UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

ESTRATÉGIAS EVOLUTIVAS (EE)

Cristiano Leite de Castro
crislcastro@ufmg.br

Departamento de Engenharia Elétrica
Universidade Federal de Minas Gerais
Belo Horizonte, Brasil

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Introdução

- Estratégias Evolutivas (*Evolution Strategies*) constituem uma classe de algoritmos evolutivos usados principalmente em otimização numérica.
- Estudos préliminares envolvendo EE foram desenvolvidos nos anos 60 na Alemanha: *Rechenberg* and *Schwefel*;

Rechenberg e Schwefel (1960)

Introdução

- **características gerais:**
  - rápida convergência;
  - excelente desempenho para problemas de otimização numérica (variáveis reais);
  - seleção determinística de sobreviventes (baseada em *aptidão*)
  - seleção aleatória dos pais (não é baseada em *aptidão*);
  - **característica especial: autoadaptação de parâmetros (coevolução)**;

UFMG

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

EE - Padrão:

Representação	Vetores reais
Recombinação	Discreta ou intermediária
Mutação	Gaussiana
Seleção para reprodução	Aleatória uniforme
Seleção para sobrevivência	(μ,λ) ou $(\mu+\lambda)$
Característica especial	Autoadaptação de atributos da mutação

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Notação

Dada uma EE com uma população de μ indivíduos que produzem λ descendentes:

- ES(μ+λ): μ indivíduos geram λ filhos; do total de μ+λ soluções, os μ melhores são selecionados para a próxima geração;

- ES(μ,λ): μ indivíduos geram λ filhos; das λ novas soluções, os μ melhores são selecionados para a próxima geração – requer μ<λ.

Representação

- Minimização de uma função n-dimensional $f: \mathbb{R}^n \to \mathbb{R}$

- Um indivíduo na estratégia evolutiva é composto pela variáveis de otimização e os parâmetros da estratégia:

$$a = \langle \boldsymbol{x} | \sigma | \alpha \rangle$$
$$= \left\langle x_1, \dots, x_n \middle| \sigma_1, \dots, \sigma_n \middle| \alpha_1, \dots, \alpha_{n(n-1)/2} \right\rangle$$

<u>Na sequência, tem-se:</u>

- Codificação do vetor solução candidata;
- Tamanhos de mutação;
- Ângulos de rotação da matriz de covariância (representam interações entre os σ_i);
- Outros componentes podem ser considerados.

Representação

- o formato do cromossomo (indivíduo) depende do tipo da estratégia de evolução selecionada:
  - Auto-adaptativa com tamanho de passo único (global):

$$\vec{a} = ((x_1, \ldots, x_n), \sigma)$$

  - Auto-adaptativa com *n* tamanhos de passo:

$$\vec{a} = ((x_1, \ldots, x_n), (\sigma_1, \ldots, \sigma_n))$$

  - Auto-adaptativa correlacionada:

$$\vec{a} = ((x_1, \ldots, x_n), (\sigma_1, \ldots, \sigma_n), (\alpha_1, \ldots, \alpha_{n(n-1)/2}))$$

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Esquema Geral

- **Autoadaptação** – o desvio padrão é parte do indivíduo, e o tamanho da mutação evolui juntamente com a solução do problema de otimização:

$$\langle x_t | \sigma_t \rangle \rightarrow \langle x_{t+1} | \sigma_{t+1} \rangle$$

$$\sigma_{t+1} = \sigma_t \exp\left[N(0,\delta)\right] \qquad \delta \propto \sqrt{\frac{1}{n}}$$

$$x_{t+1} = x_t + N(0, \sigma_{t+1} I)$$

- **A ordem é importante**: primeiro perturba-se σ, em seguida usa-se esse valor para perturbar a solução.

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Lógica de Funcionamento

- A solução é avaliada explicitamente na função-objetivo.
- O tamanho da mutação é avaliado implicitamente:
  - se o σ usado é bom, então ele tende a gerar uma boa solução que por sua vez tem um bom valor de função.
- Reverter a ordem não faz sentido!
- A nova solução gerada compete com as μ existentes pela sobrevivência.

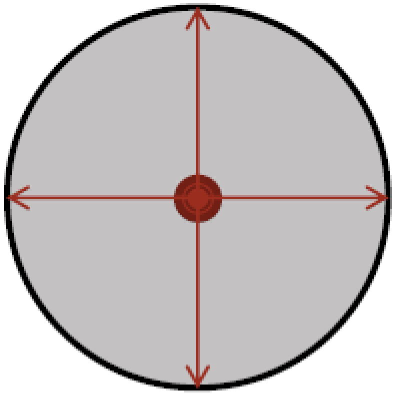
Estratégia (1)

Mutação descorrelacionada com um desvio padrão:

$$\langle x_t | \sigma_t \rangle \rightarrow \langle x_{t+1} | \sigma_{t+1} \rangle$$

