Algoritmo de Evolução Diferencial

Cristiano L. Castro

Adaptado de material produzido pelo Prof. Frederico G. Guimarães.

Sumário

Introdução

Algoritmo de Evolução Diferencial (DE):

- Proposto por Rainer Storn e Kenneth Price em 1995.
- Muito popular na otimização não linear com variáveis contínuas.
- Mecanismo básico de busca: **operador de mutação diferencial**.
- Considerado um AE, embora não seja inspirado em processo natural.
- Qualidades computacionais interessantes:
  - simplicidade de implementação;
  - robustez e eficiência;
  - autoadaptação;
  - versatilidade.

Seja o problema de otimização não linear com variáveis reais contínuas:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} f(\mathbf{x})$$
$$\text{sujeito a:} \begin{cases} \mathbf{g}(\mathbf{x}) \leq 0 \\ \mathbf{h}(\mathbf{x}) = 0 \end{cases} \tag{1}$$

Incialmente, vamos considerar o problema irrestrito, isto é, sem as funções de restrição $\mathbf{g}(\mathbf{x})$ e $\mathbf{h}(\mathbf{x})$.

Notação:

- $\mathcal{U}_{[a,b]}$: amostragem com distribuição uniforme entre a e b;
- $\mathcal{N}_{[\mu,\sigma]}$: amostragem com distribuição normal com média μ e desvio padrão σ.

Seja uma população de soluções candidatas $X_t = \{\mathbf{x}_{t,i};\ i = 1, \ldots, N\}$.

Cada indivíduo é representado por um vetor coluna:

$$\mathbf{x}_{t,i} = \begin{bmatrix} x_{t,i,1} \\ x_{t,i,2} \\ \vdots \\ x_{t,i,n} \end{bmatrix} \tag{2}$$

em que, o terceiro índice indica uma entre as n variáveis do problema.

O mecanismo de busca do DE utiliza **vetores-diferença**:

- Dois indivíduos são selecionados aleatoriamente para se criar um vetor-diferença;
- Este vetor-diferença é somado a um terceiro indivíduo, também selecionado aleatoriamente, produzindo uma solução mutante;
- A solução mutante é portanto o resultado de uma perturbação em algum indivíduo da população;
- Esta perturbação é um vetor-diferença construído aleatoriamente.

A equação a seguir ilustra esse procedimento:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right), \quad r_1, r_2, r_3 \in \{1, \ldots, N\} \tag{3}$$

Em que $\mathbf{v}_{t,i}$ representa a i-ésima solução mutante;

F é um <u>fator de escala</u> aplicado ao vetor-diferença;

O vetor $\mathbf{x}_{t,r_1}$ é denominado <u>vetor de base</u>;

A partir deste procedimento, obtém-se uma população mutante $V_t = \{\mathbf{v}_{t,i};\ i = 1, \ldots, N\}$.

A recombinação é obtida da seguinte maneira:

- Os indivíduos da população corrente X_t são recombinados com os indivíduos da população mutante V_t.
- Produz-se a descendência ou população de soluções teste U_t.
- Na versão clássica do DE, emprega-se recombinação discreta com probabilidade $C \in [0, 1]$:

$$u_{t,i,j} = \begin{cases} v_{t,i,j}, & \text{se } \mathcal{U}_{[0,1]} \leq C \ \vee \ j = \delta_i \\ x_{t,i,j}, & \text{caso contrário} \end{cases} \tag{4}$$

em que $\delta_i \in \{1, \ldots, n\}$ é um índice gerado aleatoriamente.

- O parâmetro C controla o efeito da recombinação discreta.

A Figura 1 ilustra a geração de um vetor mutante e as possíveis soluções teste obtidas após a recombinação, indicadas por ■.

