Outline

1. **Introduction**
   - Mutant Vector
   - Trial Vector
   - Selection
   - Variants of DE
2. **DE Algorithm**
   - Working Principles Through An Example
   - Initial Population
   - Evaluate Population
   - Generation of Mutant Vectors
   - Generation of Trial Vectors
   - Greedy Selection of DE
   - Graphical Example
3. **Closure**

Introduction

- Proposed by R. Storn and K. Price, "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces", Journal of Global Optimization 11: $341 - 359$, 1997.
- Differential Evolution (DE) draws idea from
  - Population-based evolutionary computation techniques that work on the population of vectors/solutions/individuals.
  - Inherit parallel-search space property for escaping the local minima.
  - Nelder and Mead's simplex search method by employing information within the vector population to alter the search space.
  - In this case, each vector in the population is perturbed by adding the difference of two randomly chosen vectors from the population.
- DE was first proposed by the authors for minimizing nonlinear and non-differentiable continuous space functions.
- As claimed by the authors, DE requires few control variables, is robust, easy to use, and function evaluations can be done in parallel.

Differential Evolution (DE)

- DE starts with a population of vectors or solutions similar to GA and PSO.
- Unlike to GA and PSO, each vector is perturbed by adding the difference of two randomly chosen vectors from the population.

DE works on using three types of vectors

Target vector	Mutant vector	Trial Vector

- **Target vector**: A vector which is being perturbed.
- **Mutant vector**: A vector is generated by adding the weighted difference between two randomly chosen vectors to a third vector from the population.
- **Trial vector**: A vector is generated by mixing the variables of target vector and mutant vector.

Nomenclature of Canonical DE

Parameters and their representations

Parameters	Canonical DE	Generalized representation
Population member/solution	i	i
Population size	NP	N
Generation counter	G	t
Maximum number of generations	G_{max}	T
Target vector i in G−th generation	$x_{i,G}$	$x_i^{(t)}$
Mutant vector i for $(G+1)$−th generation	$v_{i,G+1}$	$v_i^{(t+1)}$
Trial vector i for $(G+1)$−th generation	$u_{i,G+1}$	$u_i^{(t+1)}$
Vector size	D	n
Crossover probability	CR	p_c

- For example, Canonical DE uses $x_{i,G} = (x_{1i,G}, x_{2i,G}, \ldots, x_{Di,G})^T$
- Generalized framework: $x_i^{(t)} = (x_{i1}^{(t)}, \ldots, x_{in}^{(t)})^T$

Mutant Vector using Mutation

Purpose

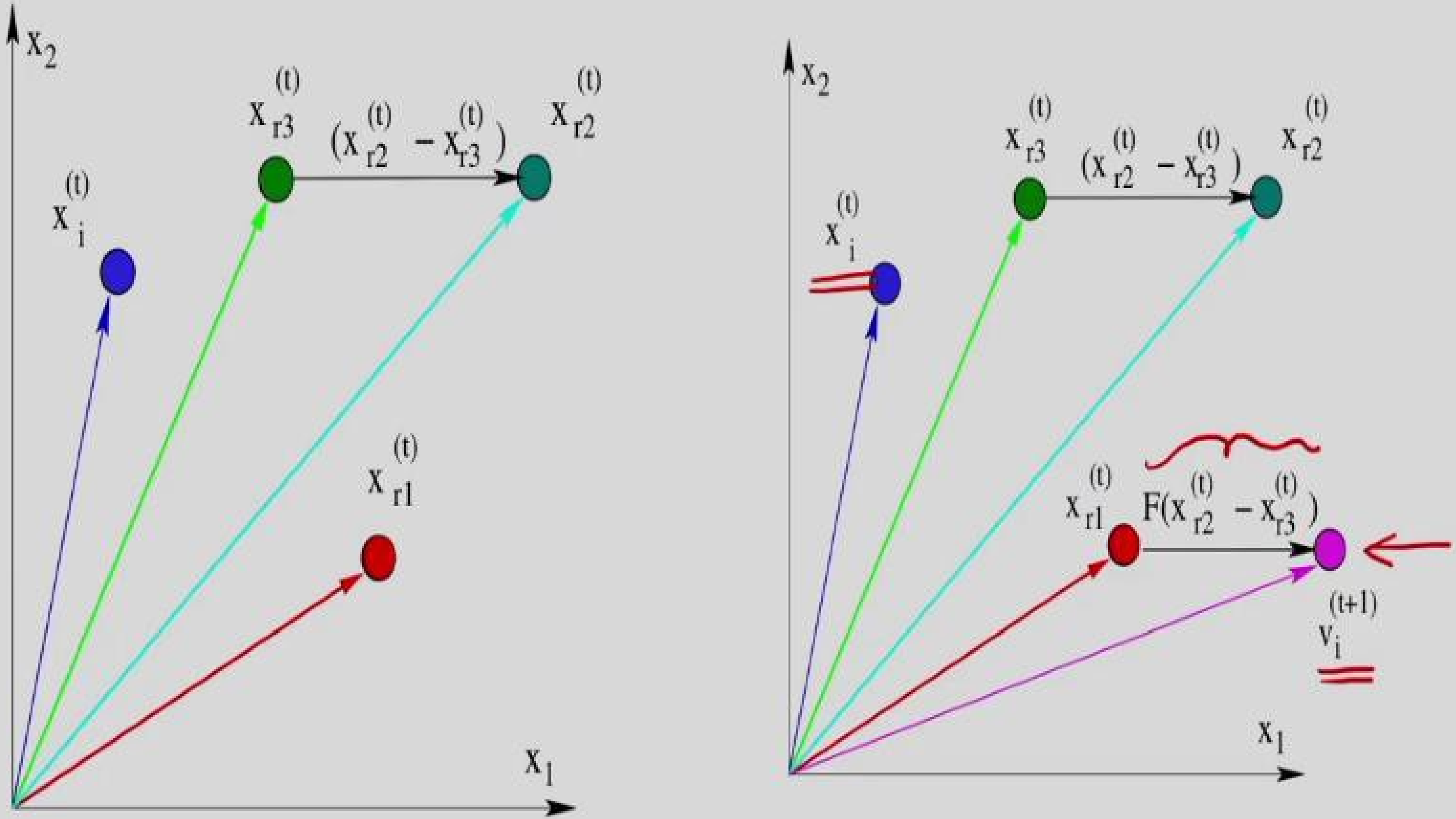
Create a new vector for exploration

- For each target vector $x_i^{(t)}$, a mutant vector $v_i^{(t+1)}$ is generated as

$$v_i^{(t+1)} = x_{r1}^{(t)} + F \times (x_{r2}^{(t)} - x_{r3}^{(t)}), \tag{1}$$

  where $r_1 \neq r_2 \neq r_3$ are randomly chosen vectors from the current population.
- Note: the target vector $i \notin \{r_1, r_2, r_3\}$ and hence, the population size of $N \geq 4$ is at least required for DE.
- The user-defined $F \in [0, 2]$ is a real and constant factor.
  - It controls the amplification of the differential variation $(x_{r2}^{(t)} - x_{r3}^{(t)})$
  - Other implementation imposes different limits on F.
- **Note**: mutation involves more than one vector

