Venkatakrishnan E
23110357
Problem 1:

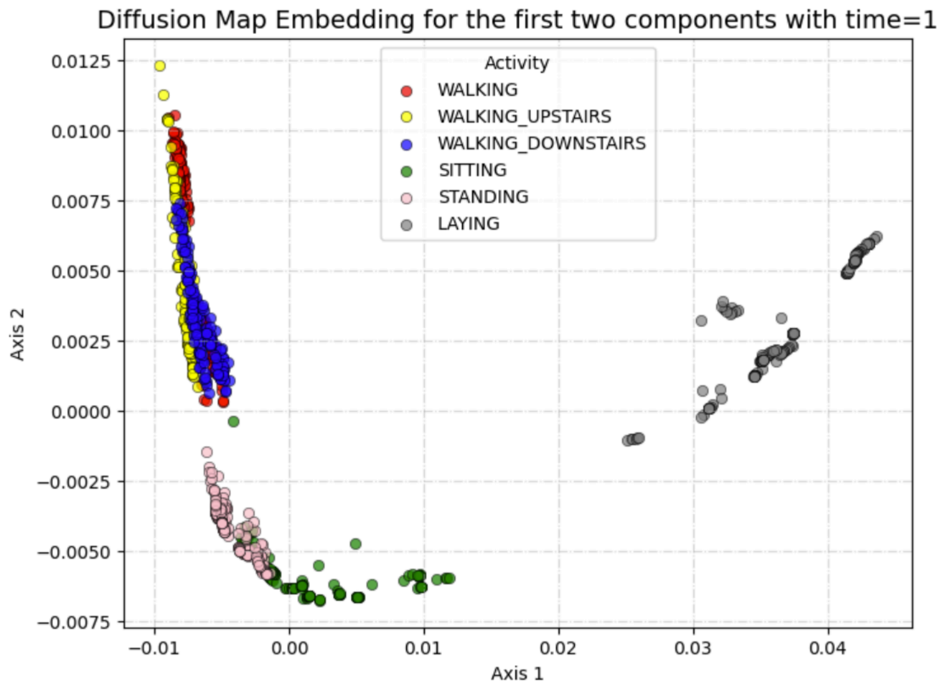
Discuss why Diffusion Maps work well for time-series clustering.

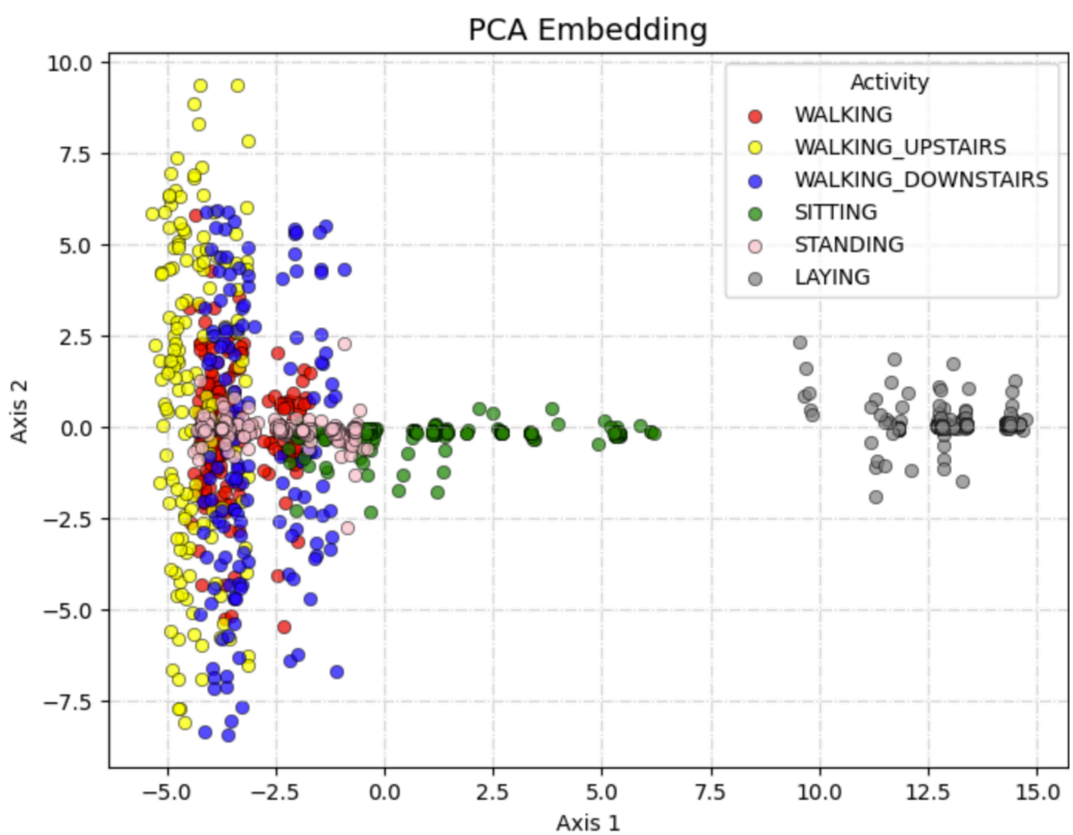
- Diffusion maps capture the underlying manifold structure of high dimensional time series data
- It constructs a graph based representation of the data, where we group points based on local similarity(which is basically the DTW distance), the edges of the graph are the transition probabilities
- We can find non linear temporal patterns with the help of diffusion maps
- Time-series data often contain noise and non-linear variations, but diffusion maps mainly focuses on the global structure thus reducing the impact of noise
- Diffusion Maps reduce the dimensionality while preserving meaningful distances.

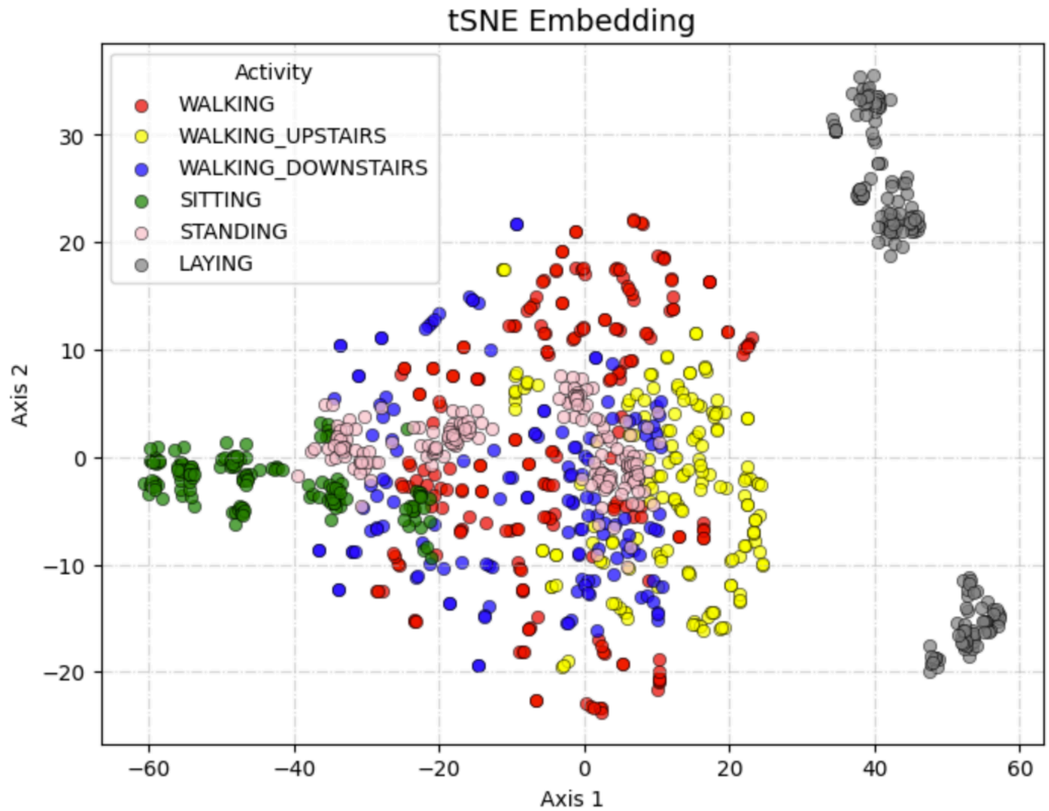
Why Diffusion Maps outperform PCA/t-SNE?