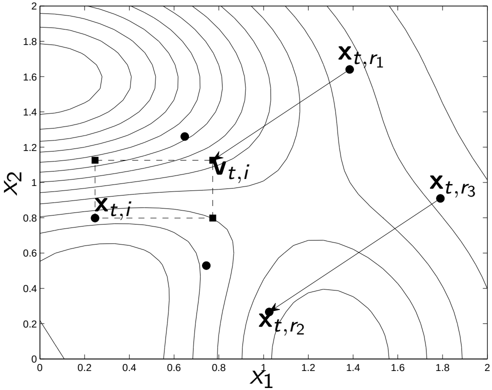


Figure : Ilustração do procedimento de geração de uma nova solução.

Os operadores do DE em forma compacta são:

$$u_{t,i,j} = \begin{cases} x_{t,r_1,j} + F\left(x_{t,r_2,j} - x_{t,r_3,j}\right), & \text{se } \mathcal{U}_{[0,1]} \leq C \ \vee \ j = \delta_i \\ x_{t,i,j}, & \text{caso contrário} \end{cases} \tag{5}$$

com $r_1, r_2, r_3 \in \{1, \ldots, N\}$, $t = 1, \ldots, t_{\max}$, $i = 1, \ldots, N$, e $j = 1, \ldots, n$.

A **seleção para sobrevivência** pode ser descrita por:

$$\mathbf{x}_{t+1,i} = \begin{cases} \mathbf{u}_{t,i}, & \text{se } f(\mathbf{u}_{t,i}) \leq f(\mathbf{x}_{t,i}) \\ \mathbf{x}_{t,i}, & \text{caso contrário} \end{cases} \tag{6}$$

Evolução Diferencial

Notação: algoritmo

Algorithm 1: Algoritmo Modelo de Evolução Diferencial

$t \leftarrow 1$;
Inicializar população $X_t = \{\mathbf{x}_{t,i};\ i = 1, \dots, N\}$;
enquanto *algum critério de parada não for satisfeito* **faça**
    para $i = 1$ *até* N **hacer**
        Selecionar vetor de base $\mathbf{x}_{t,base}$; (seleção para reprodução)
        Selecionar vetor-diferença $\{\Delta\mathbf{x}_t\}$;
        Selecionar fator de escala F;
        Gerar solução mutante usando $\mathbf{v}_{t,i} = \mathbf{x}_{t,base} + F\Delta\mathbf{x}_t$;
        Adicionar $\mathbf{v}_{t,i}$ à população mutante V_t;
    fin
    Recombinar X_t e V_t, com parâmetro C, gerando U_t;
    Aplicar seleção para sobrevivência entre X_t e U_t;
    $t \leftarrow t + 1$;
fim

Supondo uma população de tamanho N, com <u>N vetores distintos entre si</u>:

- Existem ao todo N^2 combinações de diferenças possíveis;
- N são nulas, pois correspondem a diferenças entre um mesmo vetor;
- As $N(N-1)$ diferenças restantes são não nulas;
- Essas $N(N-1)$ diferenças apresentam simetria, porque cada diferença $(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3})$ possui seu simétrico correspondente;
- A probabilidade de selecionar uma dada diferença e seu simétrico é a mesma.

Interpretação básica do mecanismo de autoadaptação do DE:

- A distribuição dos vetores-diferença depende da distribuição espacial dos indivíduos da população.
- À medida que a população se distribui de acordo com o "contorno" da função, a distribuição dos vetores-diferença também se ajusta a esse contorno.