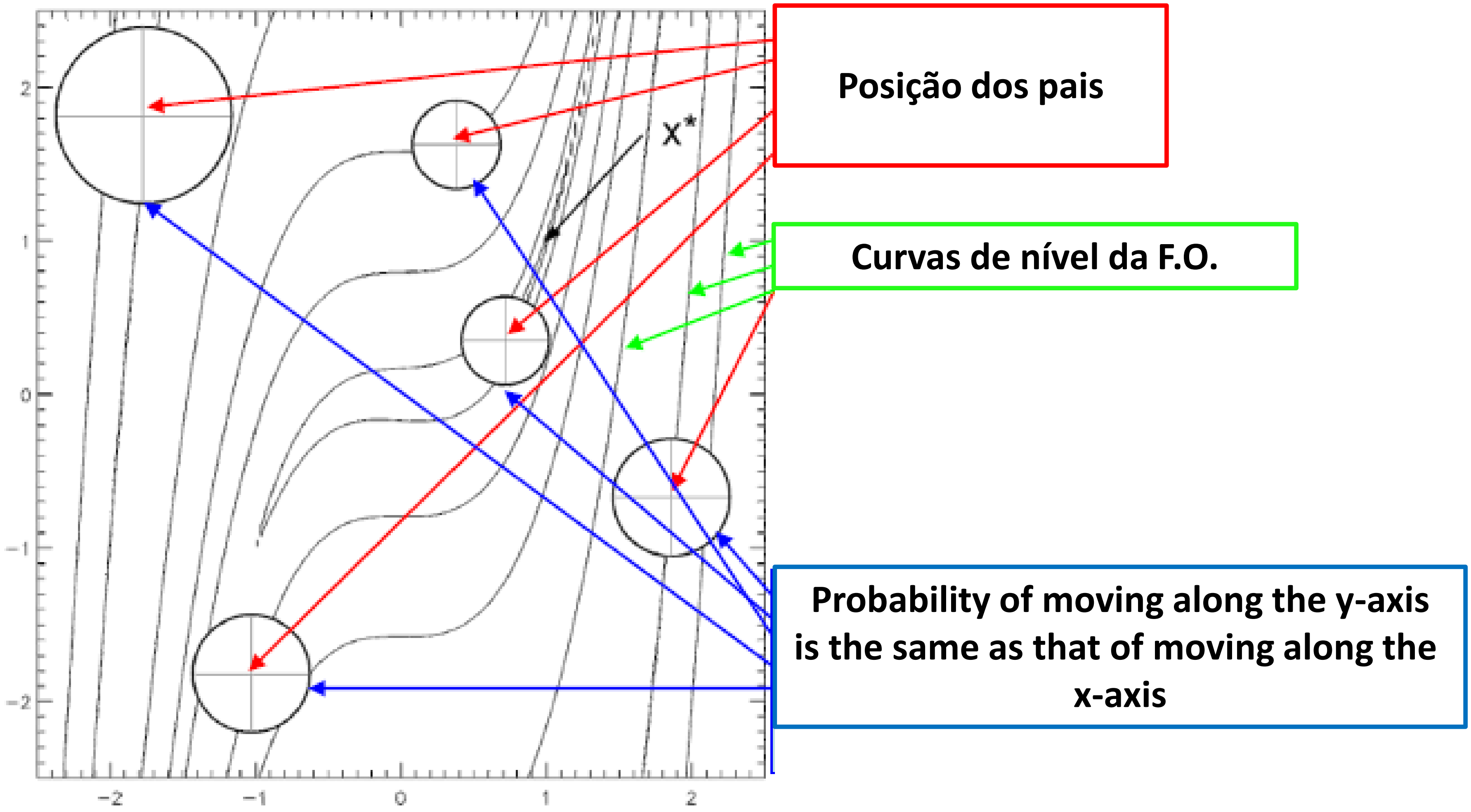
$$\sigma_{t+1} = \sigma_t \exp\left[N(0, \delta)\right] \qquad \delta \propto \sqrt{\frac{1}{n}}$$

$$x_{t+1} = x_t + N(0, \sigma_{t+1} I)$$



Exemplo Ilustrativo

- **mutação descorrelacionada com 1 desvio-padrão:**



Estratégia (2)

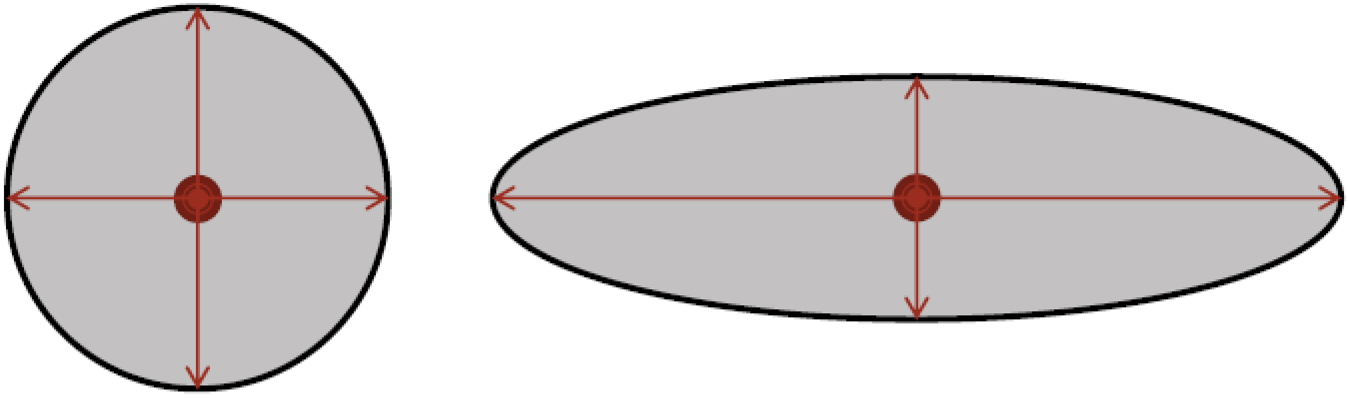
Mutação descorrelacionada com n desvios padrão:

