Artificial Intelligence Project Report: Adam Optimizer

Report by: Athang Sanjeev Bhandarkar, Harsh Gujarathi, Vinay Warrier.

Rutgers the State University of New Jersey, Camden.

04/25/2024

Abstract

In this project, we explore the application of the Adam optimization algorithm in solving the Rosenbrock function optimization problem. The Adam algorithm is a popular gradient descent optimization technique that adapts learning rates for each parameter, leading to efficient convergence in various optimization tasks. We investigate the behavior of the Adam algorithm with different hyperparameters and compare its performance with traditional gradient descent methods.

Introduction

The Rosenbrock function, also known as the Rosenbrock's valley or banana function, is a non-convex optimization problem often used as a performance test for optimization algorithms. It is defined as a sum of squared differences of consecutive pairs of variables with a global minimum at the point (1, 1). Our objective is to minimize this function using the Adam optimization algorithm.

Implementation

We implemented the Adam optimization algorithm for minimizing the Rosenbrock function. The algorithm updates parameters based on the first and second moments of gradients to adaptively adjust the learning rates. We defined the Rosenbrock function and its derivative, which are necessary for gradient descent. We tuned the hyperparameters of the Adam algorithm, including the learning rate (alpha), the decay rate of the first-moment estimate (beta1), and the decay rate of the second-moment estimate (beta2). Additionally, we adjusted the number of iterations to ensure convergence.

$$\left.\begin{aligned} V_{dw} &= \beta_1 V_{dw} + (1-\beta)\frac{\partial h}{\partial w} \\ V_{db} &= \beta_1 V_{dw} + (1-\beta)\frac{\partial h}{\partial b} \end{aligned}\right\}$$

Momentum → Smoothening factor

$$\left.\begin{aligned} S_{dw} &= \beta_2 S_{dw} + (1-\beta)\left(\frac{\partial h}{\partial w}\right)^2 \\ S_{db} &= \beta_2 S_{db} + (1-\beta)\left(\frac{\partial h}{\partial b}\right)^2 \end{aligned}\right\}$$