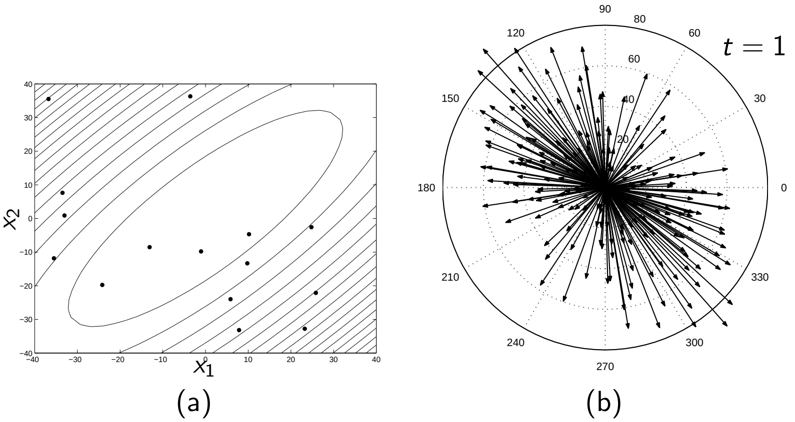
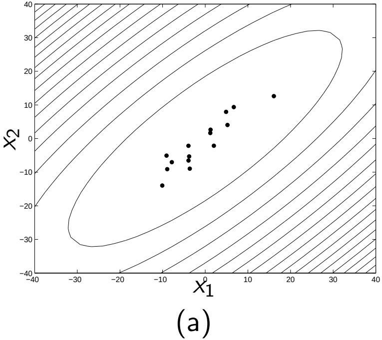


Figure : Função-objetivo quadrática. (a) Distribuição espacial da população na geração $t = 1$. (b) Distribuição dos vetores-diferença na geração $t = 1$.



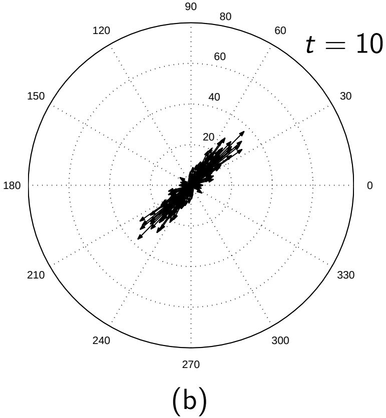


Figure : Função-objetivo quadrática. (a) Distribuição espacial da população na geração $t = 10$. (b) Distribuição dos vetores-diferença na geração $t = 10$.

Comportamento da Mutação Diferencial

Função convexa

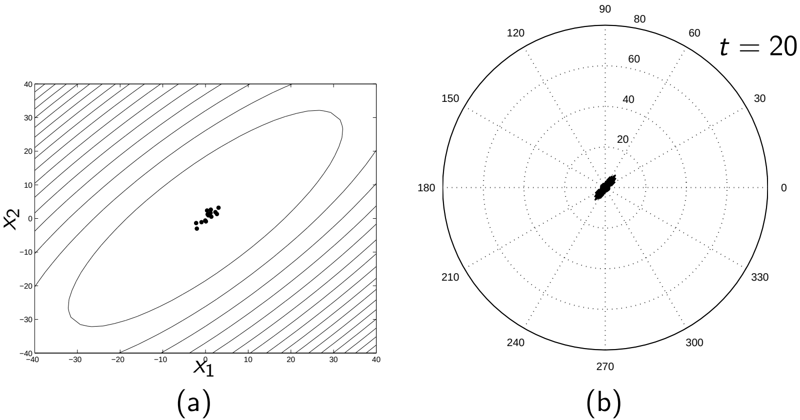


Figure : Função-objetivo quadrática. (a) Distribuição espacial da população na geração $t = 20$. (b) Distribuição dos vetores-diferença na geração $t = 20$.