Graphical Illustration of Mutation



Trial Vector using Crossover

Purpose

Create new vector for diversity

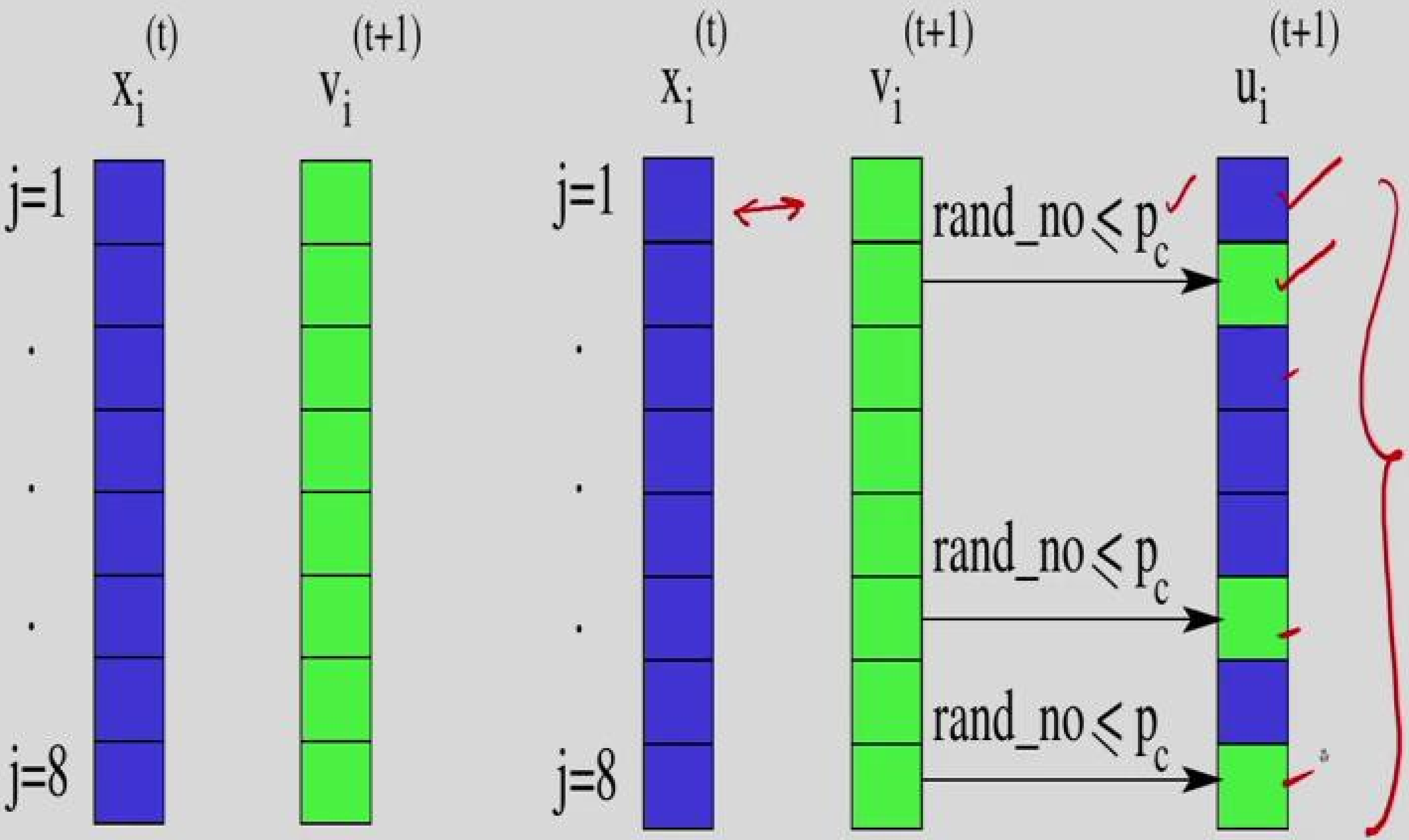
- A trial vector $u_i^{(t+1)}$ is generated as

$$u_{i_j}^{(t+1)} = \begin{cases} v_{i_j}^{(t+1)} & \text{if } (\text{rand\_no} \leq p_c) \text{ or } j = rnbr(i) \\ x_{i_j}^{(t)} & \text{if } (\text{rand\_no} > p_c) \text{ and } j \neq rnbr(i) \end{cases} \tag{2}$$

, where $j \in 1, 2, \ldots, n$

- rand_no is random number $\in [0, 1]$
- p_c: crossover rate or probability of crossover.
- $rnbr(i)$ is randomly chosen index $\in \{1, 2, \ldots, n\}$ which ensures that $u_{i_j}^{(t+1)}$ gets at least one parameter from $v_{i_j}^{(t+1)}$.

Graphical Illustration of Trial Vector



Selection

Purpose

Decide whether a trial vector $u_i^{(t+1)}$ can replace its target vector $x_i^{(t)}$ and enter into the population?

- The fitness of trial vector $u_i^{(t+1)}$ is compared with target vector $x_i^{(t)}$.
- The canonical DE uses greedy criterion for selecting one vector that to be included in the population. Assume minimization problem.

$$x_i^{(t+1)} = \begin{cases} u_i^{(t+1)} & \text{if } F(u_i^{(t+1)}) < F(x_i^{(t)}) \\ x_i^{(t)} & \text{Otherwise} \end{cases} \tag{3}$$

- Other selection criterion such as $(\mu + \lambda)-$selection strategy can be used. Meaning, generate all trial vectors and then combine with all target vectors, and choose the best N vectors.

