Task 1: Optimizer Performance on Non-Convex Functions

1. Objective

The objective of this experiment is to study and compare the performance of different optimization algorithms on non-convex functions with respect to convergence behavior, sensitivity to learning rate, final solution quality, and computational time.

The optimizers considered are Gradient Descent (GD), Stochastic Gradient Descent with Momentum, Adagrad, RMSprop, and Adam.

Each optimizer is tested with learning rates alpha = {0.01, 0.05, 0.1}.

2. Test Functions

2.1 Rosenbrock Function

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

This is a classic non-convex function with a narrow curved valley and a global minimum at (1, 1). Due to numerical instability at higher learning rates, parameter projection was applied to restrict x and y to the range [-5, 5].

2.2 Oscillatory Function

$f(x) = \sin(1/x)$, with $f(0) = 0$

This function is highly oscillatory near x = 0 and contains many local extrema. No parameter projection was applied for this function.

3. Experimental Setup

All optimizers were implemented from scratch using Python and NumPy. The same initialization was used for fair comparison: Rosenbrock was initialized at (-1.5, 1.5) and sin(1/x) was initialized at x = 0.1. The termination criteria were a gradient norm less than 1e-6 or reaching the maximum number of iterations. Execution time was measured using Python's time module.

4. Results and Observations

4.1 Rosenbrock Function Results

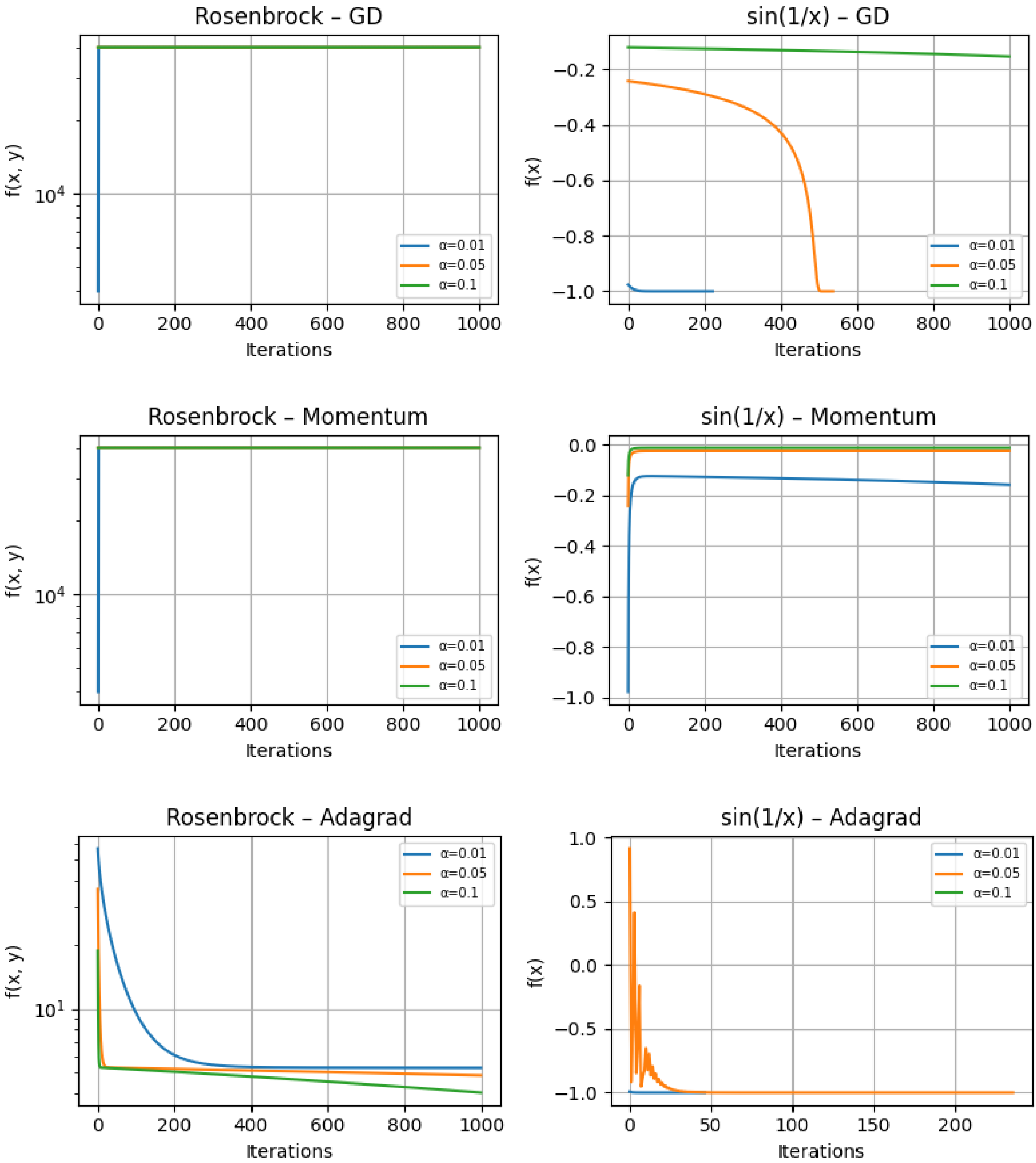
Optimizer	alpha	Final (x, y)	f(x, y)	Time (s)
GD	0.01	(-5, 5)	40036.0	0.033
GD	0.05	(-5, 5)	40036.0	0.036