$$\langle x_t | \sigma_t \rangle \rightarrow \langle x_{t+1} | \sigma_{t+1} \rangle \qquad \sigma_{t+1,i} = \sigma_{t,i} \exp\left[N(0,\tau_1) + N(0,\tau_2)\right]$$

$$x_{t+1,i} = x_{t,i} + N\left(0, \sigma_{t+1,i} I\right)$$

$$\tau_1 \propto 1 / \sqrt{2n}$$

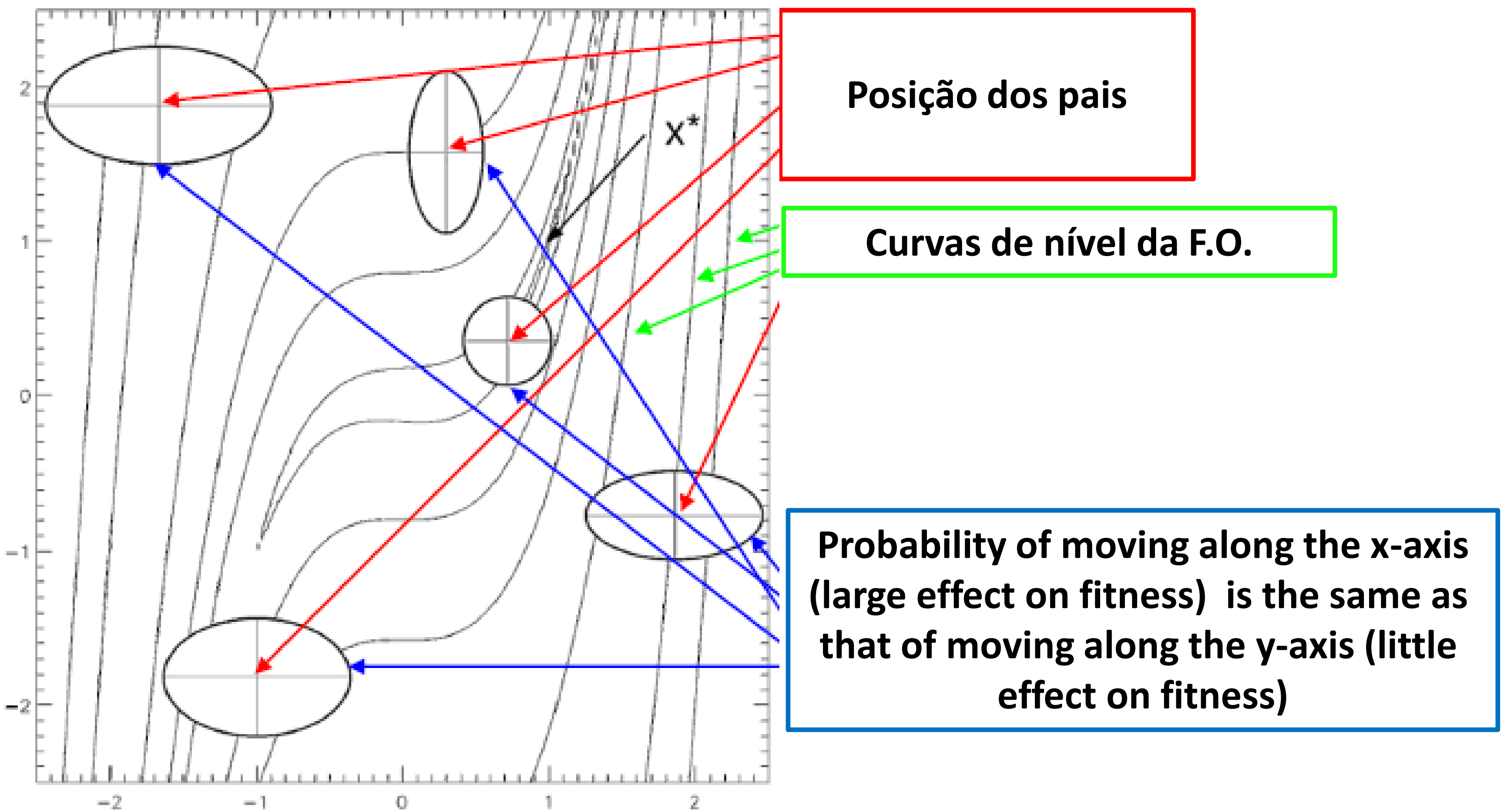
$$\tau_2 \propto 1 / \sqrt{2\sqrt{n}}$$



UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Exemplo Ilustrativo

- **mutação descorrelacionada com n desvios-padrão:**



Estratégia (3)