Variants of DE

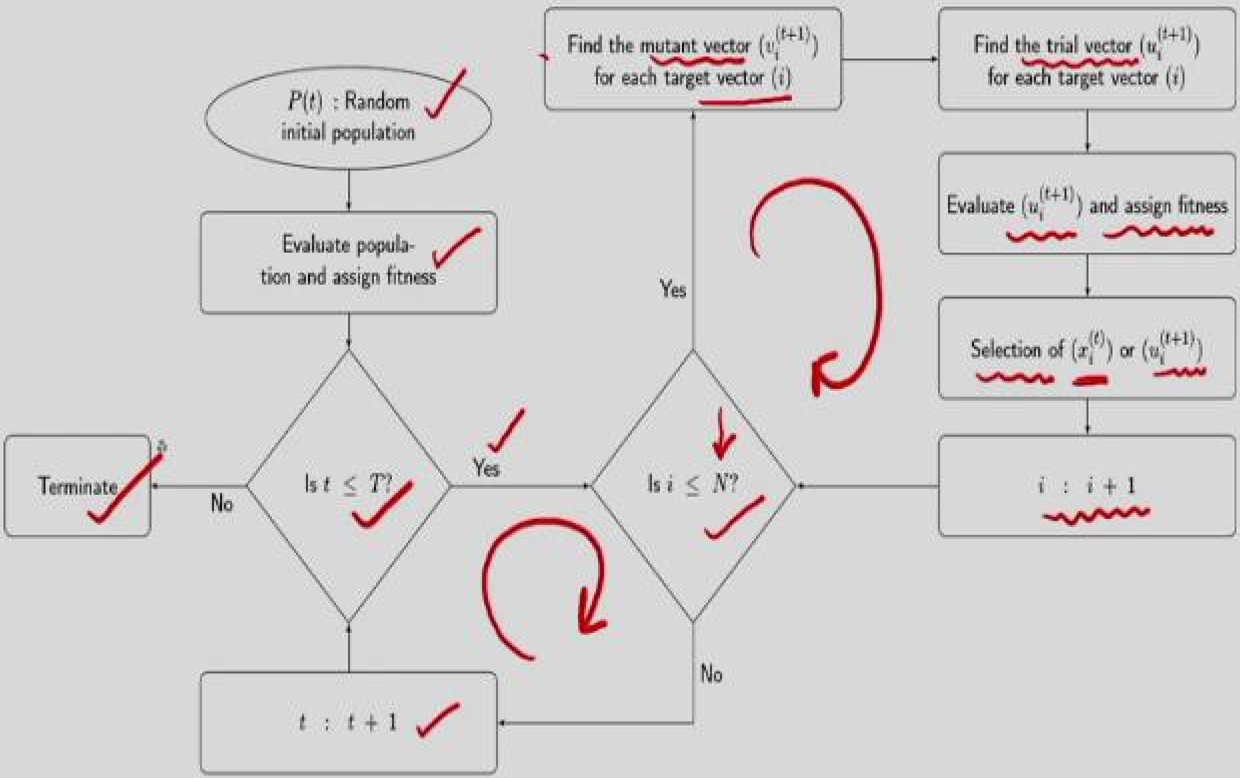
- We use the following notation for variants: DE $/x/y/z$
  - x: vector to be mutated (e.g., random or best)
  - y: is the number of vector differences used
  - z: denotes the crossover scheme
- Previous algorithm is DE $/rand/1/bin$
  - $rand$: choose a random vector for mutation ($x_{r1}^{(t)}$)
  - 1: a single vector difference is used ($x_{r2}^{(t)} - x_{r3}^{(t)}$)
  - bin: Crossover is according to independent binomial experiments (p_c)

DE $/best/2/bin$

$$v_i^{(t+1)} = x_{best}^{(t)} + F \times (x_{r1}^{(t)} + x_{r2}^{(t)} - x_{r3}^{(t)} - x_{r4}^{(t)})$$

- Mutate the best individual in the population
- Usage of 2 vector differences

Flowchart of DE



Generalized Framework of EC Techniques

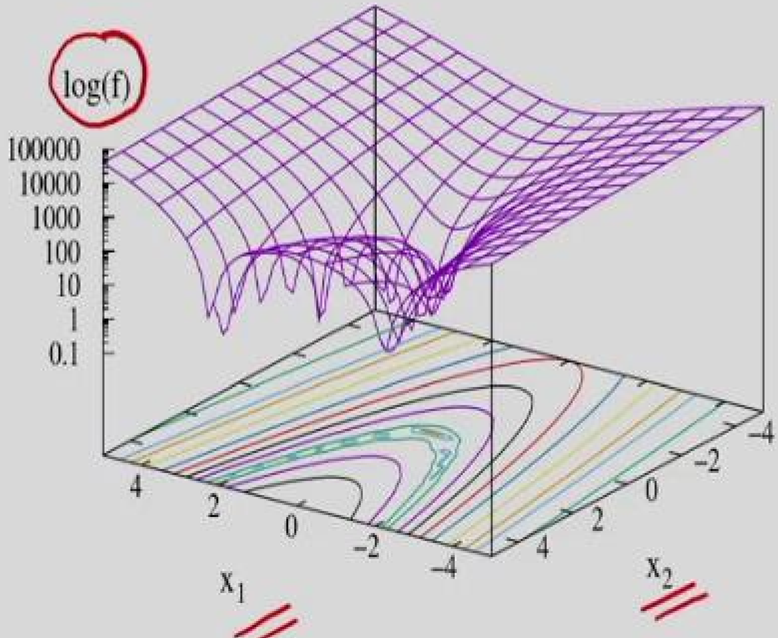
Algorithm 1 Generalized Framework for DE

Solution representation % Genetics
Input: $t := 1$ (Generation counter), Maximum allowed generation $= T$
Initialize random population $(P(t))$; % Population
Evaluate $(P(t))$; % Evaluate objective, constraints and assign fitness
while $t \leq T$ **do**
  for $(i = 1; i \leq N; i++)$ **do**
    ~~$M(t) :=$ Selection($P(t)$);~~ Find the mutant vector $(v_i^{(t+1)})$ for target vector (i); % Mutation
    ~~$Q(t) :=$ Variation($M(t)$);~~ Find the trial vector $(u_i^{(t+1)})$ for target vector (i); % Crossover
    Evaluate $(u_i^{(t+1)})$;
    $x_i^{(t+1)} :=$ Survivor$(x_i^{(t)}, u_i^{(t+1)})$; % Selection
  end for
  $t := t + 1;$
end while

Working Principles Through An Example

Rosenbrock Function

Minimize $\quad f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2,$
bounds $\quad -5 \leq x_1 \leq 5$ and $-5 \leq x_2 \leq 5.$



- Optimum solution is $x^* = (1,1)^T$ and $f(x) = 0$

Initial Population

- Let the population size is $N = 8$. Set $t = 1$.