Rms Prop ↳

Bias Correction

initial LR

$$W_t = W_{t-1} - \frac{\eta * V_{dw}}{\sqrt{S_{dw} + \epsilon}}$$

$$b_t = b_{t-1} - \frac{\eta * V_{db}}{\sqrt{S_{db} + \epsilon}}$$

Adam optimizer

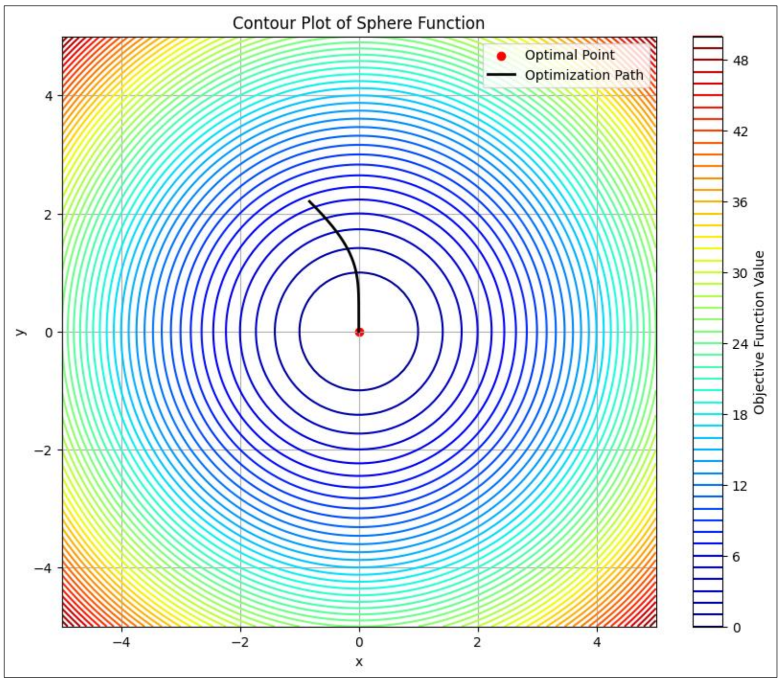
→ Smoothening factor

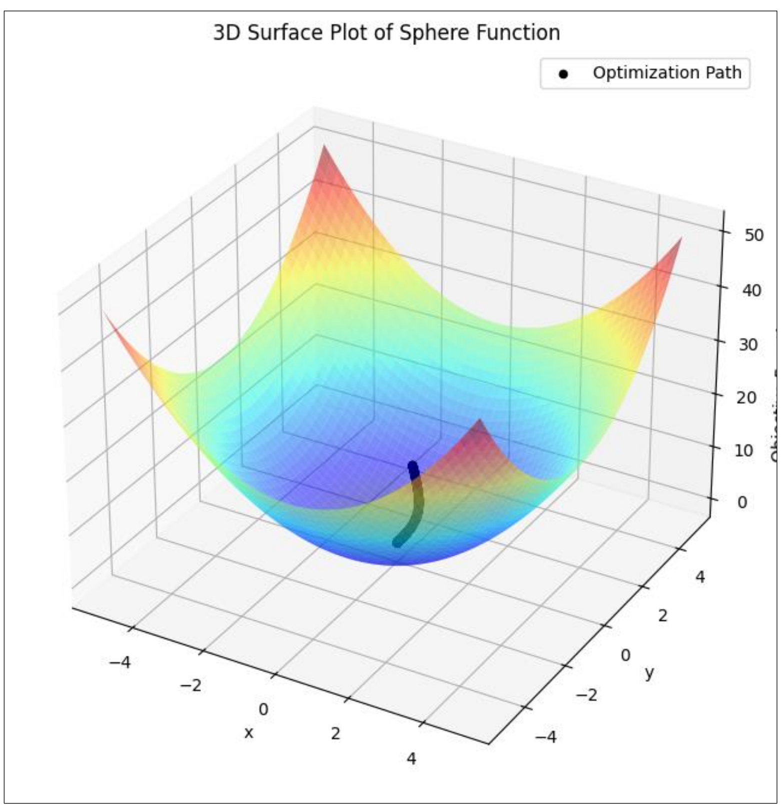
↳ Rms propu

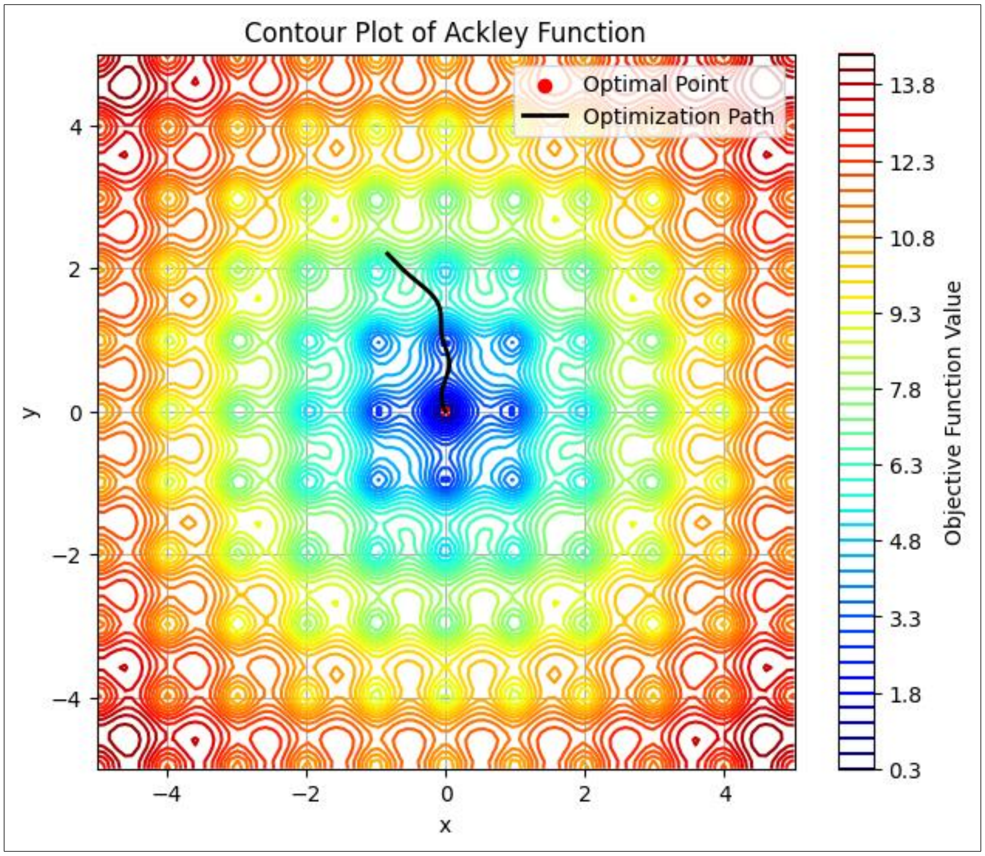
Require: α: Stepsize
Require: $\beta_1, \beta_2 \in [0, 1)$: Exponential decay rates for the moment estimates
Require: $f(\theta)$: Stochastic objective function with parameters θ
Require: θ_0: Initial parameter vector
  $m_0 \leftarrow 0$ (Initialize 1st moment vector)
  $v_0 \leftarrow 0$ (Initialize 2nd moment vector)
  $t \leftarrow 0$ (Initialize timestep)
  while θ_t not converged **do**
    $t \leftarrow t + 1$
    $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ (Get gradients w.r.t. stochastic objective at timestep t)
    $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Update biased first moment estimate)
    $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Update biased second raw moment estimate)
    $\widehat{m}_t \leftarrow m_t/(1 - \beta_1^t)$ (Compute bias-corrected first moment estimate)
    $\widehat{v}_t \leftarrow v_t/(1 - \beta_2^t)$ (Compute bias-corrected second raw moment estimate)
    $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \widehat{m}_t/(\sqrt{\widehat{v}_t} + \epsilon)$ (Update parameters)
  end while
  return θ_t (Resulting parameters)