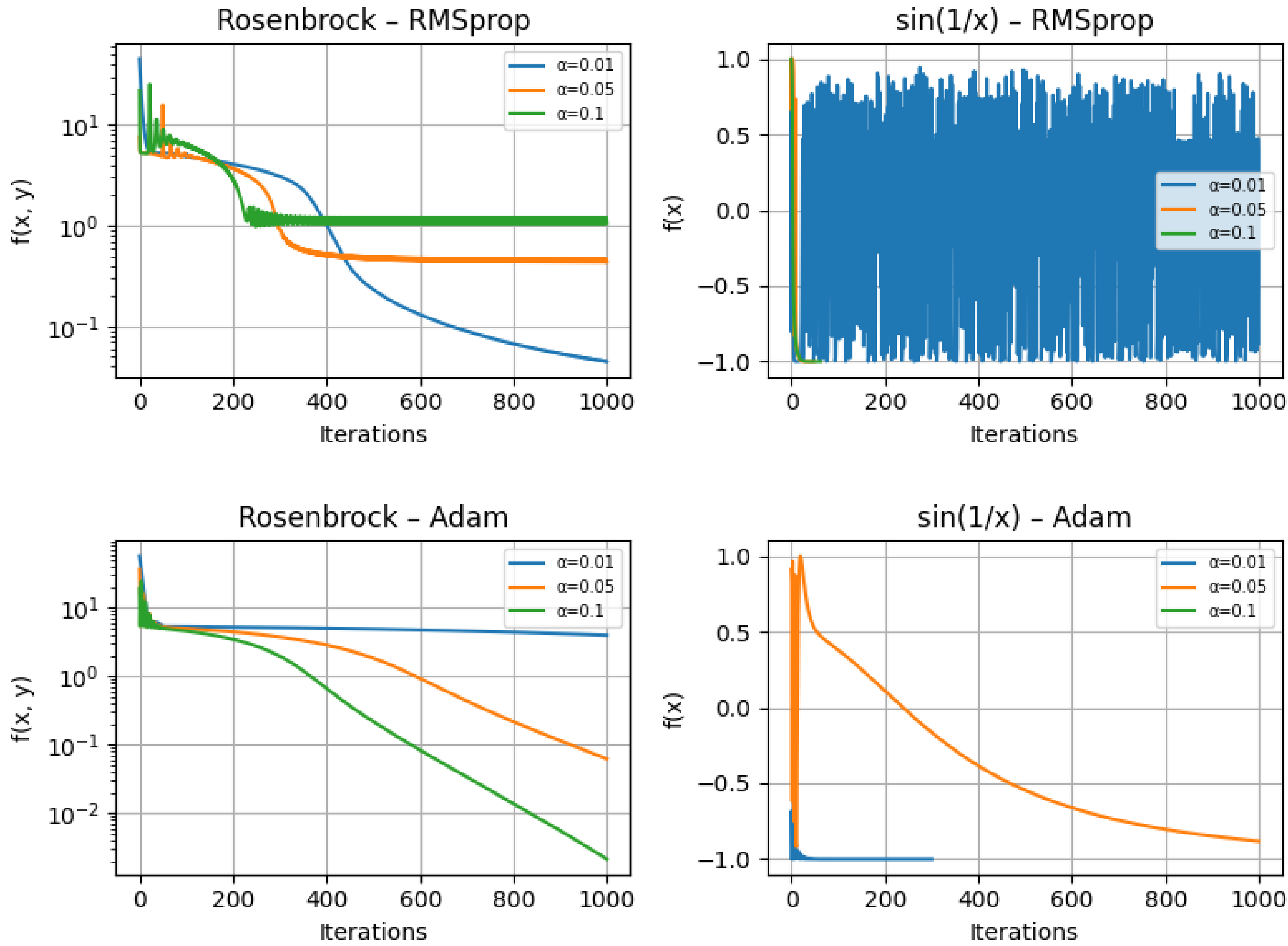
GD	0.1	(-5, 5)	40036.0	0.029
Momentum	0.01	(-5, 5)	40036.0	0.034
Momentum	0.05	(-5, 5)	40036.0	0.033
Momentum	0.1	(-5, 5)	40036.0	0.014
Adagrad	0.01	(-1.30, 1.69)	5.28	0.017
Adagrad	0.05	(-1.21, 1.46)	4.87	0.017
Adagrad	0.1	(-1.01, 1.02)	4.04	0.018
RMSprop	0.01	(0.84, 0.69)	0.045	0.019
RMSprop	0.05	(0.70, 0.43)	0.44	0.019
RMSprop	0.1	(0.30, 0.015)	1.05	0.019
Adam	0.01	(-0.99, 0.98)	3.95	0.024
Adam	0.05	(0.75, 0.56)	0.062	0.024
Adam	0.1	(0.95, 0.91)	0.0021	0.024