Initial random population

Index(i)	$x_i^{(1)}$
1	$(2.212, 3.009)^T$
2	$(-2.289, -2.396)^T$
3	$(-2.393, -4.790)^T$
4	$(-0.639, 1.692)^T$
5	$(-3.168, 0.706)^T$
6	$(0.215, -2.350)^T$
7	$(-0.742, 1.934)^T$
8	$(-4.563, 4.791)^T$

Evaluate Population

- We calculate objective function $f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$ for each solution.
- For Solution 1: $x^{(1)} = (2.212, 3.009)^T$ and $f(x^{(1)}) = 357.154$.

Initial population

Index	$x_i^{(1)}$	$f(x_i^{(1)})$
1	$(2.212, 3.009)^T$	357.154
2	$(-2.289, -2.396)^T$	5843.569
3	$(-2.393, -4.790)^T$	11066.800
4	$(-0.639, 1.692)^T$	167.414
5	$(-3.168, 0.706)^T$	8718.166
6	$(0.215, -2.350)^T$	574.796
7	$(-0.742, 1.934)^T$	194.618
8	$(-4.563, 4.791)^T$	25731.235

Termination Condition

- Since it is the first generation, we continue.

Mutant Vector

- The mutant vector is $v_i^{(t+1)} = x_{r1}^{(t)} + F \times (x_{r2}^{(t)} - x_{r3}^{(t)})$

- Assume $F = 0.5$.

Random number

Index(i)	$r1$	$r2$	$r3$
1	7	3	2
2	3	4	6
3	1	7	6
4	2	1	7
5	4	6	2
6	3	7	8
7	1	8	4
8	4	5	3

- For target vector 1, $r1 = 7$, $r2 = 3$ and $r3 = 2$.
- Therefore, $v_1^{(2)} = x_7^{(1)} + 0.5 \times (x_3^{(1)} - x_2^{(1)})$.

$$v_1^{(2)} = \begin{pmatrix} -0.742 \\ 1.934 \end{pmatrix} + 0.5 \times \left(\begin{pmatrix} -2.393 \\ -4.790 \end{pmatrix} - \begin{pmatrix} -2.289 \\ -2.396 \end{pmatrix} \right)$$

- $v_1^{(2)} = \begin{pmatrix} -0.793 \\ 0.737 \end{pmatrix}$
- For target vector 2, $r1 = 3$, $r2 = 4$ and $r3 = 6$.
- Therefore, $v_2^{(2)} = x_3^{(1)} + 0.5 \times (x_4^{(1)} - x_6^{(1)})$.

$$v_2^{(2)} = \begin{pmatrix} -2.393 \\ -4.790 \end{pmatrix} + 0.5 \times \left(\begin{pmatrix} -0.639 \\ 1.692 \end{pmatrix} - \begin{pmatrix} 0.215 \\ -2.350 \end{pmatrix} \right)$$

- $v_2^{(2)} = \begin{pmatrix} -2.820 \\ -2.769 \end{pmatrix}$

Mutant Vector

Mutant Vectors

Index(i)	$v_i^{(2)}$	$f(v_i^{(2)})$
1	$(-0.793, 0.737)^T$	4.379
2	$(-2.820, -2.769)^T$	11505.375
3	$(1.734, 5.151)^T$	460.419
4	$(-0.812, -1.858)^T$	637.545
5	$(0.613, 1.714)^T$	179.211
6	$(-0.482, -6.218)^T$	4162.865
7	$(0.250, 4.558)^T$	2021.582
8	$(-1.026, 4.439)^T$	1150.566

Mutant Vectors after modification

Index(i)	$v_i^{(2)}$	$f(v_i^{(2)})$
1	$(-0.793, 0.737)^T$	4.379
2	$(-2.820, -2.769)^T$	11505.375
3	$(1.734, 5.000)^T$	398.024
4	$(-0.812, -1.858)^T$	637.545
5	$(0.613, 1.714)^T$	179.211
6	$(-0.482, -6.218)^T$	4162.865
7	$(0.250, 4.558)^T$	2021.582
8	$(-1.026, 4.439)^T$	1150.566

Trial Vector

- A trial vector $u_i^{(t+1)}$ is generated as

$$u_{i_j}^{(t+1)} = \begin{cases} v_{i_j}^{(t+1)} & \text{if } (\text{rand\_no} \le p_c) \text{ or } j = rnbr(i) \\ x_{i_j}^{(t)} & \text{if } (\text{rand\_no} > p_c) \text{ and } j \neq rnbr(i) \end{cases}, \text{ where } j \in 1, 2, \ldots, n$$

- Let us assume that $p_c = 0.5$.

Index(i)	rand_no$_1$	rand_no$_2$
1	0.459	0.122
2	0.268	0.684
3	0.792	0.149
4	0.674	0.265
5	0.695	0.282
6	0.691	0.708
7	0.293	0.493
8	0.755	0.873

- For trial vector 1, $v_1^{(2)} = (-0.793, 0.737)^T$ and $x_1^{(1)} = (2.212, 3.009)^T$
  - Since rand_no$_1$ = 0.459 for the first variable is less than $p_c = 0.5$, copy $u_{1_1}^{(2)} = v_{1_1}^{(2)} = -0.793$.
  - Since rand_no$_2$ = 0.122 for the second variable is less than $p_c = 0.5$, copy $u_{1_2}^{(2)} = v_{1_2}^{(2)} = 0.737$.