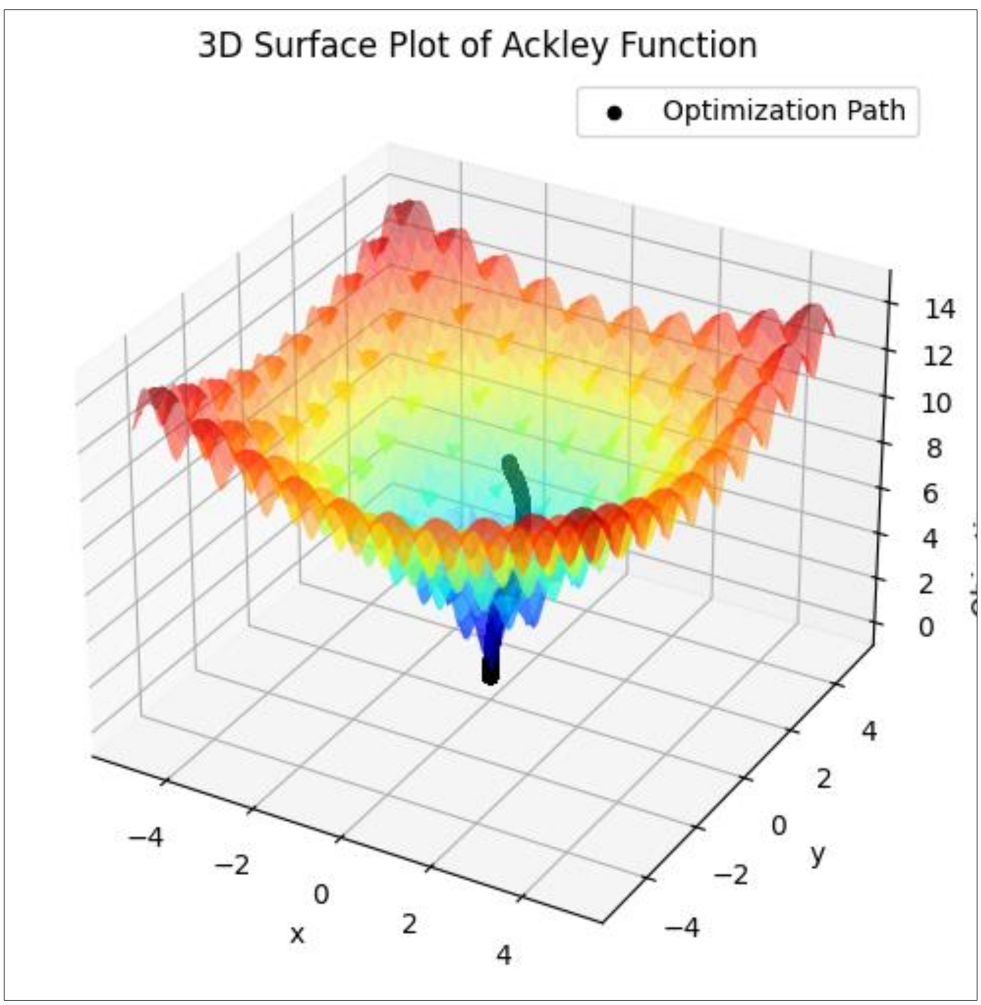
Results

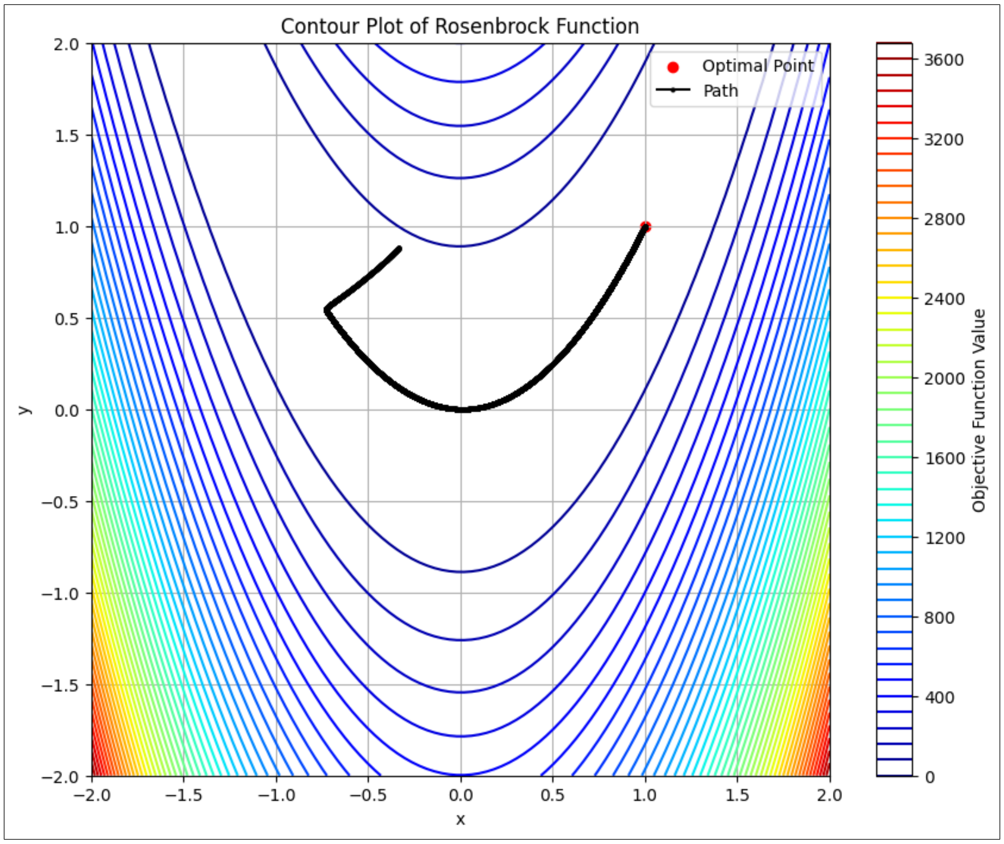
After running the Adam optimization algorithm, we observed the progress of the algorithm as it iteratively minimized the Rosenbrock function. We monitored the function value (score) at each iteration to track the convergence of the algorithm. By adjusting the hyperparameters and the number of iterations, we obtained a solution that approaches the global minimum of the Rosenbrock function.

