VIETNAM NATIONAL UNIVERSITY - HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY
FACULTY OF COMPUTER SCIENCE

REPORT ON ASSIGNMENT 2
Differential Evolution (DE) & Cross Entropy Method (CEM)

TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Student: Nguyen Duy Hoang - 22520467
Class: CS410.P21

Contents

1 Introduction

Optimization is a fundamental problem in various scientific and engineering disciplines, where the goal is to find the optimal solution that minimizes a given objective function. In this study, we compare the performance of two population-based metaheuristic optimization algorithms: Differential Evolution (DE) and the Cross-Entropy Method (CEM). Our primary objective is to evaluate the convergence speed, final solution quality, and stability of both methods when applied to a set of well-known benchmark functions.

The benchmark functions used in this study are:

- **Sphere function**: A simple convex function often used as a baseline for optimization methods.
- **Griewank function**: A complex function with multiple local minima, making it challenging for optimization.
- **Rosenbrock function**: A non-convex function with a narrow, curved valley leading to the global minimum.
- **Rastrigin function**: A multimodal function with a large number of local minima, testing the exploration ability of optimization algorithms.
- **Ackley function**: A function with an exponential component, creating a landscape with many local minima and a global optimum.

The experiments were conducted using a range of random seeds from 22520467 to 22520476 to ensure robustness of the results. The global optimal solutions and search space boundaries for each function are given by:

Function	Global optimal solution	Bounds
Sphere	(0,0)	[-5.12, 5.12]
Griewank	(0,0)	[-600, 600]
Rosenbrock	(1,1)	[-5, 10]
Rastrigin	(0,0)	[-5.12, 5.12]
Ackley	(0,0)	[-32.768, 32.768]

Table 1: Global optimal solution of benchmark functions and search space bounds for each function

The Differential Evolution (DE) algorithm is configured as follows:

```
DE(obj_function, dimension, bounds, F_scale=0.7, pop_size, max_evals, cross_pop=0.5, seed,
↪  log_filename)

CEM(obj_function, dimension, bounds, pop_size, max_evals, elite_ratio=0.5, seed, log_filename)
```

Where:

- `obj_function`: Objective function (to minimize).
- `dimension`: The number of variables defining each individual within the solution space.
- `bounds`: Search space limits.
- `F_scale`: Mutation factor (default = 0.7).
- `pop_size`: Population size.
- `max_evals`: Max function evaluations.
- `cross_pop`: Crossover rate (default = 0.5). Proportion of top solutions used for sampling new candidates in CEM (default = 0.5)
- `seed`: Random seed
- `log_filename`: Output log file.

Both DE and CEM are executed for various population sizes and max evaluations to compare their efficiency in solving the benchmark problems.

2 Results and Analysis

Significance symbols: $**$ ($p < 0.01$), $*$ ($p < 0.05$), NS ($p \geq 0.05$)

2.1 Sphere Function Performance Analysis

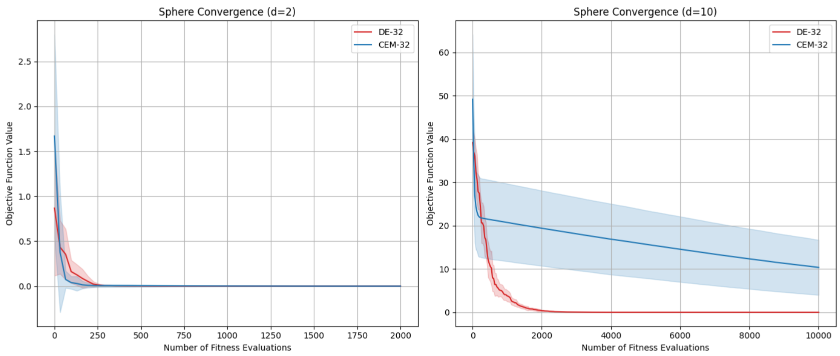

Figure 1: Convergence curve of DE vs CEM on Sphere Function

Population Size	DE (Mean ± Std)	CEM (Mean ± Std)	p-value	Significance
Sphere Function (Dimension = 2)				
8	**2.248e-48 ± 7.27e-48**	3.198 ± 4.92	0.0829	NS
16	**1.976e-24 ± 5.33e-24**	0.44 ± 0.693	0.0892	NS
32	**1.759e-12 ± 2.86e-12**	2.179e-08 ± 1.88e-08	0.0069	**
64	2.146e-07 ± 1.74e-07	**1.63e-08 ± 1.227e-08**	0.0077	**
128	0.00021 ± 9.069e-05	**3.718e-08 ± 2.241e-08**	6.1e-05	**
Sphere Function (Dimension = 10)				
8	**5.55e-18 ± 1.665e-17**	10.698 ± 8.878	0.0056	**
16	**2.16e-19 ± 3.97e-19**	12.27 ± 7.865	0.0011	**
32	**3.789e-09 ± 2.198e-09**	10.355 ± 6.34	0.0008	**
64	**0.00038 ± 0.000181**	7.887 ± 6.307	0.0045	**
128	**0.09076 ± 0.0394**	4.51 ± 6.057	0.056	NS

Table 2: Comparative Performance of DE and CEM on Sphere Function

The convergence curves presented in Figure 1 illustrate the optimization progress of DE and CEM on the Sphere function for a population size of 32, with dimensions d = 2 and d = 10. The graphs show that at d = 2, both algorithms exhibit similar convergence rates. However, at d = 10, the convergence rate of CEM is significantly slower, indicating that DE performs better in higher-dimensional spaces.

Table 2 presents the mean and standard deviation of the objective function values obtained by DE and CEM across various population sizes.. For d = 2, while DE shows excellent performance, achieving near-zero mean values, statistical significance (p-value) is only observed at larger population sizes (32, 64, 128). Notably, at population sizes 64 and 128, CEM outperforms DE, achieving better optimization results. Conversely, for d = 10, DE consistently outperforms CEM across all population sizes, with statistically significant differences (p < 0.01 or p < 0.05) observed in most cases. This aligns with the convergence curves, confirming DE's effectiveness in higher-dimensional optimization.

In summary, DE demonstrates a more robust and efficient optimization capability on the Sphere function. While DE generally converges faster and achieves better results, CEM performs better in certain cases, particularly in d = 2 with larger population sizes (64 and 128). However, in higher dimensions (d = 10), DE consistently outperforms CEM, maintaining rapid convergence and achieving superior optimization results.

2.2 Rosenbrock Function Performance Analysis

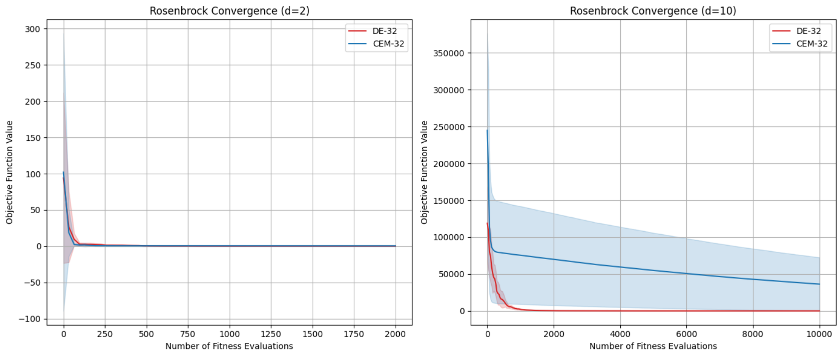

Figure 2: Convergence curve of DE vs CEM on Rosenbrock Function

Popsize N	DE (Mean ± Std)	CEM (Mean ± Std)	p-value	Significance
Rosenbrock (d=2)				
8	**0.000112 ± 0.000305**	3868.638 ± 9126.37	0.2353	NS
16	**0.00528 ± 0.0146**	2.52 ± 4.91	0.1587	NS
32	**0.0446 ± 0.0699**	0.408 ± 0.7348	0.173	NS
64	**0.0634 ± 0.0399**	0.0966 ± 0.1317	0.4849	NS
128	0.2224 ± 0.1663	**0.0215 ± 0.0281**	0.0054	**
Rosenbrock (d=10)				
8	**8.94 ± 8.388**	48448.05 ± 40512.915	0.0058	**
16	**4.569 ± 0.701**	35809.176 ± 31718.135	0.008	**
32	**6.448 ± 0.3558**	36305.61 ± 36298.525	0.0149	*
64	**30.067 ± 13.829**	16771.745 ± 21906.32	0.0475	*
128	**117.719 ± 42.136**	3342.27 ± 4884.476	0.079	NS

Table 3: Comparative Performance of DE and CEM on Rosenbrock Function

The convergence curves presented in Figure 2 illustrate the optimization progress of DE and CEM on the Rosenbrock function for a population size of 32, with dimensions d = 2 and d = 10. For d = 2, both DE and CEM exhibit similar convergence behaviors, reaching near-zero objective values after a small number of evaluations. However, DE shows slightly better consistency, as indicated by its lower standard deviation. For d = 10, DE significantly outperforms CEM, achieving much faster convergence and lower objective values. The CEM method struggles to minimize the function, as evidenced by its high objective values and large variance throughout the optimization process. This suggests that DE is more effective in handling the complexity of higher-dimensional Rosenbrock optimization.

Table 3 presents the mean and standard deviation of the objective function values obtained by DE and CEM across various population sizes. For d = 2, DE achieves the best results for most population sizes, except for N = 128, where CEM performs better. However, the differences in performance are not always statistically significant ($p > 0.05$ in most cases). For d = 10, DE consistently outperforms CEM in all population sizes, with statistically significant differences ($p < 0.05$) in most cases, reinforcing the observation from the convergence curves.

In summary, the Rosenbrock function, with its challenging narrow valley landscape, reveals a clear advantage for the DE algorithm, particularly in higher dimensions. DE's overall robustness and efficiency across varying dimensions and population sizes underscore its suitability for demanding optimization tasks.

2.3 Griewank Function Performance Analysis

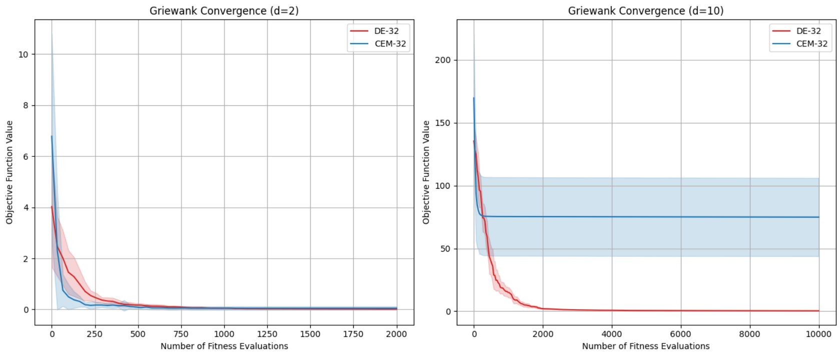

Figure 3: Convergence curve of DE vs CEM on Griewank Function

Popsize N	DE (Mean ± Std)	CEM (Mean ± Std)	p-value	Significance
Griewank (d=2)				
8	**0.0115 ± 0.0129**	20.12 ± 26.04	0.045	*
16	**0.00155 ± 0.00292**	3.28 ± 4.56	0.059	NS
32	**0.0132 ± 0.00671**	0.0466 ± 0.0728	0.202	NS
64	0.0366 ± 0.0240	**0.00935 ± 0.0151**	0.011	*
128	0.0962 ± 0.0471	**0.01146 ± 0.00596**	0.0004	**
Griewank (d=10)				
8	**0.04308 ± 0.0338**	205.092 ± 74.292	1.7e-05	**
16	**0.106 ± 0.0735**	127.156 ± 49.967	3.2e-05	**
32	**0.3578 ± 0.0522**	74.907 ± 31.164	5.2e-05	**
64	**0.5776± 0.1068**	43.732 ± 28.507	0.001	**
128	**1.259 ± 0.0569**	20.863 ± 23.882	0.036	*

Table 4: Comparative Performance of DE and CEM on Griewank Function

The convergence curves in Figure 3 illustrate the optimization progress of DE and CEM on the Griewank function for a population size of 32, with dimensions d = 2 and d = 10. For d = 2, both DE and CEM exhibit rapid convergence towards the optimal solution. However, DE demonstrates marginally better consistency, as shown by its slightly lower variability. For d = 10, DE shows a distinct advantage in convergence speed and solution quality, while CEM appears to be trapped in local minima, evidenced by its consistently higher objective values and significant variance.

Table 4 presents the mean and standard deviation of the objective function values obtained by DE and CEM across various population sizes. For d = 2, DE achieves better mean values at smaller population sizes (8 and 16), with a statistically significant difference at N = 8. Conversely, at larger population sizes (64 and 128), CEM achieves better mean values and statistically significant differences. For d = 10, DE consistently outperforms CEM across all population sizes, with highly statistically significant differences ($p < 0.05$).

In summary, the Griewank function, characterized by its numerous, regularly distributed local minima, presents a challenging optimization landscape. DE demonstrates strong optimization capabilities, particularly in higher dimensions (d = 10), where it significantly outperforms CEM. In lower dimensions (d = 2), the performance difference is less pronounced and dependent on population size; DE shows slight advantages at smaller populations, while CEM excels at larger ones. This suggests that the choice of algorithm, and potentially the appropriate population size, is critical and depends on the specific dimensionality of the Griewank function being optimized. DE's ability to effectively explore and navigate the complex, multi-modal search space of the Griewank function highlights its robustness, especially as dimensionality increases.

2.4 Rastrigin Function Performance Analysis

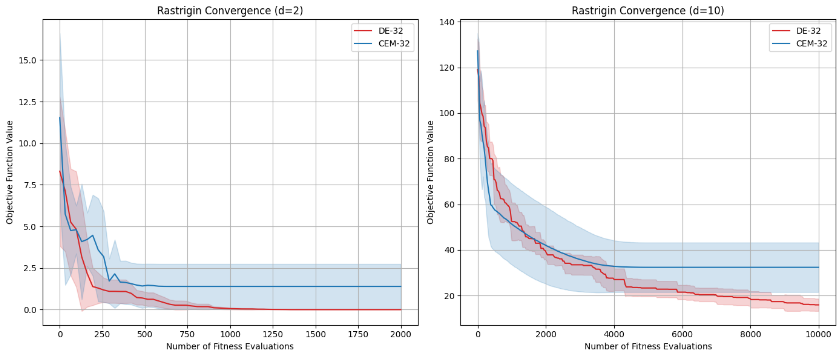

Figure 4: Convergence curve of DE vs CEM on Rastrigin Function

Popsize N	DE (Mean ± Std)	CEM (Mean ± Std)	p-value	Significance
Rastrigin (d=2)				
8	**1.09 ± 1.13**	9.00 ± 9.38	0.032	*
16	**0.0 ± 0.0**	3.40 ± 3.60	0.019	*
32	**1.335e-06 ± 2.76e-06**	1.24 ± 1.179	0.0116	*
64	**0.0618 ± 0.107**	0.4085 ± 0.501	0.070	NS
128	0.359 ± 0.3235	**0.208 ± 0.4035**	0.392	NS
Rastrigin (d=10)				
8	**5.16 ± 2.17**	75.62 ± 22.07	5e-06	**
16	**4.90 ± 3.21**	49.85 ± 15.00	6e-06	**
32	**15.898 ± 2.717**	32.338 ± 10.87	0.0012	**
64	24.12 ± 2.867	**23.186 ± 6.73**	0.708	NS
128	30.30 ± 5.564	**14.43 ± 7.409**	8.7e-05	**

Table 5: Comparative Performance of DE and CEM on Rastrigin Function

The convergence curves in Figure 4 illustrate the optimization progress of DE and CEM on the Rastrigin function for a population size of 32, with dimensions d = 2 and d = 10. For d = 2, while both algorithms show initial oscillations due to the function's multi-modal nature, DE demonstrates superior consistency and convergence to near-zero objective values, unlike CEM which appears trapped in a local minimum. In d = 10, DE significantly outperforms CEM, exhibiting faster convergence and lower objective values, while CEM's performance is hindered by high objective values and large variance, indicating its susceptibility to local minima and difficulty in exploring the high-dimensional search space.

Table 5 presents the mean and standard deviation of objective function values obtained by DE and CEM across various population sizes. For d = 2, DE achieves better mean values at smaller population sizes (8, 16, 32), with statistically significant differences ($p < 0.05$). At larger population sizes (64, 128), the performance difference is less pronounced, with no statistically significant difference at N=64 and CEM performing slightly better at N=128. For d = 10, DE consistently outperforms CEM at smaller population sizes (8, 16, 32), with highly statistically significant differences ($p < 0.01$). However, at larger population sizes (64, 128), CEM performs better, with a statistically significant difference at N=128.

In summary, the Rastrigin function, characterized by its numerous local minima, presents a significant optimization challenge. While DE demonstrates strong performance at smaller population sizes, particularly in higher dimensions (d=10), CEM exhibits competitive performance at larger population sizes (64, 128).