PCA assumes a linear structure, but the time series data generally is non linear, this is evident from the lower ARI and silhouette score compared to diffusion maps.
t-SNE is designed for visualization and relies on pairwise similarities to map data points into a lower-dimensional space.
t-SNE can distort the global structure thus making it not suitable for clustering of time series data.
Higher ARI scores indicate better clustering alignment with true labels, higher values of Silhouette score indicate better separation between clusters.

Technique	ARI score	Silhouette score
Diffusion Maps	0.53	0.52
PCA	0.39	0.45
tSNE	0.24	0.47


Diffusion Map Embedding for the first two components with time=1
Activity
WALKING
WALKING_UPSTAIRS
WALKING_DOWNSTAIRS
SITTING
STANDING
LAYING
Axis 1
Axis 2
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000
−0.0025
−0.0050
−0.0075
−0.01
0.00
0.01
0.02
0.03
0.04


PCA Embedding
Activity
WALKING
WALKING_UPSTAIRS
WALKING_DOWNSTAIRS
SITTING
STANDING
LAYING
Axis 1
Axis 2
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
12.5
15.0

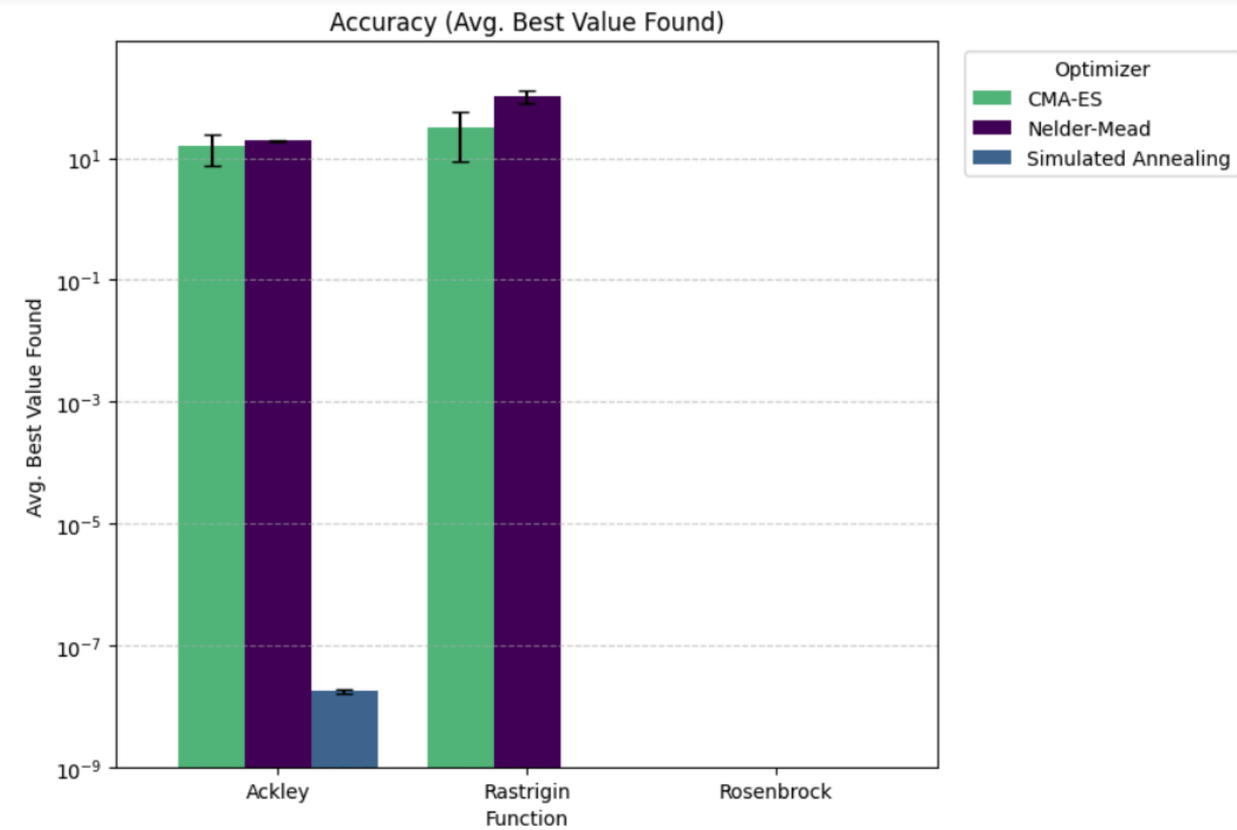


Why using DTW

- In time-series clustering we mostly care about the shape of trends, not the raw values at a timestamp.
- DTW focuses on overall similarity by warping the time axis to align similar patterns thus making it more meaningful than Euclidean distance.
- Euclidean distance compares points one to one and thus it is very sensitive to misalignment.

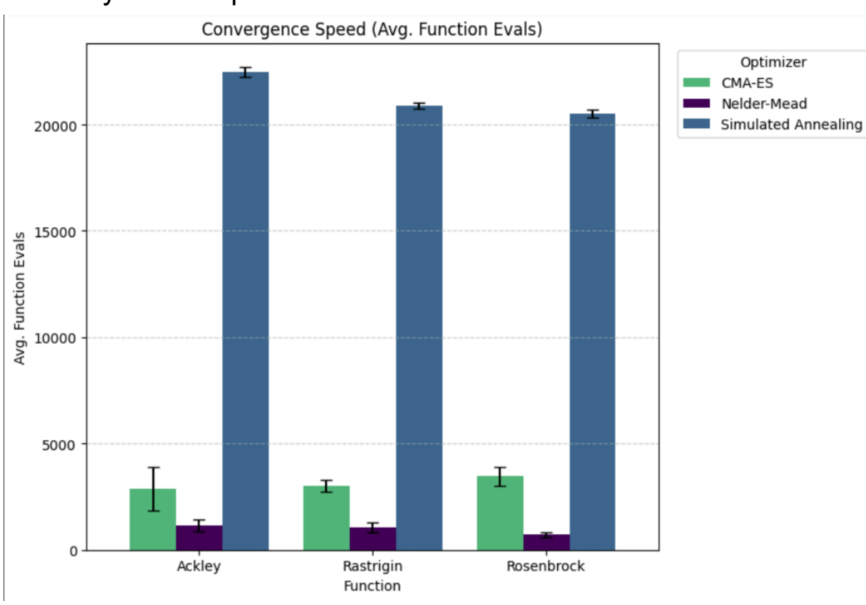
Problem 2:

```
  Performance Analysis: Accuracy & Speed
---------------------------------------------

--- Accuracy Comparison (Mean Best Value Found) ---
Lower values indicate better accuracy.
  Function           Optimizer  Mean_Best_Value  Std_Best_Value
    Ackley Simulated Annealing       1.7829e-08      1.3768e-09
    Ackley              CMA-ES       1.5526e+01      8.1870e+00
    Ackley         Nelder-Mead       1.9436e+01      3.6033e-01
 Rastrigin Simulated Annealing       4.6896e-14      2.7662e-14
 Rastrigin              CMA-ES       3.2535e+01      2.3634e+01
 Rastrigin         Nelder-Mead       1.0268e+02      2.4063e+01
Rosenbrock              CMA-ES       7.6300e-15      9.8886e-15
Rosenbrock Simulated Annealing       7.5070e-12      3.5575e-12
Rosenbrock         Nelder-Mead       2.3910e-10      2.4092e-10

--- Convergence Speed Comparison (Mean Function Evaluations) ---
Lower values indicate faster convergence (fewer function calls).
  Function           Optimizer  Mean_Function_Evals  Std_Function_Evals
    Ackley         Nelder-Mead                 1154                 265
    Ackley              CMA-ES                 2886                1032
    Ackley Simulated Annealing                22472                 220
 Rastrigin         Nelder-Mead                 1058                 218
 Rastrigin              CMA-ES                 3020                 293
 Rastrigin Simulated Annealing                20880                 138
Rosenbrock         Nelder-Mead                  726                 121
Rosenbrock              CMA-ES                 3456                 436
Rosenbrock Simulated Annealing                20511                 184
```



