The methodologies employed in this study can also be expanded to take into account additional methods and constraints. Future research could also look into using these optimization methods to solve challenges that arise in the real world, like improving machine learning algorithms, financial systems, and supply chain systems.

In summary, the results of this research provide a comprehensive overview of the effectiveness of the numerical optimization techniques used to unconstrained problems. The work done in this study can be utilized to lay the groundwork for future studies aimed at improving and expanding the optimization approaches used.

References

[1] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed. Springer, 2006, pp. 24-27.

[2] J. Nocedal and S. J. Wright, Numerical Optimization, 2nd ed. Springer, 2006, pp. 27-34, 48-52.

[3] Conn, A.R., Gould, N.I.M, Toint, P., "Testing a Class of Methods for Solving Minimization Problems with Simple Bounds on the Variables", Mathematics of Computation, Vol. 50, pp. 399-430, 1988.

[4] Luksan, Ladislav Vlček, Jan. "Test Problems for Unconstrained Optimization". (2003).

[5] Roose, A., Kulla, V., Lomp, M., Meressoo, T., "Test Examples for systems of nonlinear equations", Estonian Software and Computer Service Company, Tallin 1990