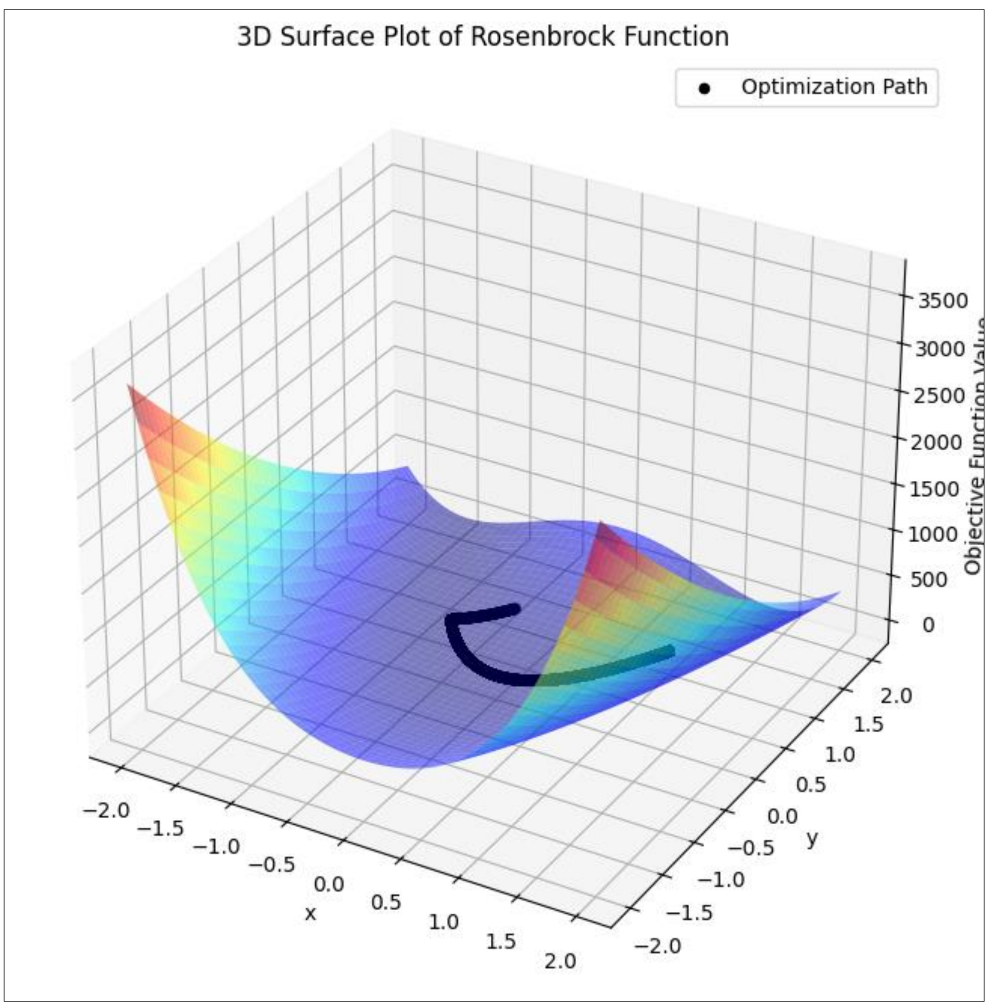





Contour Plot of Ackley Function
Optimal Point
Optimization Path
Objective Function Value
13.8
12.3
10.8
9.3
7.8
6.3
4.8
3.3
1.8
0.3
x
y


3D Surface Plot of Ackley Function
Optimization Path
x
y


Contour Plot of Rosenbrock Function
Optimal Point
Path
Objective Function Value
x
y


3D Surface Plot of Rosenbrock Function
Optimization Path
x
y

Benchmark Functions:

1. Sphere Function:

- The Sphere function is one of the simplest optimization benchmark functions used in optimization algorithms.
- The Sphere function is often used to test optimization algorithms due to its simplicity and well-defined global minimum.

2. Rosenbrock Function:

- The Rosenbrock function is a non-convex optimization problem used as a performance test problem for optimization algorithms.
- The function has a narrow, banana-shaped valley with steep sides, making it challenging for optimization algorithms to converge to the global minimum.

3. Ackley Function:

- The Ackley function is another widely used optimization benchmark function known for its multimodality and lack of regularity.
- It has multiple local minima and one global minimum, making it challenging for optimization algorithms to find the global optimum.

Conclusion

In conclusion, the Adam optimization algorithm demonstrates its effectiveness in minimizing the Rosenbrock function. By adaptively adjusting learning rates, Adam efficiently navigates the complex landscape of the Rosenbrock function and converges to its global minimum. Through this project, we gained insights into the behavior of the Adam algorithm and its applicability in solving non-convex optimization problems.

References

- Adam, A Method for Stochastic Optimization: https://arxiv.org/pdf/1412.6980
- Adam Optimizer: https://youtu.be/tuU59-G1PgU?si=h6lHLrHTg_ZmOkWN