Comportamento da Mutação Diferencial

Função multimodal

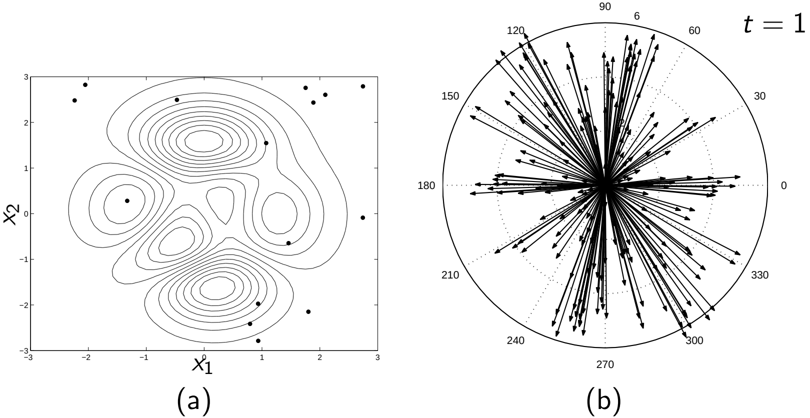


Figure : Função-objetivo multimodal. (a) Distribuição espacial da população na geração $t = 1$. (b) Distribuição dos vetores-diferença na geração $t = 1$.

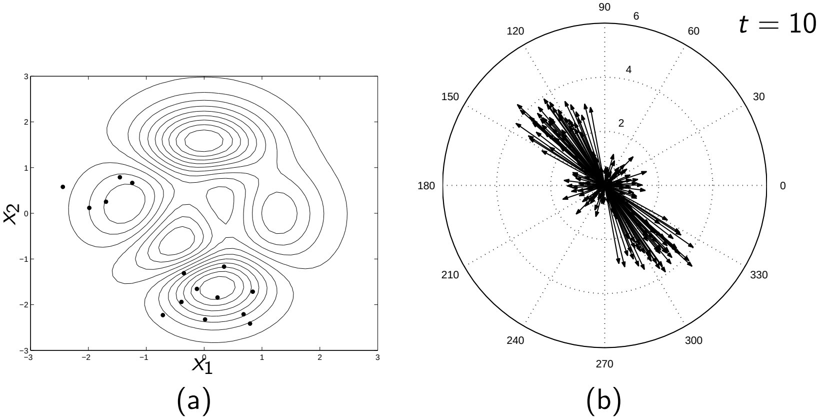
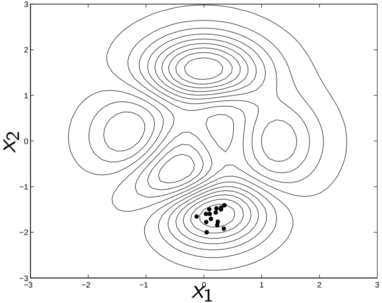
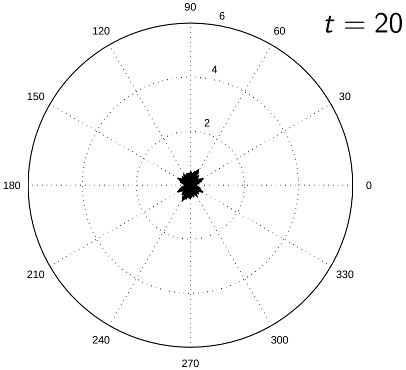


Figure : Função-objetivo multimodal. (a) Distribuição espacial da população na geração $t = 10$. (b) Distribuição dos vetores-diferença na geração $t = 10$.



(a)



(b)

Figure : Função-objetivo multimodal. (a) Distribuição espacial da população na geração $t = 20$. (b) Distribuição dos vetores-diferença na geração $t = 20$.

Neste caso, usaremos a função de Rosenbrock:

$$f(\mathbf{x}) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2 \tag{7}$$

- É unimodal e suas curvas de nível formam conjuntos não convexos.
- O ponto de mínimo se localiza em um vale estreito e longo.
- Essa região apresenta uma inclinação pequena que dificulta a convergência de métodos de otimização baseados em derivadas.