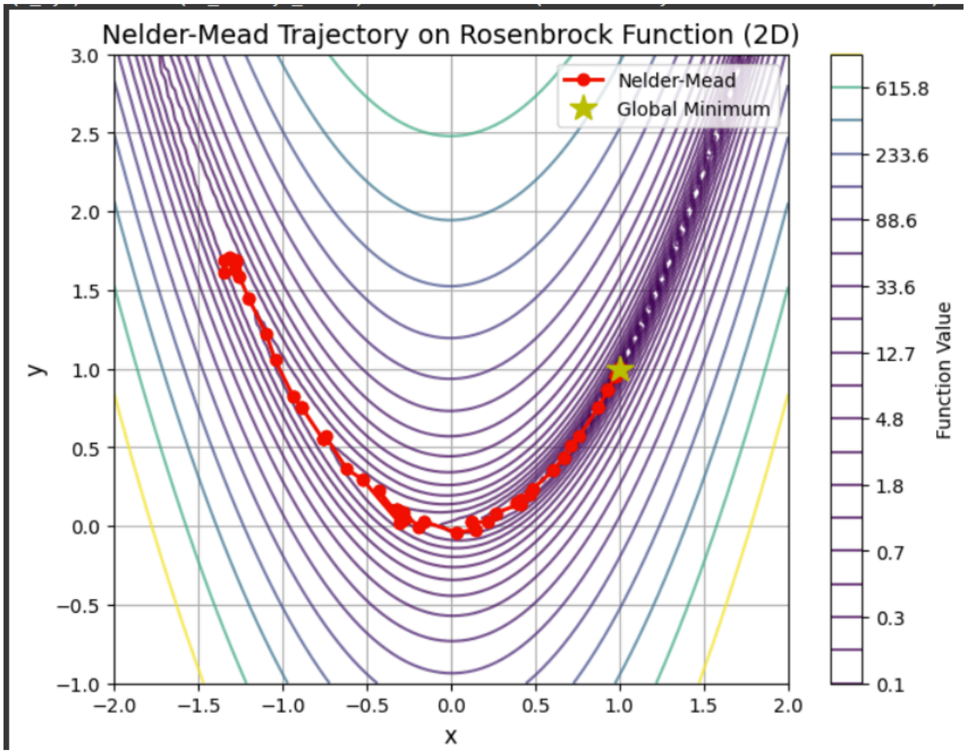
Here for Rosenbrock the value is <10^-9, it nearly converged to 0 for all three optimization techniques, as Rosenbrock was fairly straightforward compared to Ackey and the Rastrigin.

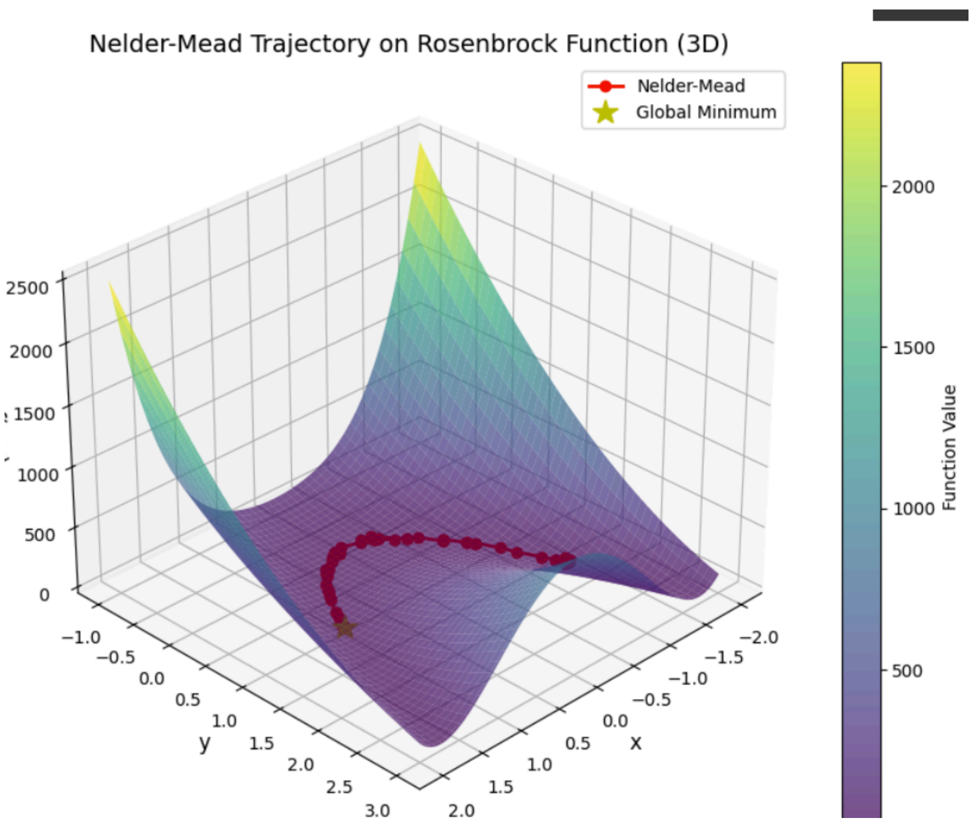
For all 3 functions, simulated annealing has given the least value which indicates the best accuracy as our optimal solution is 0.