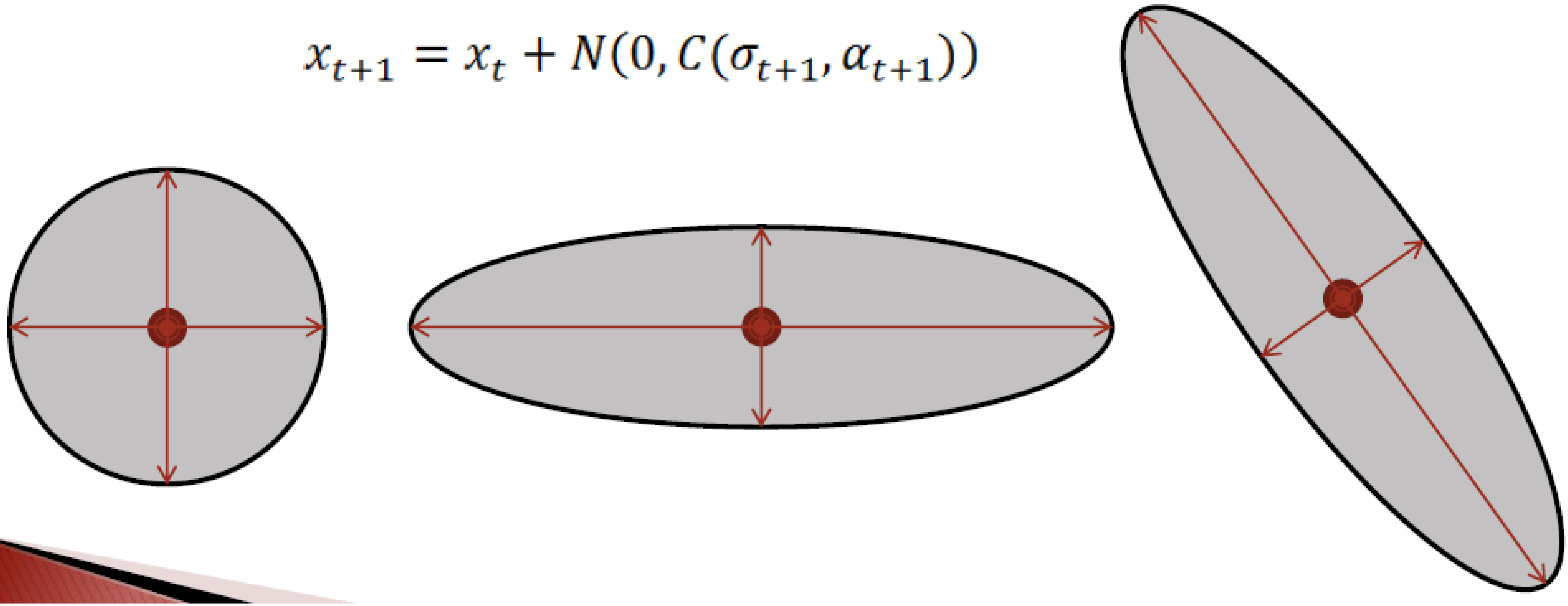
Mutação correlacionada com n desvios padrão e ângulos de rotação:

$$\langle x_t | \sigma_t | \alpha_t \rangle \rightarrow \langle x_{t+1} | \sigma_{t+1} | \alpha_{t+1} \rangle$$

$$\sigma_{t+1,i} = \sigma_{t,i} \exp\left[N(0,\tau_1) + N(0,\tau_2)\right]$$

$$\alpha_{t+1,i} = \alpha_{t,i} + \beta N(0,1)$$

$$x_{t+1} = x_t + N(0, C(\sigma_{t+1}, \alpha_{t+1}))$$



UFMG

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Estratégia (3)

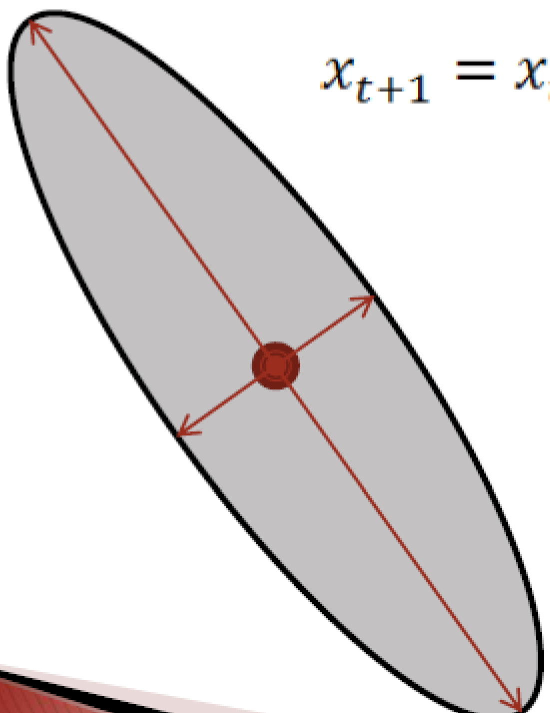
Mutação correlacionada com n desvios padrão e ângulos de rotação:

$$\langle x_t | \sigma_t | \alpha_t \rangle \rightarrow \langle x_{t+1} | \sigma_{t+1} | \alpha_{t+1} \rangle$$

$$\sigma_{t+1,i} = \sigma_{t,i} \exp\left[N(0,\tau_1) + N(0,\tau_2)\right]$$

$$\alpha_{t+1,i} = \alpha_{t,i} + \beta N(0,1)$$

$$x_{t+1} = x_t + N(0, C(\sigma_{t+1}, \alpha_{t+1}))$$



$$c_{ii} = \sigma_i^2$$

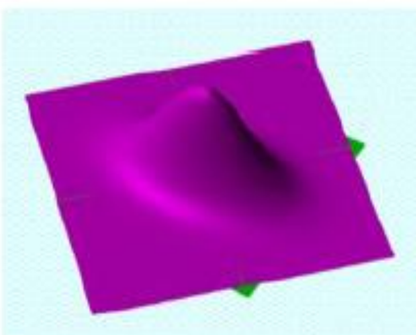
$$c_{ij,i\neq j} = \begin{cases} 0 & \text{sem correlações} \\ 0.5(\sigma_i^2 - \sigma_j^2)\tan(2\alpha_{ij}) & \text{correlações} \end{cases}$$

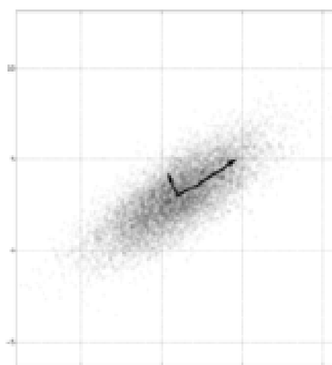
$$\tan(2\alpha_{ij}) = \frac{2c_{ij}}{\sigma_i^2 - \sigma_j^2}$$

$$\beta \approx 5^\circ$$

Gaussiana Multivariada

$$\vec{N}(\vec{\mu}, C\ ) = \quad f_X(\vec{x}) = \frac{1}{(2\pi)^{n/2}|C|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu})' C^{-1}(\vec{x}-\vec{\mu})\right)$$





$\vec{\mu}$ = vetor de médias
C = matriz de covariância
|C| = determinante de C
C^{-1} = inversa de C

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Matriz de Covariância

$$C = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

C: simétrica e definida positiva

σ_{ii}: variâncias de cada variável x_i

σ_{ij}: covariância entre as variáveis x_i e x_j

$\sigma_{ij} > 0 \rightarrow$ Ex: $x_i = \text{comprimento}$ e $x_j = \text{peso}$

$\sigma_{ij} < 0 \rightarrow$ Ex: $x_i = \text{venda de carro}$ e $x_j = \text{indice de desemprego}$

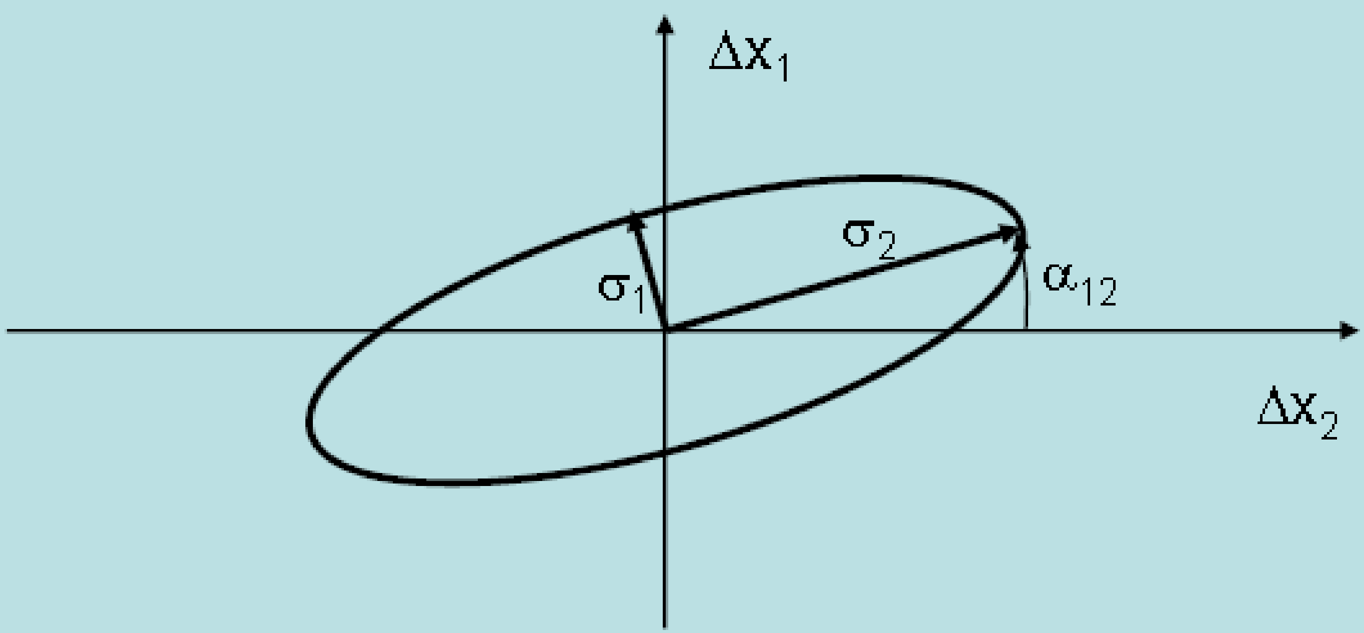
$\sigma_{ij} = 0 \rightarrow x_i$ e x_j são estatisticamente independentes