Comportamento da Mutação Diferencial

Função unimodal não convexa

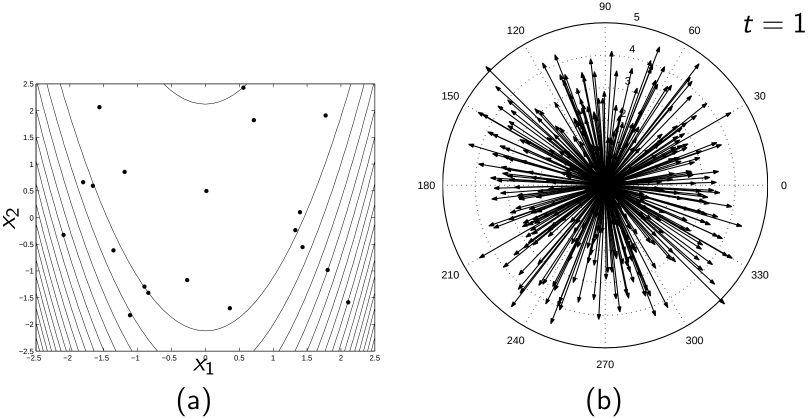
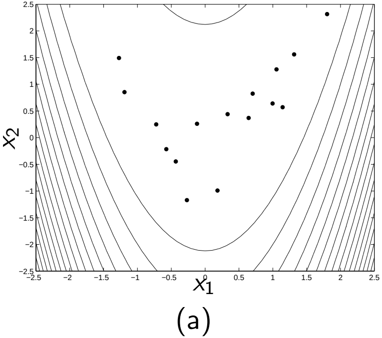


Figure : Função-objetivo unimodal não convexa. (a) Distribuição espacial da população em $t = 1$. (b) Distribuição dos vetores-diferença em $t = 1$.



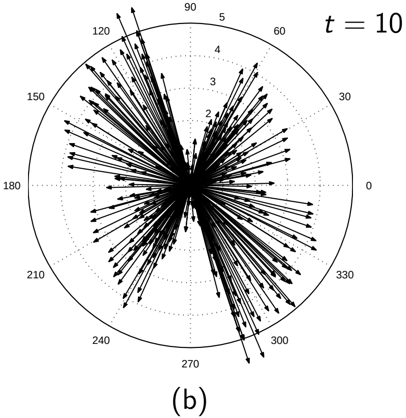


Figure : Função-objetivo unimodal não convexa. (a) Distribuição espacial da população em $t = 10$. (b) Distribuição dos vetores-diferença em $t = 10$.

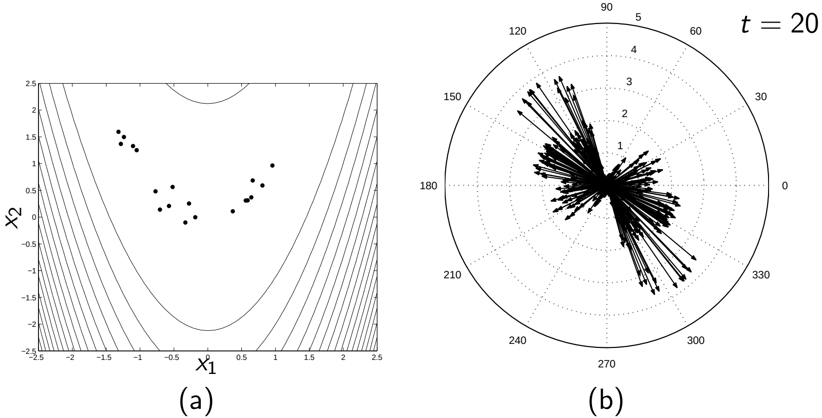


Figure : Função-objetivo unimodal não convexa. (a) Distribuição espacial da população em $t = 20$. (b) Distribuição dos vetores-diferença em $t = 20$.

Os índices r_1, r_2, r_3 são sorteados aleatoriamente para cada i.

O que ocorre quando algum par dos índices i, r_1, r_2, r_3 é coincidente?

- Se $r_2 = r_3$:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1}$$

  Não há perturbação aplicada ao vetor de base.

  Ocorre apenas recombinação de $\mathbf{x}_{t,i}$ com $\mathbf{x}_{t,r_1}$.