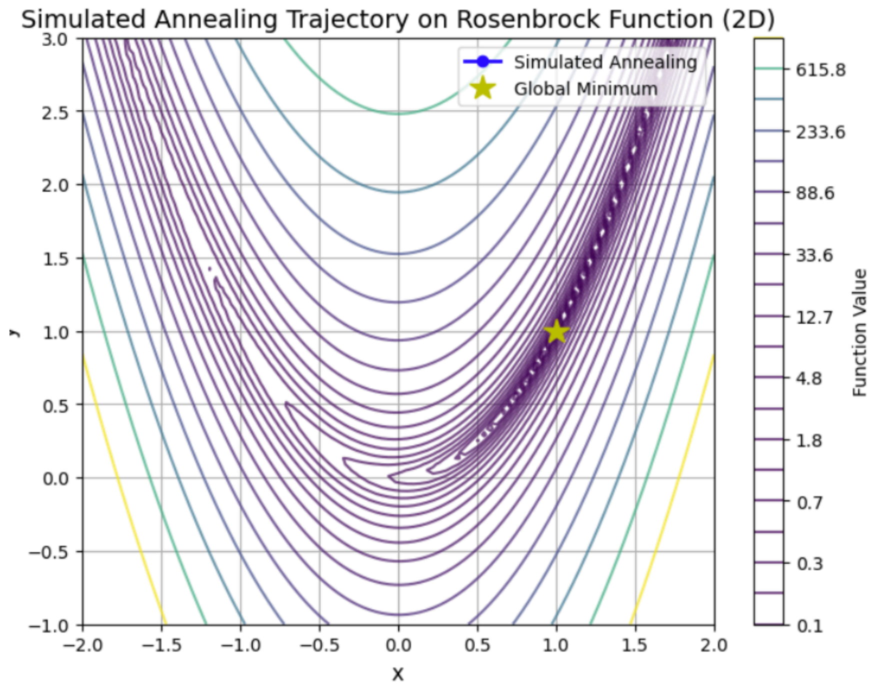


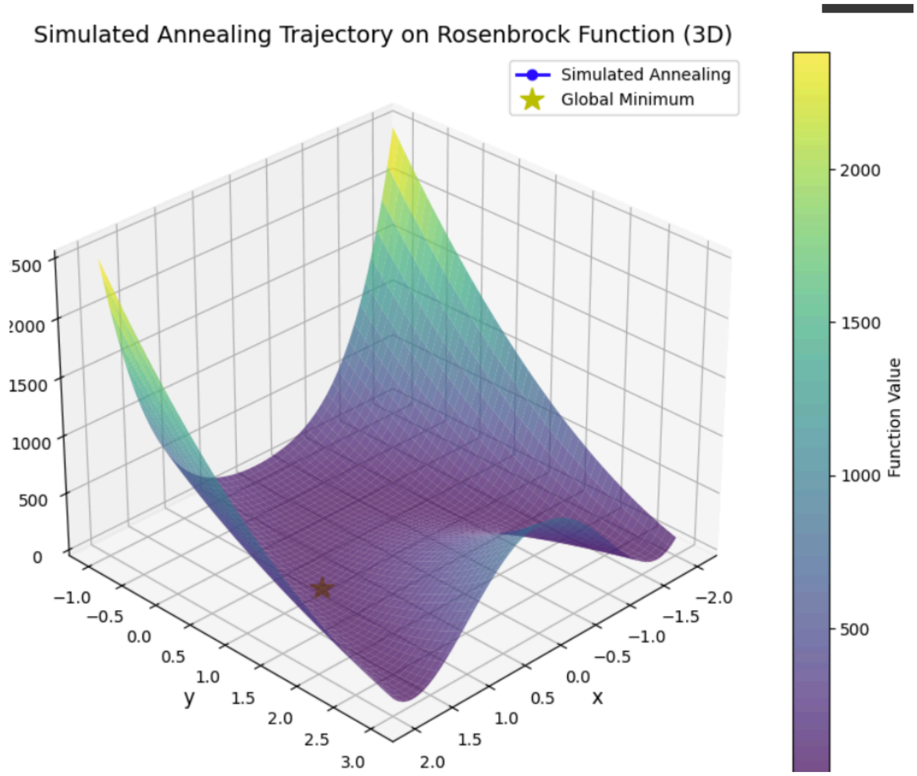
Now in the sense of time taken for convergence, simulated annealing even though it gave us the best results has taken a lot of time to converge, nearly 20000 function evaluations.

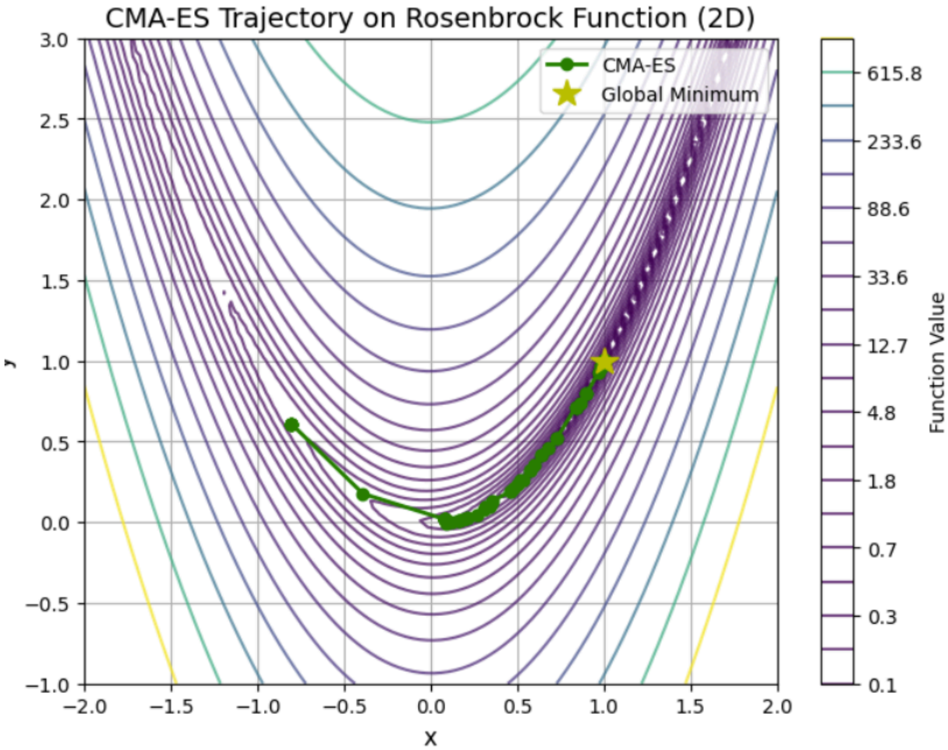
The plots for the optimization trajectories(These plots are also present in the collab notebook)