2.5 Ackley Function Performance Analysis

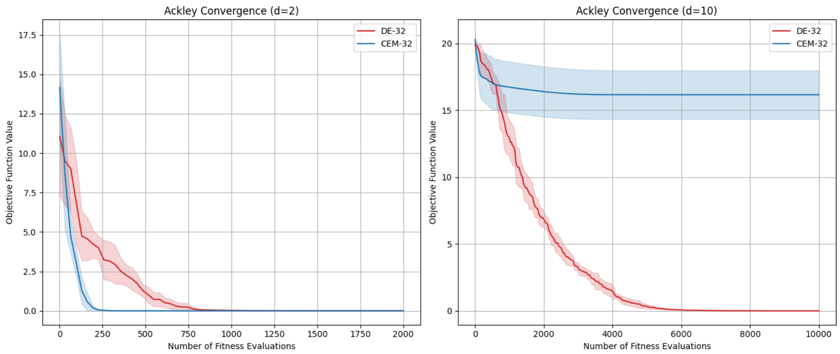

Figure 5: Convergence curve of DE vs CEM on Ackley Function

Popsize N	DE (Mean ± Std)	CEM (Mean ± Std)	p-value	Significance
Ackley (d=2)				
8	**8.615e-15 ± 2.335e-14**	12.15 ± 6.187	0.0002	**
16	**4.452e-11 ± 6.843e-11**	5.347 ± 6.815	0.043	*
32	**1.515e-05 ± 1.517e-05**	0.000287 ± 0.000187	0.0018	**
64	0.00828 ± 0.00461	**0.000381 ± 0.000152**	0.0006	**
128	0.749 ± 0.371	**0.000471 ± 0.000203**	0.00019	**
Ackley (d=10)				
8	**0.8849 ± 1.585**	19.139 ± 0.507	0	**
16	**1.64e-09 ± 1.27e-09**	18.225 ± 0.634	0	**
32	**0.000597 ± 0.000158**	16.168 ± 1.816	0	**
64	**0.318 ± 0.0867**	13.441 ± 2.79	0	**
128	**4.39 ± 0.609**	8.725 ± 3.985	0.098	**

Table 6: Comparative Performance of DE and CEM on Ackley Function

The Ackley function, with its nearly flat outer region and numerous local minima surrounding a central hole, presents a challenging optimization landscape. The convergence curves in Figure 5 illustrate the optimization progress of DE and CEM on the Ackley function for a population size of 32, with dimensions d = 2 and d = 10. For d = 2, both algorithms show a similar initial descent. However, for d = 2, CEM demonstrates a faster initial convergence, reaching a lower objective value with fewer evaluations compared to DE. For d = 10, DE demonstrates a clear superiority, achieving convergence around 6000 evaluations, whereas CEM becomes trapped in local minima early on, maintaining high objective values and large variance, highlighting its inability to escape the complex landscape in the high-dimensional space.

Table 6 presents the mean and standard deviation of objective function values obtained by DE and CEM across various population sizes. For d = 2, DE achieves significantly better mean values at smaller population sizes (8, 16, 32), with statistically significant differences ($p < 0.05$). However, at larger population sizes (64, 128), CEM achieves better mean values, suggesting it benefits from larger population diversity in navigating the function's complex landscape. For d = 10, DE consistently outperforms CEM across all population sizes, with highly statistically significant differences ($p < 0.01$), indicating DE's robustness in handling the increased dimensionality.

In summary, DE demonstrates a strong ability to navigate the complex Ackley function, particularly in high-dimensional spaces (d=10), where it significantly outperforms CEM. In lower dimensions (d=2), while DE shows advantages at smaller population sizes, CEM exhibits better performance at larger population sizes. This suggests that DE is more effective in escaping local minima and finding the global optimum, especially as dimensionality increases. However, CEM's performance improves with larger population sizes, potentially due to enhanced exploration capabilities in the search space.

3 Conclusion

In this assignment, we compared the performance of two optimization algorithms, DE and CEM, on benchmark functions. Based on the experimental results, we observed that DE generally exhibits better performance than CEM with smaller population sizes and lower dimensions (d = 2). However, as the population size increases, CEM's performance improves and becomes competitive with DE. In higher dimensions (d=10), DE consistently outperforms CEM.

Specifically, on benchmark functions with numerous local minima, CEM, due to its random initialization based on mean and covariance matrix updates from previous generations, tends to get trapped in local minima, particularly with smaller population sizes. In contrast, DE, with its crossover and mutation operators, demonstrates a superior ability to escape these local minima.

For the cases where d=2, animation GIFs illustrating the population movement across generations, along with the function's contour, were created using a random seed of 22520467 and a population size of N=32. These GIFs, which also mark the global optimum to visualize the population's proximity to the solution after 5,000 function evaluations, are available at the following Google Drive Link. The log files used to generate these GIFs, containing detailed information about the best solutions and objective function values across generations, are also available at the following Google Drive Link.

Additionally, log files containing the best solution and objective function values (DE_{best_i} fitness(best_i) and CEM_{best_i} fitness(best_i)) found by DE and CEM at each generation, along with the cumulative number of function evaluations ($\text{num\_of\_evaluations}_i$), are also available at the following Google Drive Link.