Trial Vector

- For trial vector 2, $v_2^{(2)} = (-2.820, -2.769)^T$ and $x_2^{(1)} = (-2.289, -2.396)^T$
  - Since rand_no$_1 = 0.268$ for the first variable is less than $p_c = 0.5$, copy $u_{2_1}^{(2)} = v_{2_1}^{(2)} = -2.820$.
  - Since rand_no$_2 = 0.684$ for the second variable is more than $p_c = 0.5$, copy $u_{2_2}^{(2)} = x_{2_2}^{(2)} = -2.396$.

Trial Vectors

Index(i)	$u_i^{(2)}$	$f(u_i^{(2)})$
1	$(-0.793, 0.737)^T$	4.379
2	$(-2.820, -2.396)^T$	10718.538
3	$(-2.393, 5.000)^T$	64.003
4	$(-0.639, -1.858)^T$	516.457
5	$(-3.168, 1.714)^T$	6937.656
6	$(0.215, -2.350)^T$	574.796
7	$(0.250, 4.558)^T$	2021.582
8	$(-4.563, 4.791)^T$	25731.235

Selection

- Greedy selection of DE

$$x_i^{(t+1)} = \begin{cases} u_i^{(t+1)} & \text{if } F(u_i^{(t+1)}) < F(x_i^{(t)}) \\ x_i^{(t)} & \text{Otherwise} \end{cases}$$

- Let us assume that the fitness value is the same as function value.

Fitness of target and trial vectors

Index(i)	$f(x_i^{(1)})$	$f(u_i^{(2)})$
1	357.154	4.379
2	5843.569	10718.538
3	11066.800	64.003
4	167.414	516.457
5	8718.166	6937.656
6	574.796	574.796
7	194.618	2021.582
8	25731.235	25731.235

New Target Vectors for the next generation

Index(i)	$x_i^{(2)}$	$f(x_i^{(2)})$
1	$(-0.793, 0.737)^T$	4.379
2	$(-2.289, -2.396)^T$	5843.569
3	$(-2.393, 5.000)^T$	64.003
4	$(-0.639, 1.692)^T$	167.414
5	$(-3.168, 1.714)^T$	6937.656
6	$(0.215, -2.350)^T$	574.796
7	$(-0.742, 1.934)^T$	194.618
8	$(-4.563, 4.791)^T$	25731.235

Compare Vectors After One generation

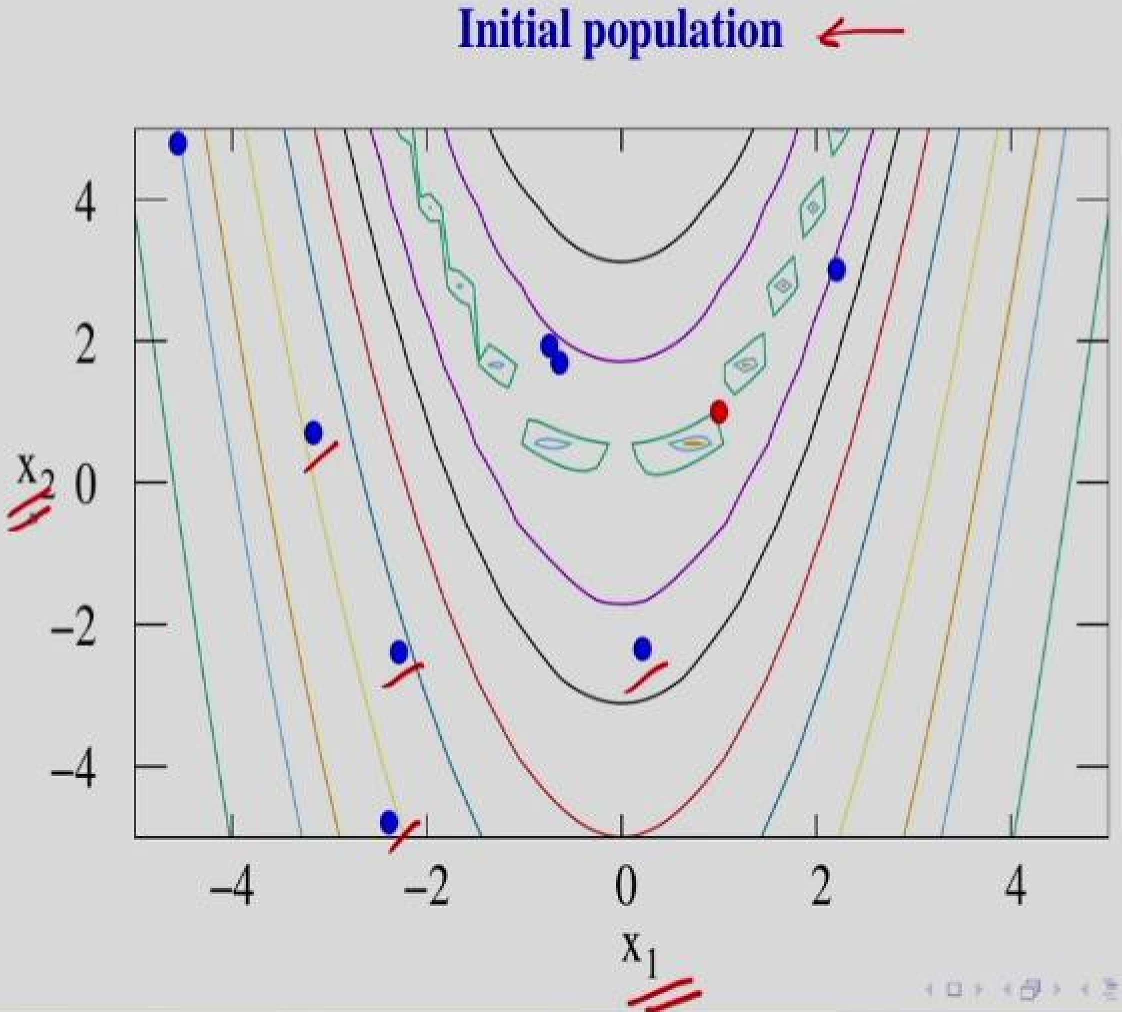
Initial population

Index	$x_i^{(1)}$	$f(x_i^{(1)})$
1	$(2.212, 3.009)^T$	357.154
2	$(-2.289, -2.396)^T$	5843.569
3	$(-2.393, -4.790)^T$	11066.800
4	$(-0.639, 1.692)^T$	$\boxed{167.414}$
5	$(-3.168, 0.706)^T$	8718.166
6	$(0.215, -2.350)^T$	574.796
7	$(-0.742, 1.934)^T$	194.618
8	$(-4.563, 4.791)^T$	25731.235