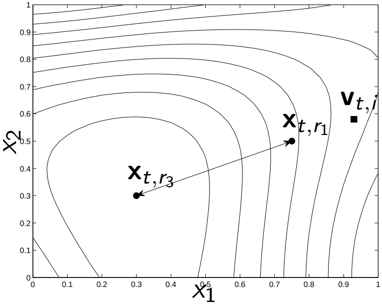
- Se $r_1 = r_2$ ou $r_1 = r_3$, tem-se uma recombinação aritmética.
- Para $r_1 = r_2$, temos:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + F\left(\mathbf{x}_{t,r_1} - \mathbf{x}_{t,r_3}\right) = (1+F)\mathbf{x}_{t,r_1} - F\mathbf{x}_{t,r_3}$$

Para $r_1 = r_3$, ocorre uma recombinação aritmética tradicional entre $\mathbf{x}_{t,r_1}$ e $\mathbf{x}_{t,r_2}$ (se $F \in [0,1]$):

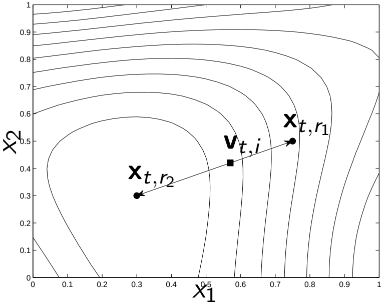
$$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_1}\right) = (1-F)\mathbf{x}_{t,r_1} + F\mathbf{x}_{t,r_2}$$

Se $F > 1$ os cruzamentos são degenerados.



(a)

Figure : Quando $r_1 = r_2$, a nova solução é gerada fora do segmento que une as duas soluções recombinantes. No exemplo acima, $F = 0.4$.



(b)

Figure : Quando $r_1 = r_3$, ocorre uma recombinação aritmética tradicional. No exemplo acima, $F = 0.4$.

- Se $i = r_1$. Uma perturbação é aplicada a $\mathbf{x}_{t,i}$ de acordo com:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,i} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right)$$

A solução teste $\mathbf{u}_{t,i}$ é o resultado da recombinação da solução $\mathbf{x}_{t,i}$ com sua versão perturbada.

O parâmetro C passa a ter o significado de um parâmetro de mutação, controlando quantas variáveis de $\mathbf{x}_{t,i}$ serão perturbadas.

- Se $i = r_2$ ou $i = r_3$:

  Nesse caso o vetor-diferença está na direção que liga $\mathbf{x}_{t,i}$ a algum outro vetor da população, $\mathbf{x}_{t,r_2}$ ou $\mathbf{x}_{t,r_3}$.

  Essa combinação não é necessariamente indesejável.

- Estas combinações degeneradas, mesmo que tenham baixa probabilidade de ocorrência, **são em geral indesejáveis e podem prejudicar o desempenho do algoritmo**.

- Por essa razão, recomenda-se adotar índices mutuamente distintos, isto é, $i \neq r_1 \neq r_2 \neq r_3$, em implementações mais práticas do algoritmo de evolução diferencial.

Algorithm 2: Pseudocódigo para a seleção de índices mutuamente distintos

```
repita
    Selecione aleatoriamente r1 ∈ {1, ..., N};
até r1 ≠ i;
repita
    Selecione aleatoriamente r2 ∈ {1, ..., N};
até r2 ≠ i ∧ r2 ≠ r1;
repita
    Selecione aleatoriamente r3 ∈ {1, ..., N};
até r3 ≠ i ∧ r3 ≠ r1 ∧ r3 ≠ r2;
```

- O algoritmo de evolução diferencial em sua versão básica utiliza:

(i) seleção aleatória do vetor base com probabilidade uniforme;

(ii) um (01) vetor-diferença para a mutação, e

(iii) recombinação discreta entre a solução corrente e seu vetor mutante correspondente.

- Entretanto, muitas variações desse esquema básico são possíveis.