Se a matriz de covariância é diagonal, então $\vec{N}(\vec{\mu}, C)$ se reduz ao produto de distribuições normais univariadas para cada componente x_i

Matriz de Covariância

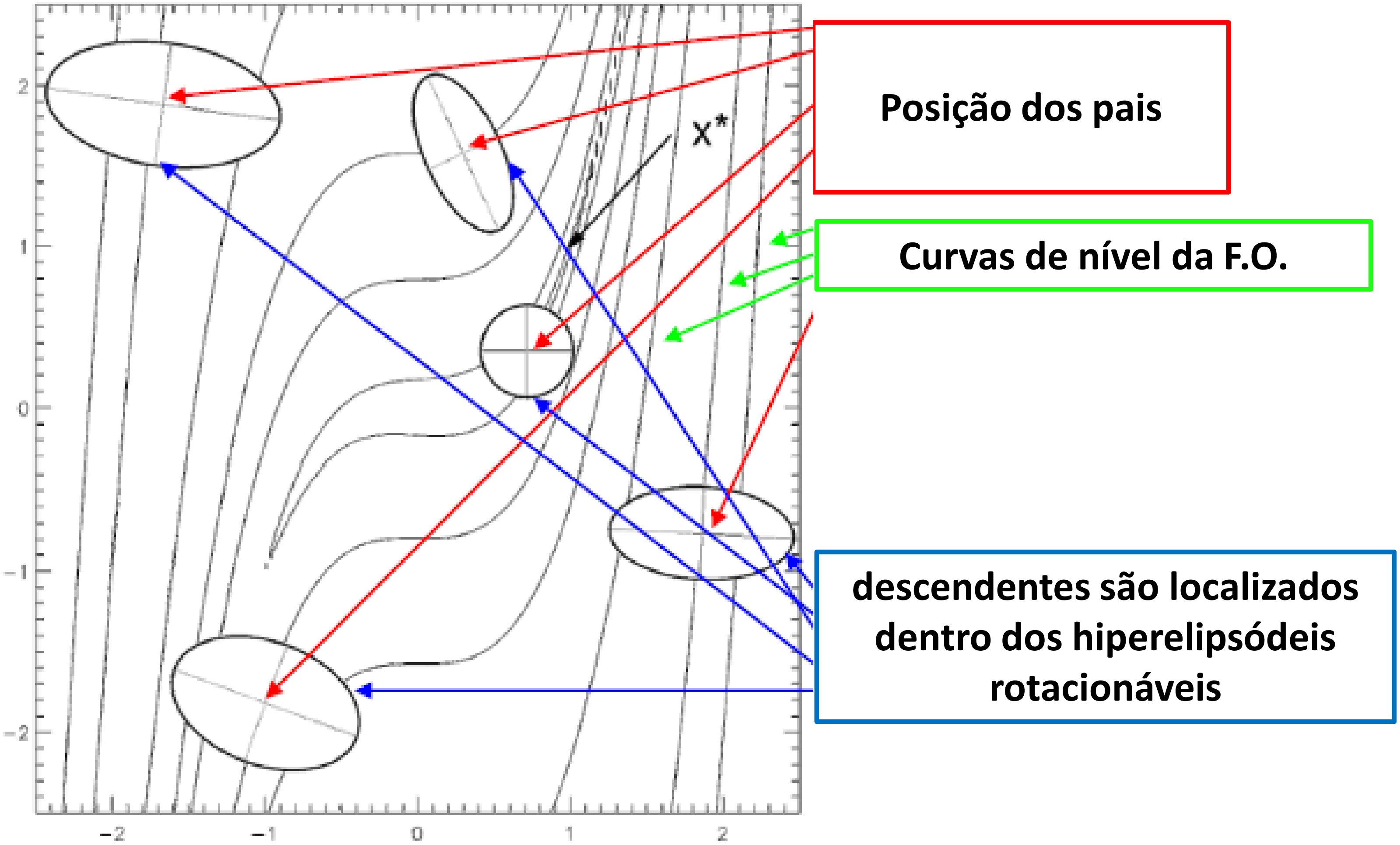
- Interpretação dos ângulos de rotação entre cada par de variáveis:

$$c_{ij(i \neq j)} = \frac{1}{2}(\sigma_i^2 - \sigma_j^2)\tan(2\alpha_{ij})$$



Exemplo Ilustrativo

- **mutação correlacionada com n desvios-padrão e angulos de rotação:**



UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Recombinação

- A recombinação atua tanto nas variáveis de otimização quanto nos parâmetros da estratégia.