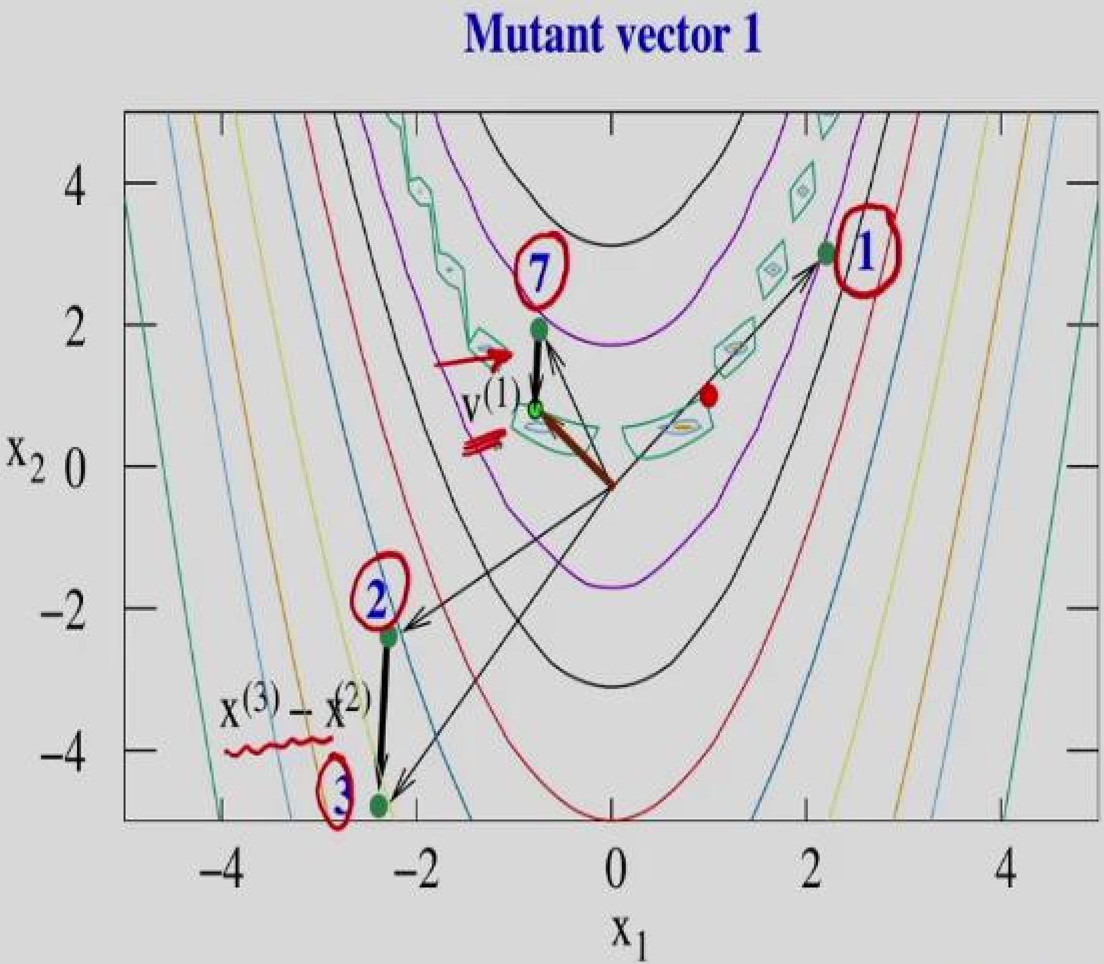
New Target Vectors

Index(i)	$x_i^{(2)}$	$f(x_i^{(2)})$
1	$(-0.793, 0.737)^T$	4.379
2	$(-2.289, -2.396)^T$	5843.569
3	$(-2.393, 5.000)^T$	64.003
4	$(-0.639, 1.692)^T$	167.414
5	$(-3.168, 1.714)^T$	6937.656
6	$(0.215, -2.350)^T$	574.796
7	$(-0.742, 1.934)^T$	194.618
8	$(-4.563, 4.791)^T$	25731.235

