Optimization of the Rosenbrock Function

Alejandro Alonso Sanchez

February 11, 2025

Abstract The Rosenbrock function is a well-known non-convex function used to test optimization algorithms. In this report, I investigate the performance of Newton's Method, Gradient Descent, and BFGS in minimizing the Rosenbrock function. The step size is optimized using Strong Wolfe conditions. The results confirm that all methods converge to the optimal solution at $(1, 1, 0)$ when initialized from $(-1.5, 2)$.

Keywords Rosenbrock function · Optimization · Newton's Method · Gradient Descent · BFGS · Wolfe Conditions

1 Introduction

The Rosenbrock function, also called the Rosenbrock valley or banana function, is a common test problem for optimization algorithms. It is defined as:

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2. \tag{1}$$

The function has a narrow, curved valley leading to its global minimum at $(1, 1)$. This makes optimization challenging, as algorithms must efficiently navigate the curved landscape.

2 Optimization Methods

In this section, I define the three optimization methods used.

Universidad de Guanajuato Campus Irapuato-Salamanca

2.1 Gradient Descent

Gradient descent iteratively updates the solution using the negative gradient:

$$x_{k+1} = x_k - \alpha \nabla f(x_k), \tag{2}$$

where α is the step size, chosen to satisfy the Strong Wolfe conditions.

2.2 Newton's Method

Newton's Method uses second-order information to accelerate convergence:

$$x_{k+1} = x_k - H^{-1} \nabla f(x_k), \tag{3}$$

where H is the Hessian matrix of second derivatives.

2.3 BFGS (Quasi-Newton)

BFGS is an iterative method that approximates the inverse Hessian:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}, \tag{4}$$

where $s_k = x_{k+1} - x_k$ and $y_k = \nabla f_{k+1} - \nabla f_k$.

3 Step Size Optimization

The step size α is chosen using the Strong Wolfe conditions:

$$f(x_k + \alpha d_k) \leq f(x_k) + c_1 \alpha \nabla f(x_k)^T d_k, \tag{5}$$

$$|\nabla f(x_k + \alpha d_k)^T d_k| \leq c_2 |\nabla f(x_k)^T d_k|. \tag{6}$$

These conditions ensure a balance between sufficient descent and stable step sizes.

4 Results

I initialized the optimization algorithms at $(-1.5, 2)$. The following plots show the function surface and optimization paths:

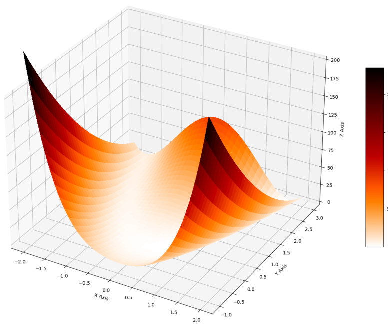

Fig. 1 3D Surface Plot of the Rosenbrock Function

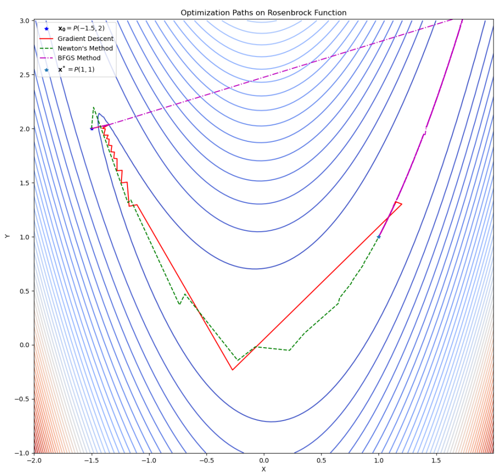

Fig. 2 Optimization Paths for Newton, Gradient Descent, and BFGS

All methods successfully converged to $(1, 1)$. Newton's method converged in the fewest iterations (40), followed by gradient descent (150) and finally, BFGS was the slowest (1100) and had a strange trajectory (see Figure 3).

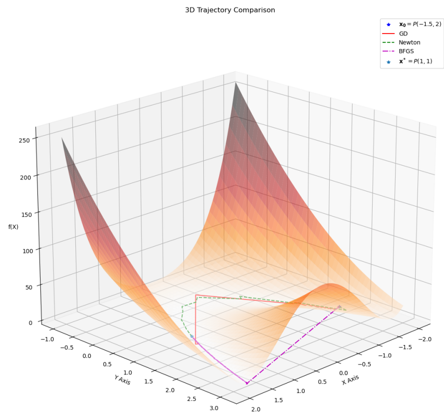

Fig. 3 3D view of Optimization paths for Newton, Gradient Descent and BFGS

5 Conclusion

This report analyzed the optimization of the Rosenbrock function using three methods. By aplying the Strong Wolfe conditions, optimal step sizes were produced. The results confirm successful convergence to the global minimum.

References

1. Rosenbrock, H. H. An Automatic Method for Finding the Greatest or Least Value of a Function. *The Computer Journal*, 1960.
2. Nocedal, J., Wright, S. Numerical Optimization. *Springer*, 2006.