- O procedimento de recombinação em EEs difere do cruzamento típico em algoritmos genéticos.

- Notação ES($\mu/\rho,\lambda$) ou ES($\mu/\rho+\lambda$):
  - indica quantos pais são usados na operação de recombinação – ρ pais são selecionados aleatoriamente e formam um filho que, em seguida, sofre mutação.

- **Local recombination**: ($\rho=2$) um filho é criado a partir da recombinação de dois pais selecionados aleatoriamente entre os μ pais presentes na população.

- **Global recombination**: ($2<\rho\leq\mu$) mais de dois pais são selecionados aleatoriamente para produzir um único filho.

- **No recombination**: ($\rho=1$)

UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

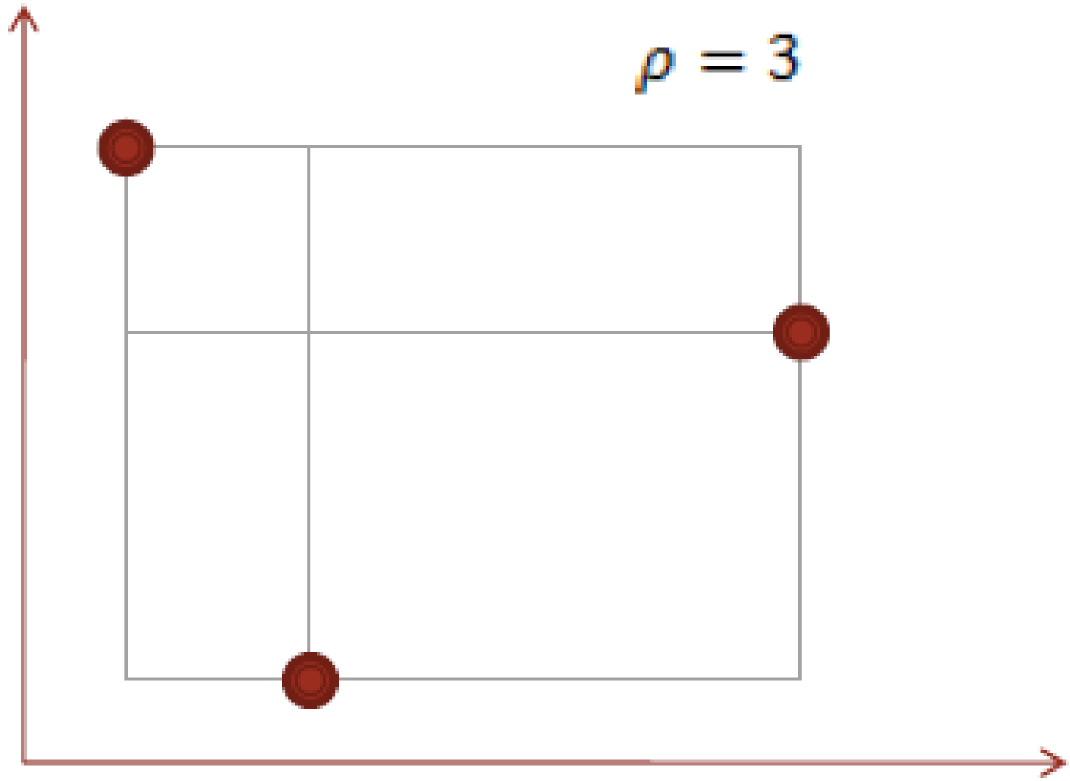
Recombinação

	LOCAL: Two fixed parents	GLOBAL: Multi-parent (more used)
INTERMEDIARY:	Local intermediary	Global intermediary
DISCRETE:	Local discrete	Global discrete

Obs.: Acording to Eiben & Smith:
- Discrete recombination is more used for the variable part of chromossome;
- Intermediary recombination is more used for parameter part of chromossome;

Recombinação Discreta

- **Discrete recombination**: Os alelos dos pais são usados para gerar um filho. Para cada componente que compõe o filho, um dos ρ pais é selecionado aleatoriamente e o componente correspondente é herdado deste.