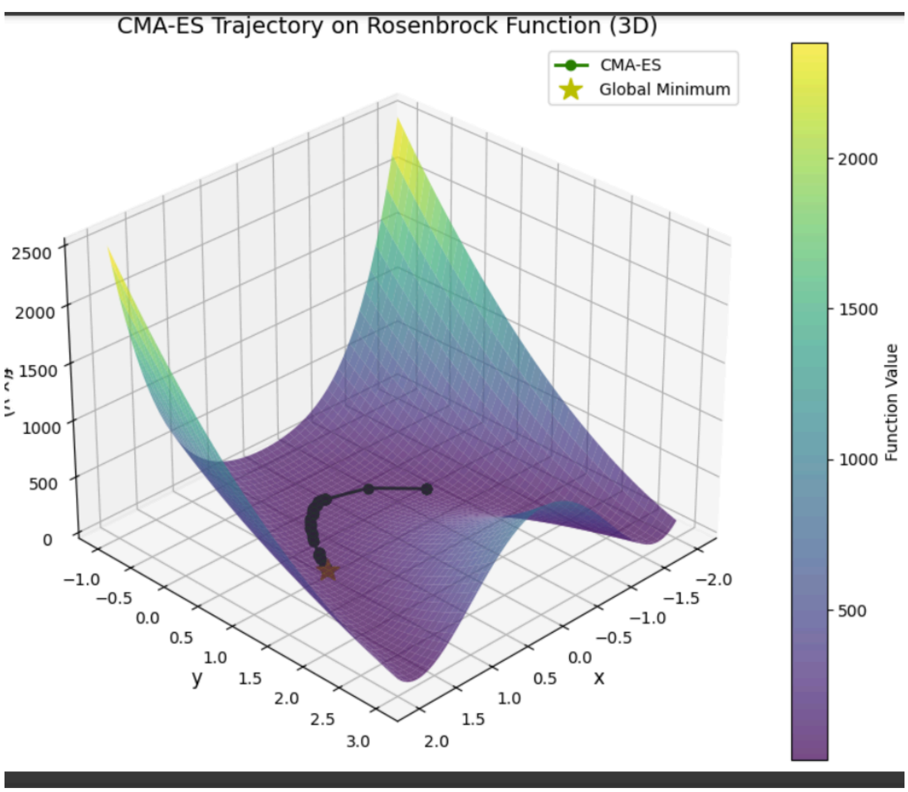


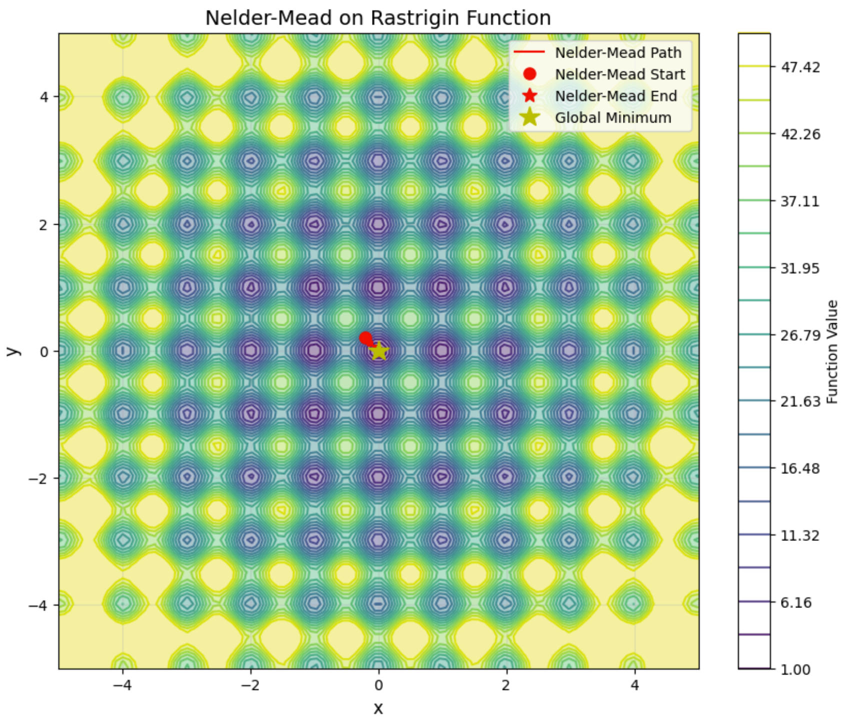


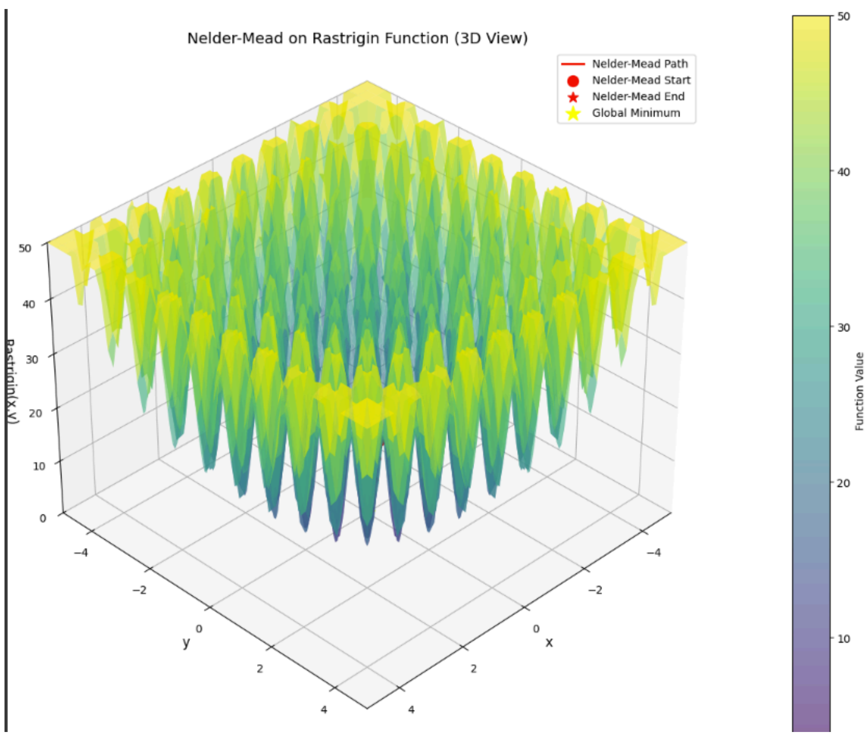

Simulated Annealing Trajectory on Rosenbrock Function (2D)
Simulated Annealing
Global Minimum
x
y
Function Value
615.8
233.6
88.6
33.6
12.7
4.8
1.8
0.7
0.3
0.1

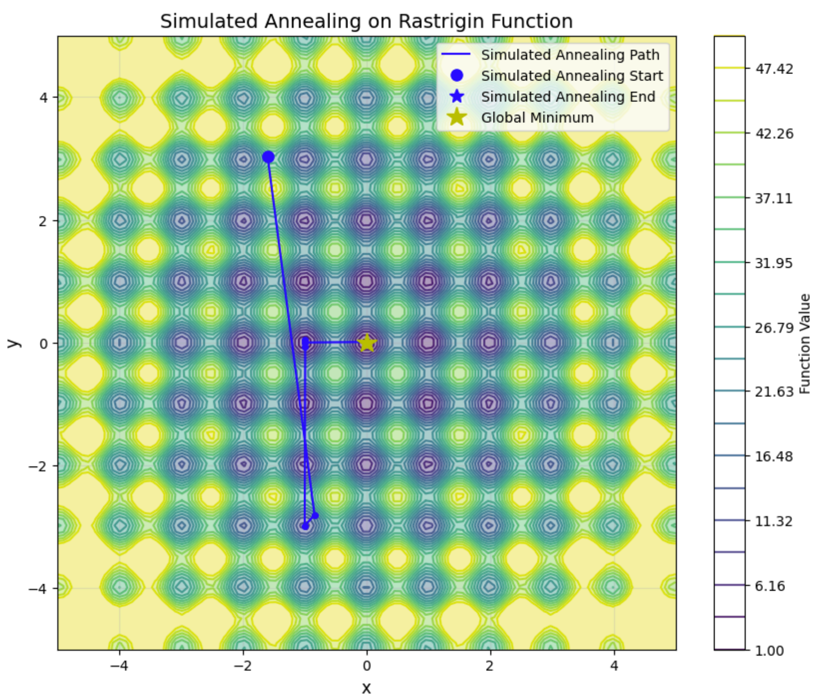

Simulated Annealing Trajectory on Rosenbrock Function (3D)
Simulated Annealing
Global Minimum
x
y
Function Value
2000
1500
1000
500

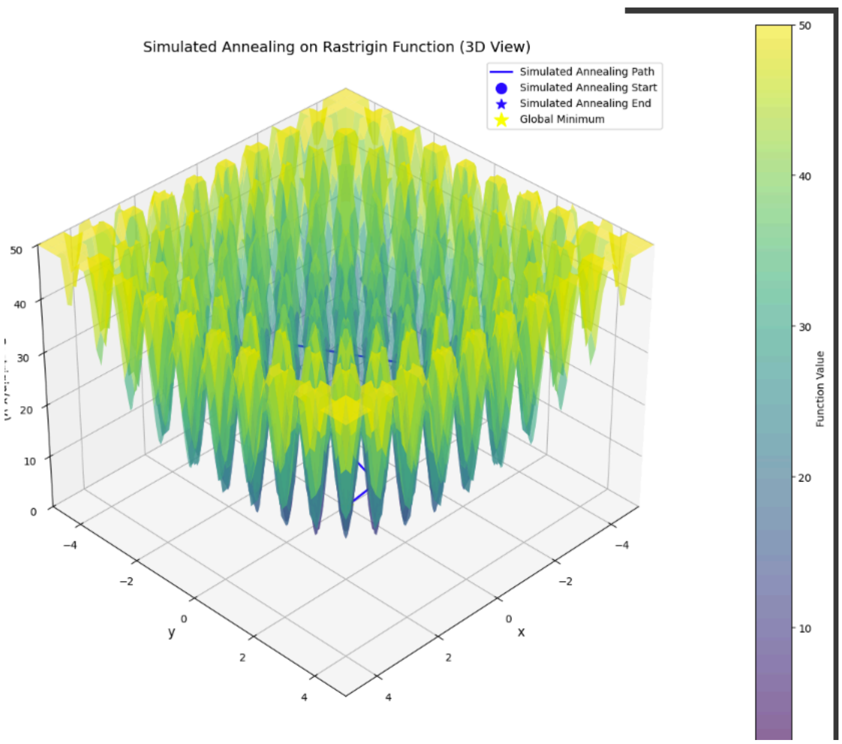

CMA-ES Trajectory on Rosenbrock Function (2D)
CMA-ES
Global Minimum
x
y
Function Value
615.8
233.6
88.6
33.6
12.7
4.8
1.8
0.7
0.3
0.1

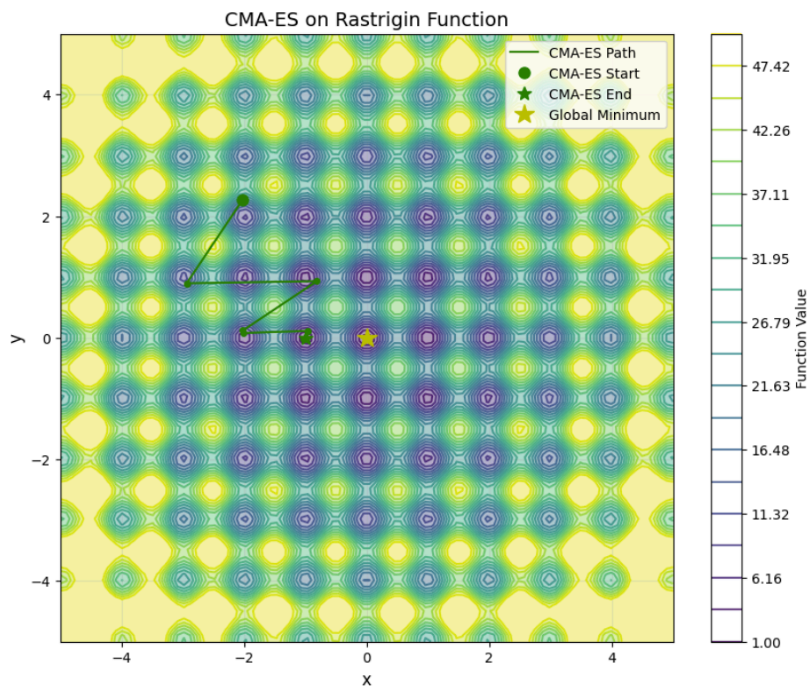

CMA-ES Trajectory on Rosenbrock Function (3D)
CMA-ES
Global Minimum
x
y
Function Value

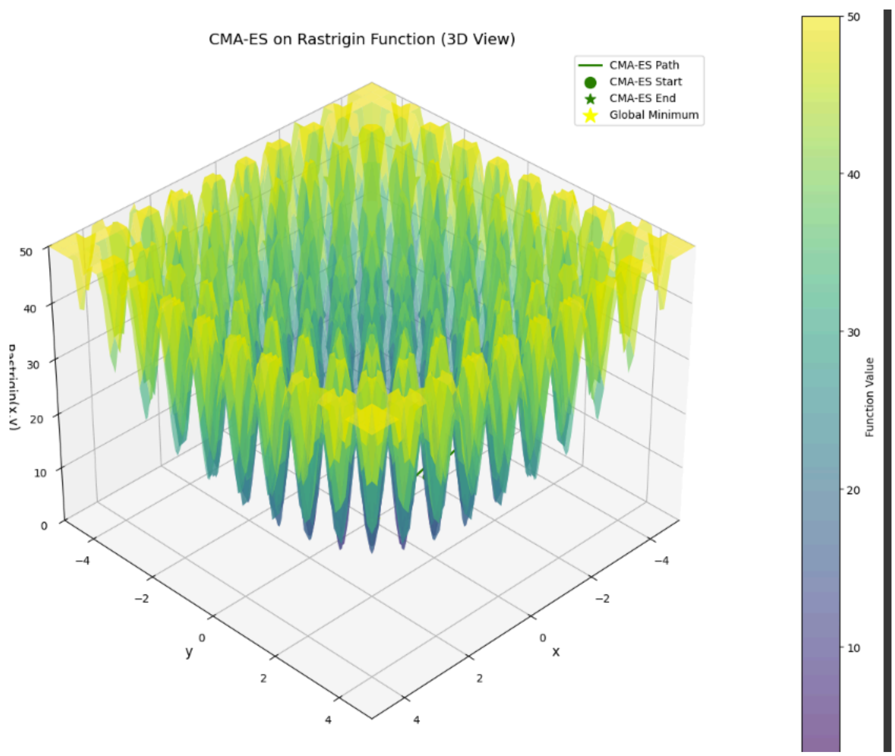

Nelder-Mead on Rastrigin Function
Nelder-Mead Path
Nelder-Mead Start
Nelder-Mead End
Global Minimum
x
y
Function Value

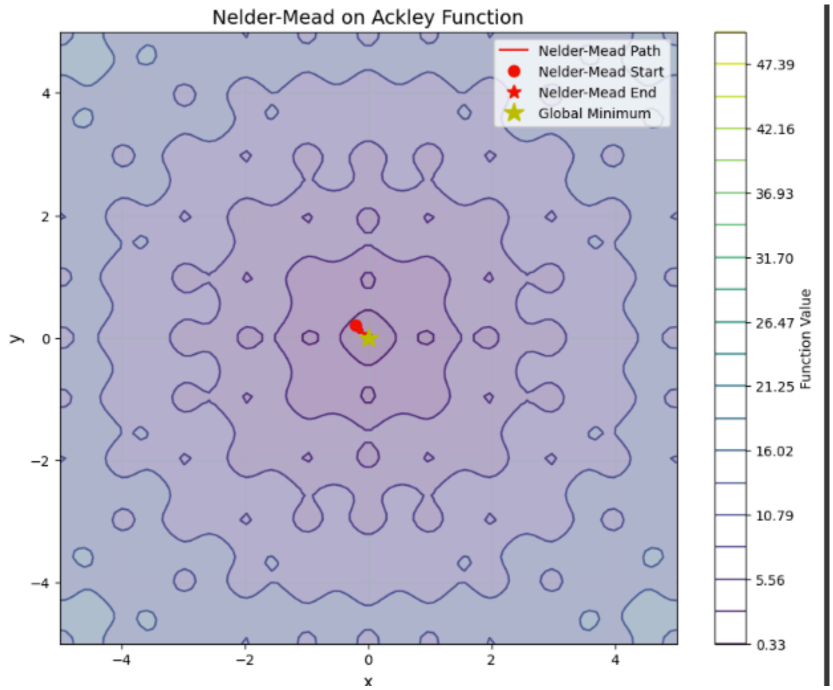

Nelder-Mead on Rastrigin Function (3D View)
Nelder-Mead Path
Nelder-Mead Start
Nelder-Mead End
Global Minimum
x
y
Rastrigin(x,y)
Function Value

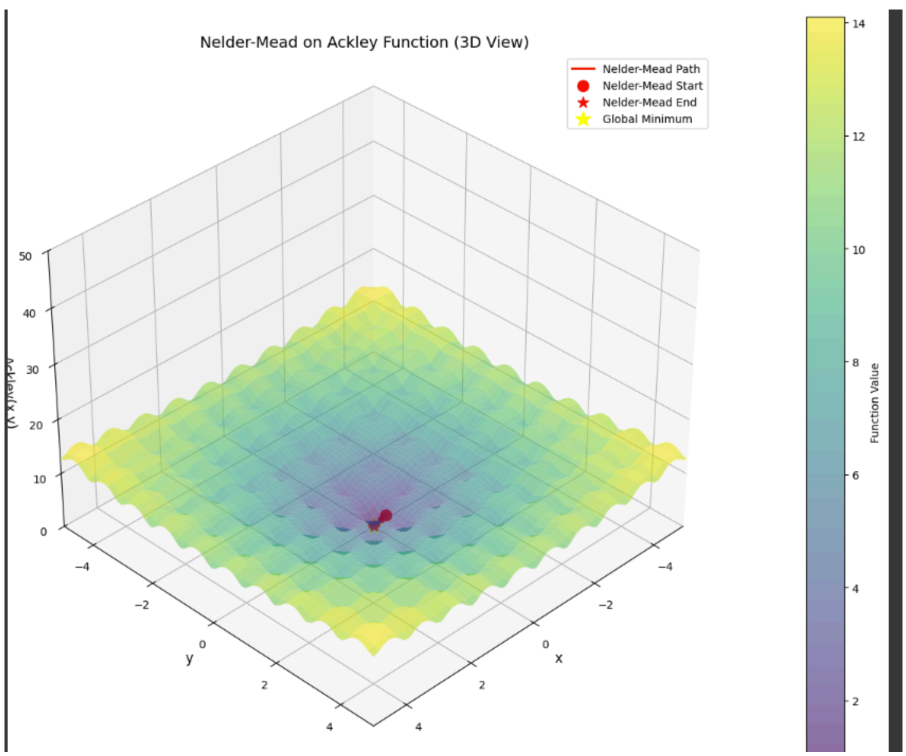

Simulated Annealing on Rastrigin Function
Simulated Annealing Path
Simulated Annealing Start
Simulated Annealing End
Global Minimum
x
y
Function Value
47.42
42.26
37.11
31.95
26.79
21.63
16.48
11.32
6.16
1.00

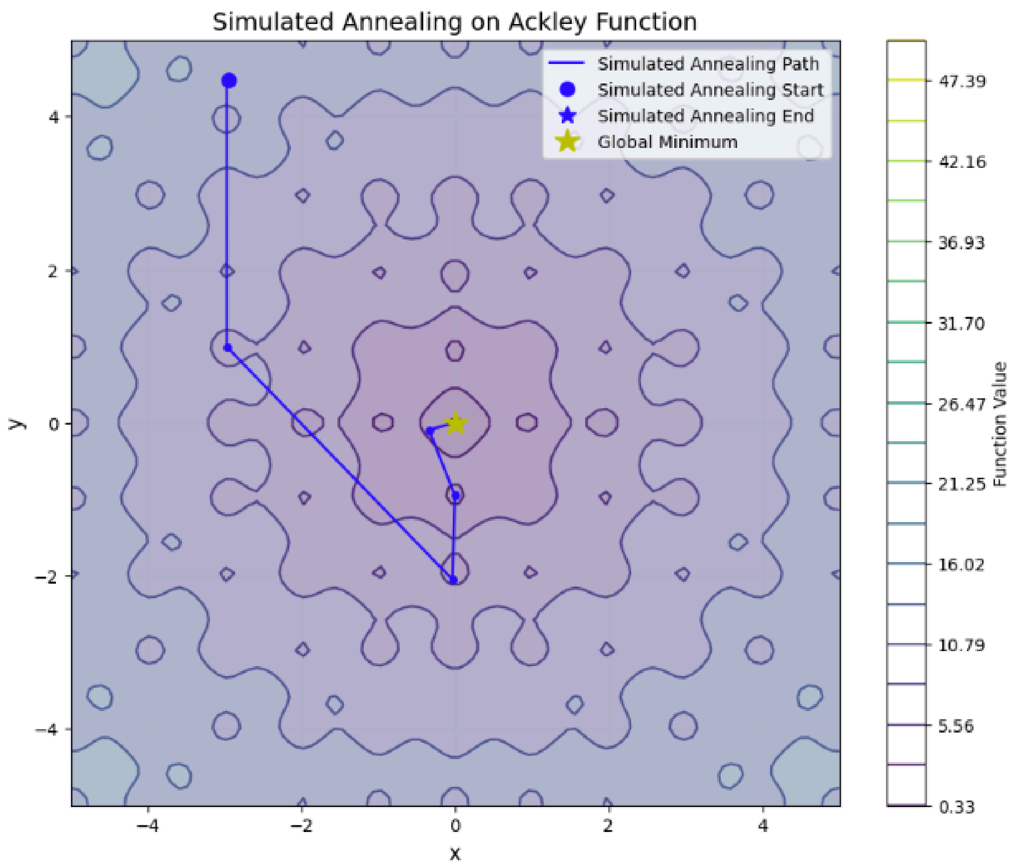

Simulated Annealing on Rastrigin Function (3D View)
Simulated Annealing Path
Simulated Annealing Start
Simulated Annealing End
Global Minimum
x
y
Function Value

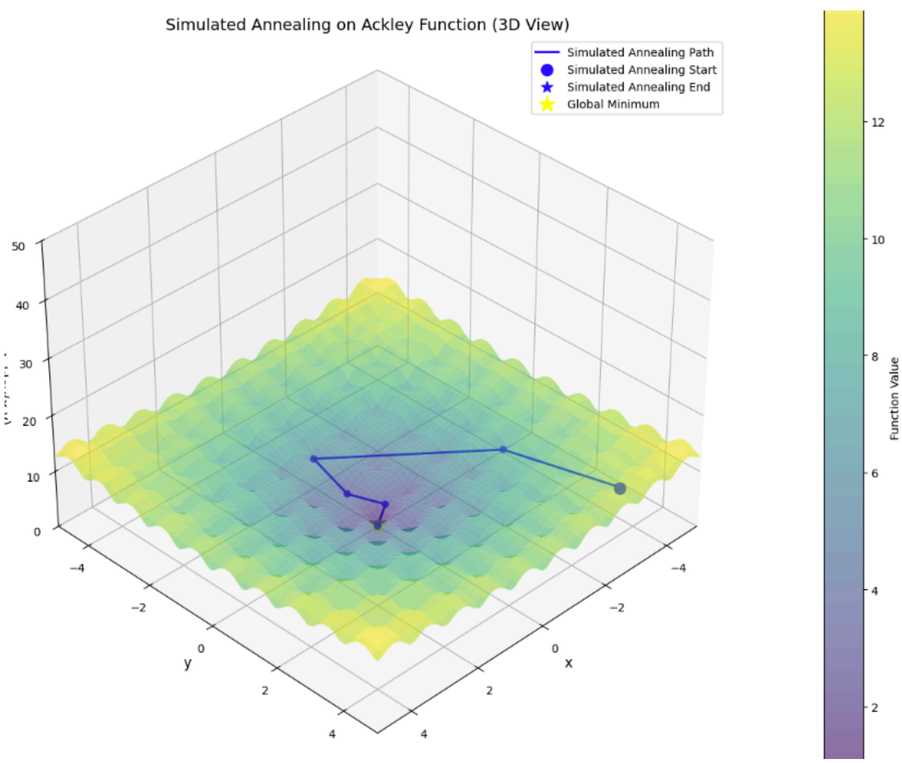

CMA-ES on Rastrigin Function
CMA-ES Path
CMA-ES Start
CMA-ES End
Global Minimum
x
y
Function Value
47.42
42.26
37.11
31.95
26.79
21.63
16.48
11.32
6.16
1.00

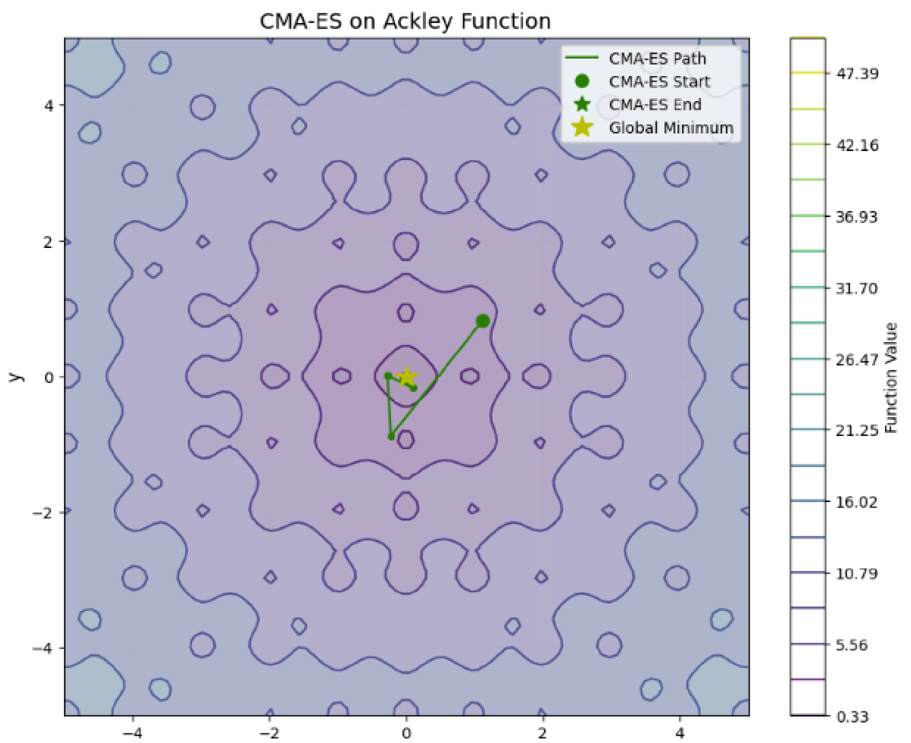

CMA-ES on Rastrigin Function (3D View)
CMA-ES Path
CMA-ES Start
CMA-ES End
Global Minimum
x
y
Function Value

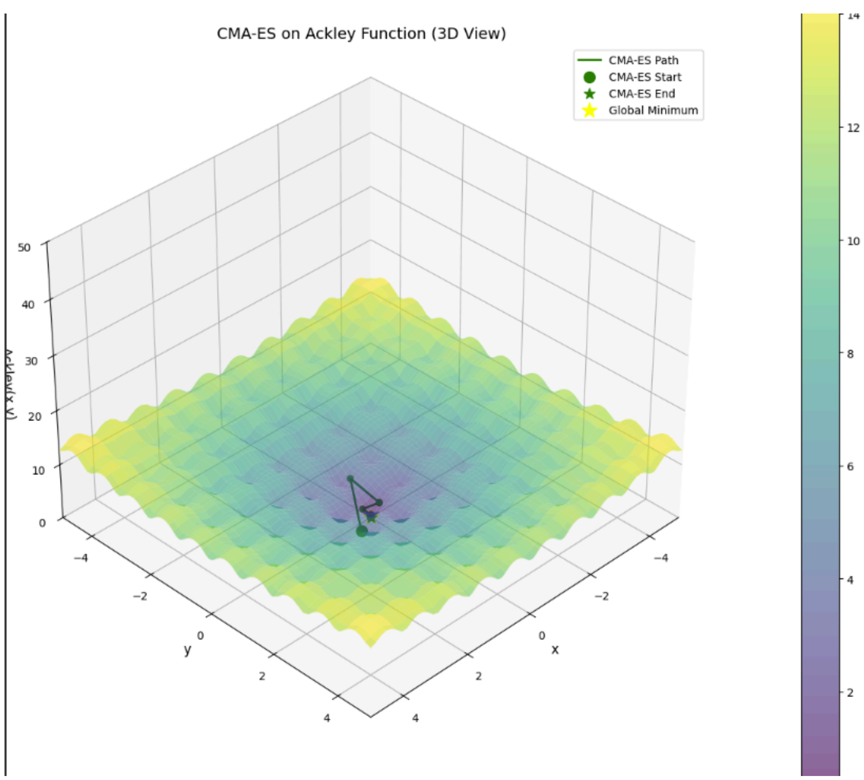

Nelder-Mead on Ackley Function
Nelder-Mead Path
Nelder-Mead Start
Nelder-Mead End
Global Minimum
x
y
Function Value
47.39
42.16
36.93
31.70
26.47
21.25
16.02
10.79
5.56
0.33


Nelder-Mead on Ackley Function (3D View)
Nelder-Mead Path
Nelder-Mead Start
Nelder-Mead End
Global Minimum
x
y
Function Value


Simulated Annealing on Ackley Function
Simulated Annealing Path
Simulated Annealing Start
Simulated Annealing End
Global Minimum
x
y
Function Value
47.39
42.16
36.93
31.70
26.47
21.25
16.02
10.79
5.56
0.33


Simulated Annealing on Ackley Function (3D View)
Simulated Annealing Path
Simulated Annealing Start
Simulated Annealing End
Global Minimum
x
y
Function Value


CMA-ES on Ackley Function
CMA-ES Path
CMA-ES Start
CMA-ES End
Global Minimum
y
4
2
0
−2
−4
−4
−2
0
2
4
Function Value
47.39
42.16
36.93
31.70
26.47
21.25
16.02
10.79
5.56
0.33


CMA-ES on Ackley Function (3D View)
CMA-ES Path
CMA-ES Start
CMA-ES End
Global Minimum
50
40
30
20
10
0
y
x
−4
−2
0
2
4
−4
−2
0
2
4
Function Value
14
12
10
8
6
4
2

Trade-offs of each method.

Nelder Mead
Advantages:
- Works well for non-differentiable functions.
- Performs well on low-dimensional problems.

Disadvantages:
- Inefficient for high-dimensional problems.
- Can get stuck in a local minima.
- It's very sensitive to the initial guess, it did not converge when for the Ackely function when the initial guess was (-1, 1) but did so when it was (-0.2, 0.2).
- If the function is very noisy and rapidly changes, nelder mead finds it difficult to find the optimal solution.

Simulated Annealing
Advantages:
- Works well for non-differentiable functions.
- It can escape local minima, which nelder mead generally struggles with.
- Robust in non-smooth environments, and works well even in noisy environments.
- Good for black-box functions.
- Performs well on low-dimensional problems.

Disadvantages:
- Slow convergence, it takes a lot of iterations to converge as seen in the convergence speeds plots above.
- Performance degrades for very large search spaces, not efficient in higher dimensional problems.

CMA-ES
Advantages:
- Works well for non-differentiable functions.
- Works well even in high dimensional space, which the above two optimization methods failed.
- Robust in non-smooth environments, and works well even in noisy environments.
- Good for black-box functions.

Disadvantages:
- Slow convergence, it takes a lot of iterations to converge as seen in the convergence speeds plots above.
- Stores a full covariance matrix, making it costly for very high dimensions.
- It's an overkill for small search spaces.

Final Conclusion

- Use Nelder-Mead if you need a quick, low-dimensional local optimization and the given function is not very complex.

- Use Simulated Annealing if there are a lot of minimas and we want to find the global minima, even though it takes time, but it eventually reaches the optimal solution.

- If we are working with higher dimensional data and we have the memory capacity(to store the covariance matrix) we can use CMA-ES.