Optimizer Performance on Non-Convex Functions

Mohd Darish Khan, Hardik Singh, Suryansh Jaiswal (Group - 4)

2101MC29, 2101MC19, 2101MC41

Introduction

In this report, we optimize two non-convex functions using various optimization algorithms. The optimization is performed with different learning rates to evaluate the performance of each algorithm.

Non-Convex Function Optimization

Function Definitions

We consider the following non-convex functions:

- **Function 1:** $f(x, y) = (1 - x)^2 + 100(y - x^2)^2$
- **Function 2:** $f(x) = \sin\left(\frac{1}{x}\right)$ with $f(0) = 0$

Optimization Algorithms

The following optimization algorithms are used:

- Gradient Descent
- Stochastic Gradient Descent (SGD)
- Adam
- RMSprop
- Adagrad

Learning Rates

The learning rates (α) used for optimization are:

- $\alpha = 0.01$
- $\alpha = 0.05$
- $\alpha = 0.1$

Results

Optimization of Function 1: $f(x, y) = (1 - x)^2 + 100(y - x^2)^2$

Convergence Behavior:

- *Gradient Descent:*

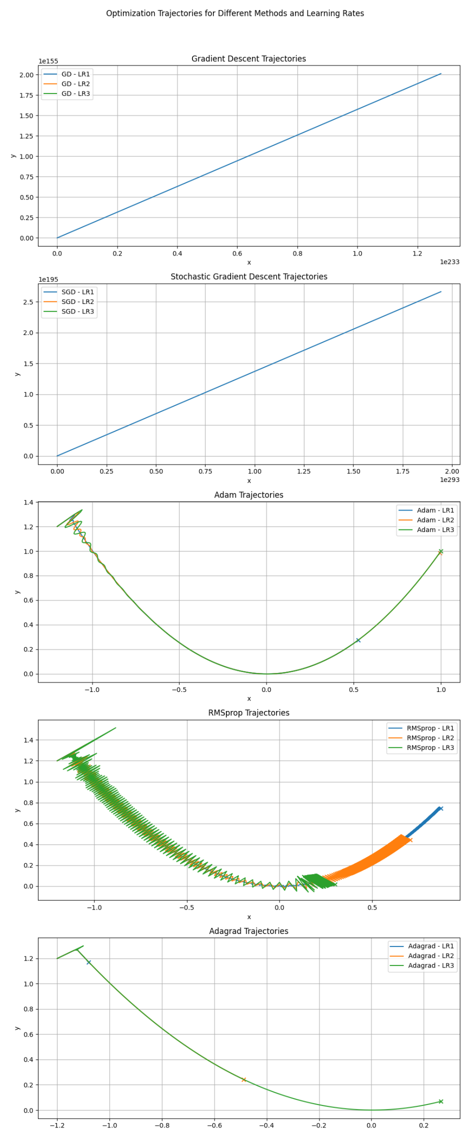

Figure 1:

- *SGD:*
- *Adam:*
- *RMSprop:*
- *Adagrad:*

Inference: The convergence of Adam and AdaGrad were better than others.

Impact of Learning Rates

Inference: Smaller learning rate made convergence easier.

Optimization of Function 2: $f(x) = \sin\left(\frac{1}{x}\right)$

Convergence Behavior:

- *Gradient Descent:*
- *SGD:*
- *Adam:*
- *RMSprop:*
- *Adagrad:*

Inference: The convergence of Adam and AdaGrad were better than others.

Impact of Learning Rates

Inference: Smaller learning rate made convergence easier.

Discussion

The first function converges at (1, 1), while the second one converges at around 0.63 for substantial positive values.

Conclusion

Adam and Adagrad were the best performing algorithms.

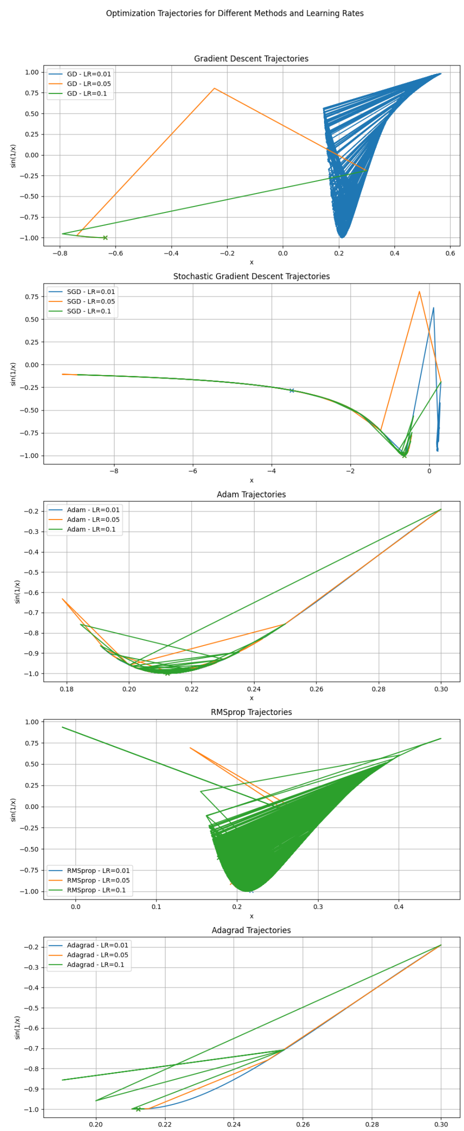
Optimization Trajectories for Different Methods and Learning Rates
Gradient Descent Trajectories
GD - LR=0.01
GD - LR=0.05
GD - LR=0.1
sin(1/x)
x
Stochastic Gradient Descent Trajectories
SGD - LR=0.01
SGD - LR=0.05
SGD - LR=0.1
sin(1/x)
x
Adam Trajectories
Adam - LR=0.01
Adam - LR=0.05
Adam - LR=0.1
sin(1/x)
x
RMSprop Trajectories
RMSprop - LR=0.01
RMSprop - LR=0.05
RMSprop - LR=0.1
sin(1/x)
x
Adagrad Trajectories
Adagrad - LR=0.01
Adagrad - LR=0.05
Adagrad - LR=0.1
sin(1/x)
x

Figure 2: