Comparing Differential Evolution to Classical and Evolutionary Optimization

STAT 540 Project

Adaku Uchendu

Problem Definition

- **Main Objective**

For an objective function $f : X \subset \mathbb{R}^D \to \mathbb{R}$ where the feasible region $X \neq \emptyset$, the minimization problem is to find $x* \in X$ such that $f(x*) \leq f(x) \quad \forall x \in X$ where $f(x*) \neq -\infty$.

- **Data Description**
    - Car MSRP Kaggle dataset
    - Dimensions: 8084 observations x 16 features
    - Used only 8 Features: Number of doors, Engine HP, highway MPG, city mpg, Engine cylinders, Year, Popularity
    - MSRP group: *Ordinary, Deluxe, Super-deluxe, Luxury, Super-luxury*
- **Regression model**

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_7 x_7 + \epsilon$$

Optimization Algorithms

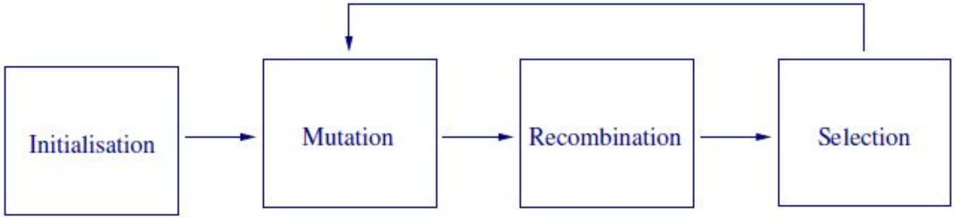
- Classical Algorithms
  - Gradient Descent: *first-order*
  - Stochastic Gradient Descent: *first-order*
  - Newton-Raphson: *second-order*
  - Quasi-Newton (‘BFGS’): *second-order*
- Evolutionary Algorithms
  - Particle Swarm Optimization: *direct search or zero-order*
  - **Differential Evolution:** *direct search or zero-order*

Differential Evolution

Generate initial population $P^0 = \{\vec{x}_1^0, \vec{x}_2^0, \ldots, \vec{x}_N^0\}$
Let $t = 0$
repeat
for each individual $\vec{x}_i^t$ in the population P^t **do**
Generate three random integers r_1, r_2 and
$r_3 \in \{1,2,\ldots,N\} \setminus i$, with $r_1 \neq r_2 \neq r_3$
Generate a random integer $j_{rand} \in \{1,2,\ldots,D\}$
for each parameter j **do**

$$u_{i,j}^{t+1} = \begin{cases} x_{r_3,j}^t + F \times (x_{r_1,j}^t - x_{r_2,j}^t), \\ \quad \text{if } (rand \leq CR \,||\, j = rand[1, D]) \\ x_{i,j}^t \quad , \text{otherwise} \end{cases}$$

end for
Replace $\vec{x}_i^t$ with the child $\vec{u}_i^{t+1}$ in the population P^{t+1},
if $\vec{u}_i^{t+1}$ is better, otherwise $\vec{x}_i^t$ is retained
end for
$t = t + 1$
until the termination condition is achieved



Initialization and Mutation

- **Initialization:** Define upper and lower bounds for each parameter:

$$x_j^L \leq x_{j,i,1} \leq x_j^U$$

  Then randomly select the initial parameter values uniformly on the intervals $[x_j^L, x_j^U]$.

- **Mutation:** For a given parameter $x_{i,G}$ randomly select 3 vectors - $x_{r1,G}$, $x_{r2,G}$, and $x_{r3,G}$, such that $i, r1, r2$, and $r3$ are distinct. Next, perform

$$v_{i,G+1} = x_{r1,G} + F(x_{r2,G} - x_{r3,G})$$

  where mutation factor $F \in [0, 2]$ and $v_{i,G+1}$ is called the donor vector.

Recombination and Selection

- **Recombination:** The trial vector $u_{i,G+1}$ is developed from the elements of the target vector, $x_{i,G}$, and the elements of the donor vector, $v_{i,G+1}$. Elements of the donor vector enter the trial vector with probability CR;

$$v_{j,i,G+1} = \begin{cases} v_{j,i,G+1} & if\ rand_{j,i} \leq CR\ or\ j = I_{rand} \\ x_{j,i,G} & if\ rand_{j,i} > CR\ and\ j \neq I_{rand} \end{cases}$$

  where $rand_{i,j}\ U[0,1], I_{rand}$ is a random integer from $[1, \ldots, D]$ and I_{rand} ensures that $v_{i,G+1} \neq x_{i,G}$.