- Pode-se utilizar um mesmo vetor como vetor de base em todas as operações de mutação diferencial. Assim, temos:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,best} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right) \tag{8}$$

onde $\mathbf{x}_{t,best}$ representa o melhor indivíduo na população na geração t.

- Pode-se adotar o vetor correspondente à média da distribuição espacial da população como vetor base.

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,mean} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right) \tag{9}$$

com:

$$\mathbf{x}_{t,mean} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{t,i} \tag{10}$$

- Outra possibilidade é gerar soluções na direção do melhor indivíduo para servirem como vetores de base.

$$\mathbf{v}_{t,i} = \underbrace{\mathbf{x}_{t,i} + \lambda \left(\mathbf{x}_{t,best} - \mathbf{x}_{t,i}\right)}_{\mathbf{x}_{t,base}} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right) \qquad (11)$$

com $\lambda \in [0, 1]$.

- O vetor de base corresponde a um ponto gerado aleatoriamente sobre o segmento que liga $\mathbf{x}_{t,i}$ e $\mathbf{x}_{t,best}$.
- Essa forma de seleção de $\mathbf{x}_{t,base}$ apresenta menor pressão seletiva que em (8), embora haja uma polarização na direção da melhor solução.

- A mutação diferencial pode ser generalizada para empregar mais vetores-diferença na criação do vetor mutante:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + \sum_{k=1}^{d} F_k \Delta \mathbf{x}_{t,k} \tag{12}$$

em que a perturbação aplicada ao vetor de base é composta pela soma de d vetores-diferença da forma:

$$\Delta \mathbf{x}_{t,k} = \left(\mathbf{x}_{t,r_{k+1}} - \mathbf{x}_{t,r_{k+1+d}} \right) \tag{13}$$

- Por exemplo, usando o mesmo valor de F_k para todo k e $d = 3$:

$$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + F \left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_5} + \mathbf{x}_{t,r_3} - \mathbf{x}_{t,r_6} + \mathbf{x}_{t,r_4} - \mathbf{x}_{t,r_7} \right) \tag{14}$$

- O uso de mais de um vetor-diferença na mutação diferencial **aumenta a capacidade de geração de diversidade no algoritmo**.

- Contudo, essa estratégia reduz a capacidade de alinhamento dos vetores-diferença com o contorno da função.

- O uso de mais vetores-diferença aumenta a busca global no algoritmo e prejudica a busca local.

- Existe uma notação sintética para representar as variações do DE.

  Essa notação padrão segue o formato `DE/base/d/rec`:

@ `base` indica a forma como o vetor de base é selecionado.

@ `d` indica o número de vetores-diferença.

@ `rec` faz referência ao operador de recombinação utilizado.

- Por exemplo, a versão clássica do DE pode ser representada pela notação `DE/rand/1/bin`.

Notação	Mutação diferencial
`DE/rand/1/bin`	$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right)$
`DE/best/1/bin`	$\mathbf{v}_{t,i} = \mathbf{x}_{t,best} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right)$
`DE/mean/1/bin`	$\mathbf{v}_{t,i} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_{t,k} + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right)$
`DE/rand-to-best/1/bin`	$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + \lambda\left(\mathbf{x}_{t,best} - \mathbf{x}_{t,r_1}\right) + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right)$
`DE/current-to-best/1/bin`	$\mathbf{v}_{t,i} = \mathbf{x}_{t,i} + \lambda\left(\mathbf{x}_{t,best} - \mathbf{x}_{t,i}\right) + F\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_3}\right)$
`DE/rand/2/bin`	$\mathbf{v}_{t,i} = \mathbf{x}_{t,r_1} + F_1\left(\mathbf{x}_{t,r_2} - \mathbf{x}_{t,r_4}\right) + F_2\left(\mathbf{x}_{t,r_3} - \mathbf{x}_{t,r_5}\right)$

Table : Algumas instâncias do algoritmo de evolução diferencial.