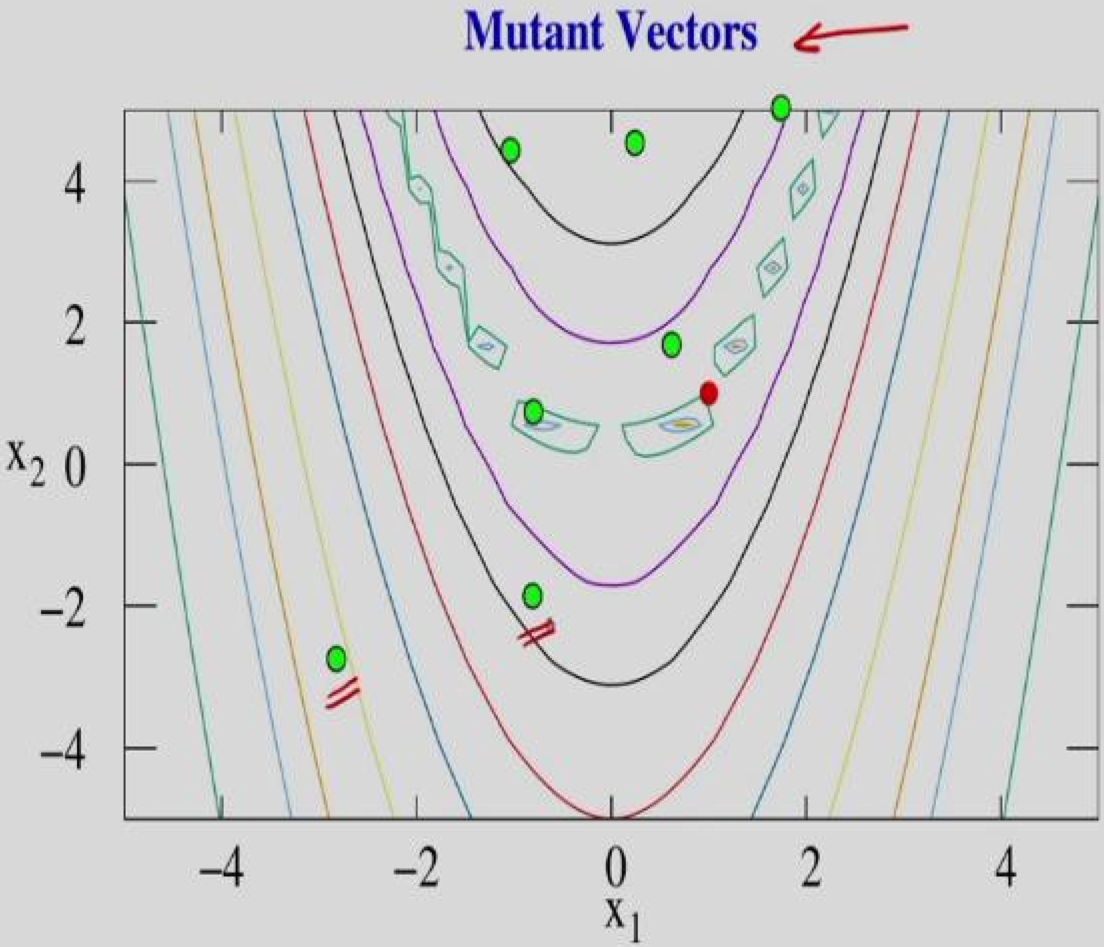
Graphical Example



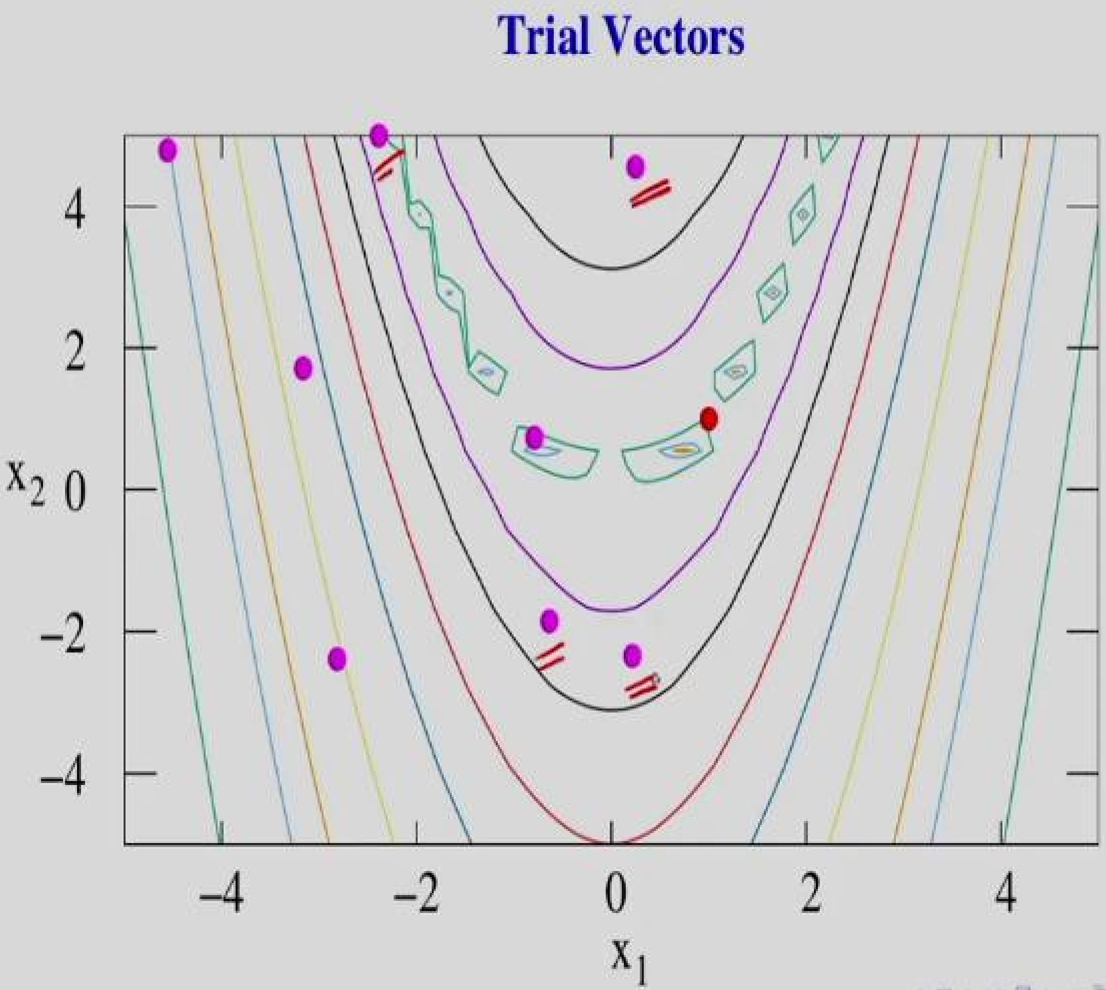
Graphical Example



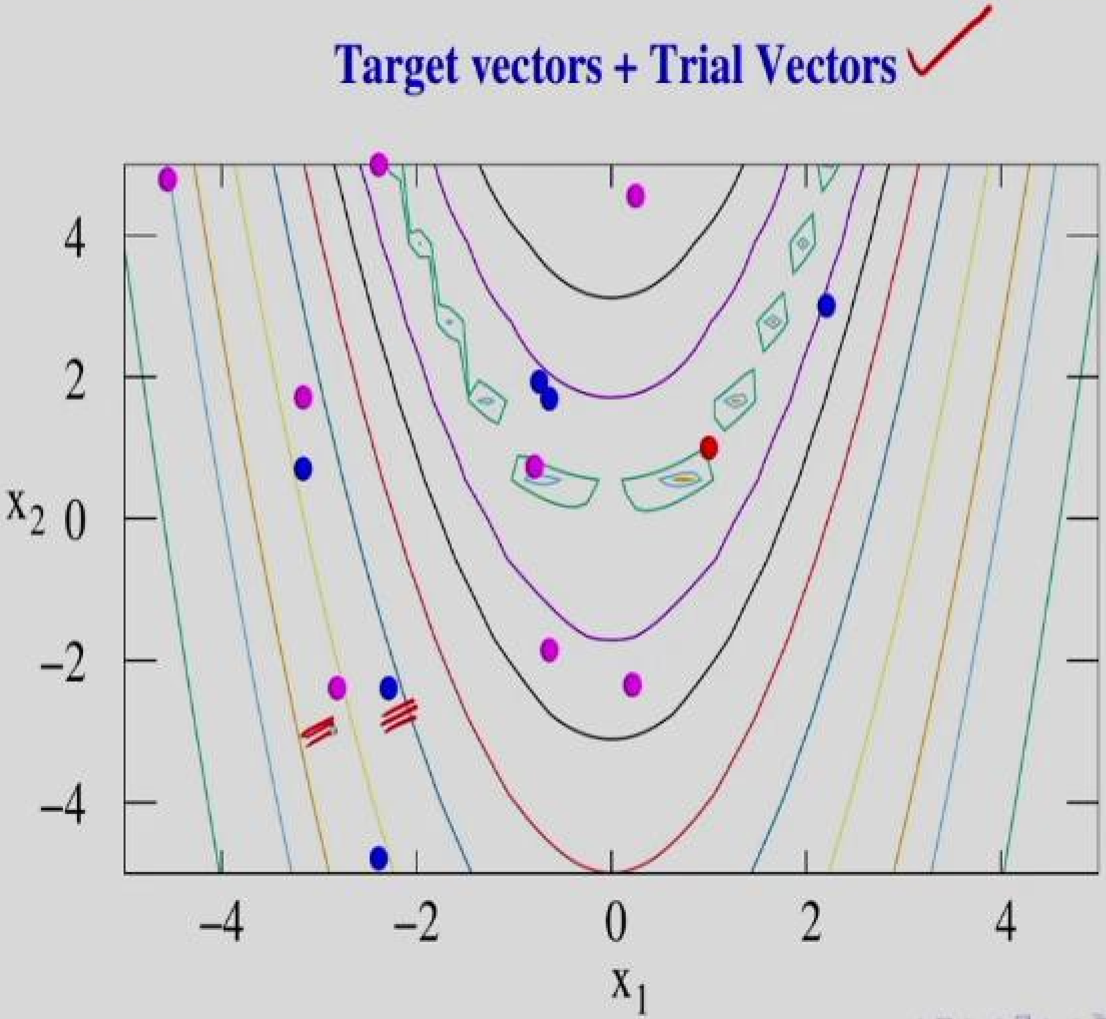
Graphical Example



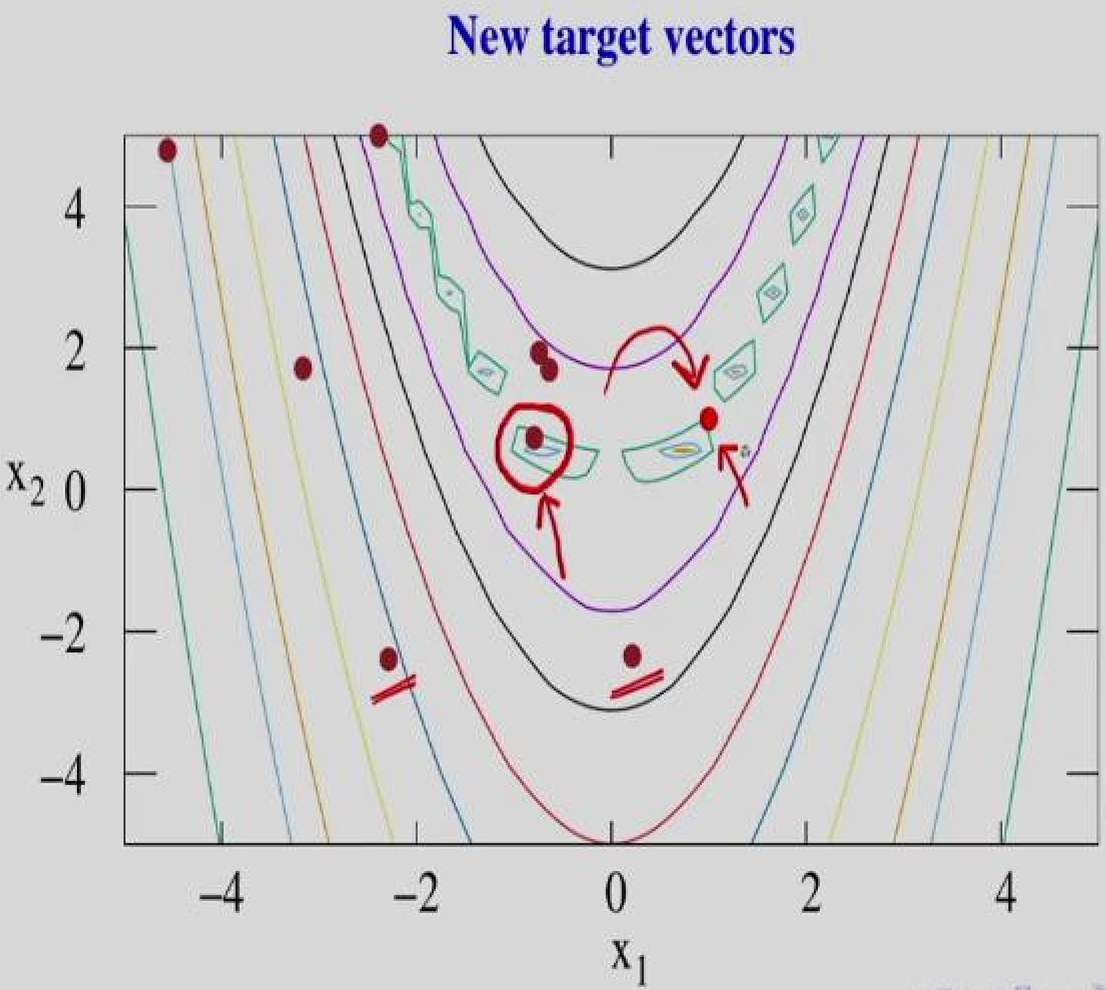
Graphical Example



Graphical Example



Graphical Example



Closure

- Differential evolution (DE)
  - Introduction
  - Mutant vector
  - Trial vector
  - Greedy selection of canonical DE
- Flow chart of DE
- DE on the generalized framework of EC techniques
- Working principles through an example
- Graphical example