4.2 sin(1/x) Function Results

Optimizer	alpha	Final x	f(x)	Time (s)
GD	0.01	-0.6366	-1.00	0.0018
GD	0.05	-0.6366	-1.00	0.0041
GD	0.1	-6.49	-0.154	0.0084
Momentum	0.01	-6.29	-0.158	0.0094
Momentum	0.05	-41.52	-0.024	0.0095
Momentum	0.1	-83.64	-0.012	0.011
Adagrad	0.01	0.0909	-1.00	0.0007
Adagrad	0.05	0.0909	-1.00	0.0032
Adagrad	0.1	~0	0	0.00004
RMSprop	0.01	0.056	-0.880	0.017
RMSprop	0.05	0.212	-1.00	0.0007
RMSprop	0.1	-0.6366	-1.00	0.0009
Adam	0.01	0.0909	-1.00	0.0063
Adam	0.05	-0.485	-0.882	0.024
Adam	0.1	~0	0	0.00009

4.3 Convergence Plots


Rosenbrock – GD
sin(1/x) – GD
Rosenbrock – Momentum
sin(1/x) – Momentum
Rosenbrock – Adagrad
sin(1/x) – Adagrad
f(x, y)
f(x)
Iterations
α=0.01
α=0.05
α=0.1



5. Conclusion

The experiment demonstrates that adaptive optimization algorithms outperform traditional gradient-based methods on non-convex functions. Adam provides the best overall performance in terms of convergence speed, stability, and final solution quality, making it the most suitable optimizer among those tested.