- **Selection:** The target vector, $x_{i,G}$ is compared to the trial vector, $v_{i,G+1}$ and the one with the lowest function value is admitted to the next generation:

$$x_{i,G+1} = \begin{cases} u_{i,G+1} & if\ f(u_{i,G+1}) \leq f(x_{i,G}),\ i = 1, 2, \ldots, N \\ x_{i,G} & otherwise \end{cases}$$

Variants of Differential Evolution

Variant	**Mathematical Formulation**
Best/1/Exp	$x_{i,j,G+1} = best_{j,G} + F.(x_{r1,j,G} - x_{r2,j,G})$
Rand/1/Exp	$x_{i,j,G+1} = x_{r1,j,G} + F.(x_{r2,j,G} - x_{r3,j,G})$
RandToBest/1/Exp	$x_{i,j,G+1} = x_{i,j,G} + F.(best_{i,G} - x_{i,j,G}) + F.(x_{r1,j,G} - x_{r2,j,G})$
Best/2/Exp	$x_{i,j,G+1} = best_{i,G} + F.(x_{r1,i,G} + x_{r2,i,G} - x_{r3,i,G} - x_{r4,j,G})$
Rand/2/Exp	$x_{i,j,G+1} = x_{r1,i,G} + F.(x_{r2,i,G} + x_{r3,i,G} - x_{r4,i,G} - x_{r5,i,G})$
Best/1/Bin	$x_{j,i,G+1} = best_{i,G} + F.(x_{r1,i,G} - x_{r2,i,G})$
Rand/1/Bin	$x_{j,i,G+1} = x_{r1,j,G} + F.(x_{r2,j,G} - x_{r3,j,G})$
RandToBest/1/Bin	$x_{j,i,G+1} = x_{i,j,G} + F.(best_{i,G} - x_{i,j,G}) + F.(x_{r1,j,G} - x_{r2,j,G})$
Best/2/Bin	$x_{j,i,G+1} = best_{i,G} + F.(x_{r1,i,G} + x_{r2,i,G} - x_{r3,i,G} - x_{r4,i,G})$
Rand/2/Bin	$x_{j,i,G+1} = x_{r1,i,G} + F.(x_{r2,i,G} + x_{r3,i,G} - x_{r4,i,G} - x_{r5,i,G})$

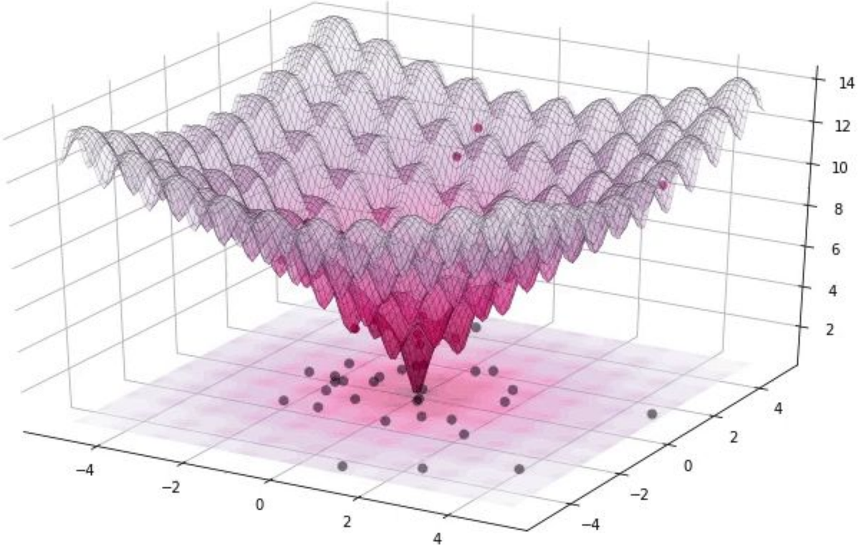
Table 1: Differential Evolution variants

DE with Ackley's function

$$f(x_0 \cdots x_n) = -20exp(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}) - exp(\frac{1}{n}\sum_{i=1}^{n} cos(2\pi x_i)) + 20 + e$$

$$-32 \leq x_i \leq 32$$

$$\text{minimum at } f(0, \cdots, 0) = 0$$



Results

f	Function	Definition	Bound	Global minimum
f_1	Rosenbrock	$F(\vec{x_i}) = \sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2$	$-15 \leq x_i \leq 15$	$x_i = [1.0, 1.0]$
f_2	Ackley	$F(\vec{x_i}) = -20 \exp(-0.2\sqrt{\frac{1}{n} \sum_{i=1}^{n} x_i^2}) - \exp(\frac{1}{n} \sum_{i=1}^{n} \cos 2\pi x_i) + 20 + e$	$-32 \leq x_i \leq 32$	$x_i = [0, 0]$
f_3	Sphere	$F(\vec{x_i}) = \sum_{i=1}^{n} x_i^2$	$-100 \leq x_i \leq 100$	$x_i = \vec{0}$

Table 2: Test Objective functions

Algorithm	Iteration	f_1	f_2	f_3
DE	100	0.434	$5.610e^{-12}$	$4.326e^{-25}$
PSO	100	0.325	$4.441e^{-16}$	$1.706e^{-21}$
GD	100	0.132	$4.501e^{-05}$	$4.315e^{-05}$
SGD	100	0.123	$1.965e^{-06}$	$2.167e^{-05}$
QN	100	0.999	$3.911e^{-10}$	$2.885e^{-07}$
NR	100	0.999	$-7.943e^{-09}$	$1.961e^{-13}$

Table 3: Results of the Test Objective functions with the Optimization algorithms

	RMSE
Baseline	0.4094
DE	1318.6819
PSO	SR
GD	0.5465
SGD	0.5859
NR	NA
QN	NA

Table 5: Results of the Optimization algorithm with the Linear Regression model

Summary

- One of the most popular Evolutionary algorithm
- Performs well on problems with large dimensions
- Does not guarantee convergence to global minima
- More efficient and accurate than Genetic algorithms
- Applications:
  - Black-box Adversarial attack (i.e. One-Pixel)
  - Design of digital filters
  - Optimization of fermentation of alcohol