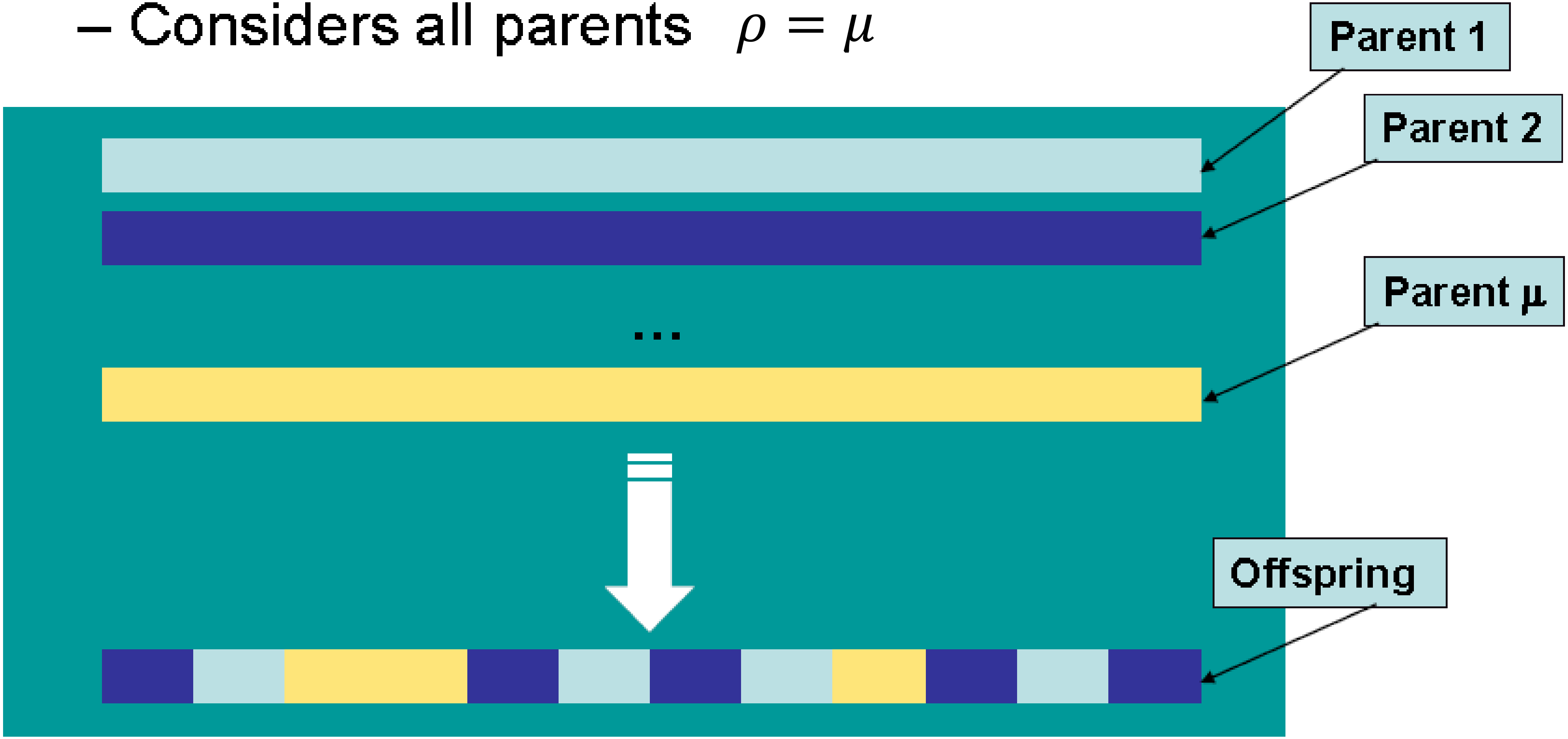
UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Recombinação Discreta Local

– Variable at position i will be copied at random (uniformly distr.) from parent 1 or parent 2, position i.

Recombinação Discreta Global

– Considers all parents $\rho = \mu$



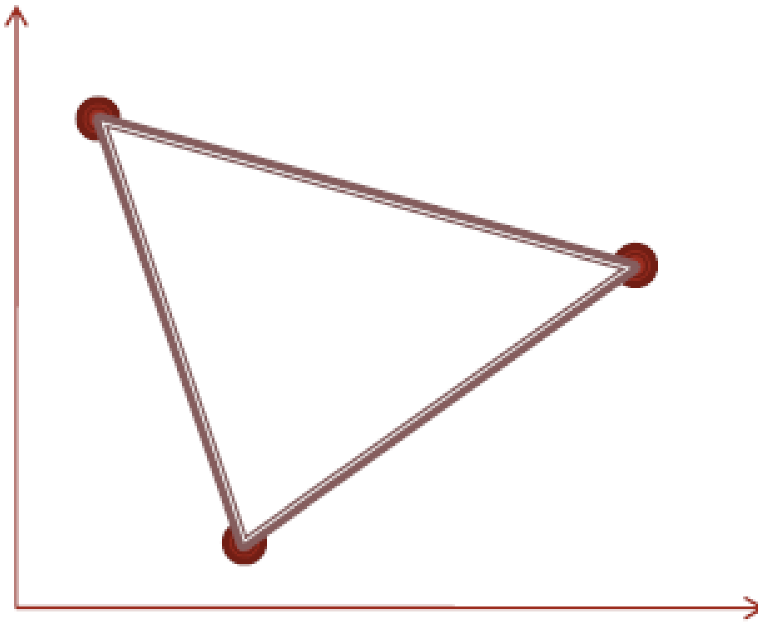
UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Recombinação Intermediária

- **Intermediate recombination**: Os valores médios dos alelos dos pais são usados para gerar um filho.

$$\chi_{t,k} = \langle x_{t,k} | \sigma_{t,k} | \alpha_{t,k} \rangle$$

$$\chi_{t,k} = \left\langle \frac{1}{\rho}\sum_{i=1}^{\rho} x_{t,i} \middle| \frac{1}{\rho}\sum_{i=1}^{\rho} \sigma_{t,i} \middle| \frac{1}{\rho}\sum_{i=1}^{\rho} \alpha_{t,i} \right\rangle$$

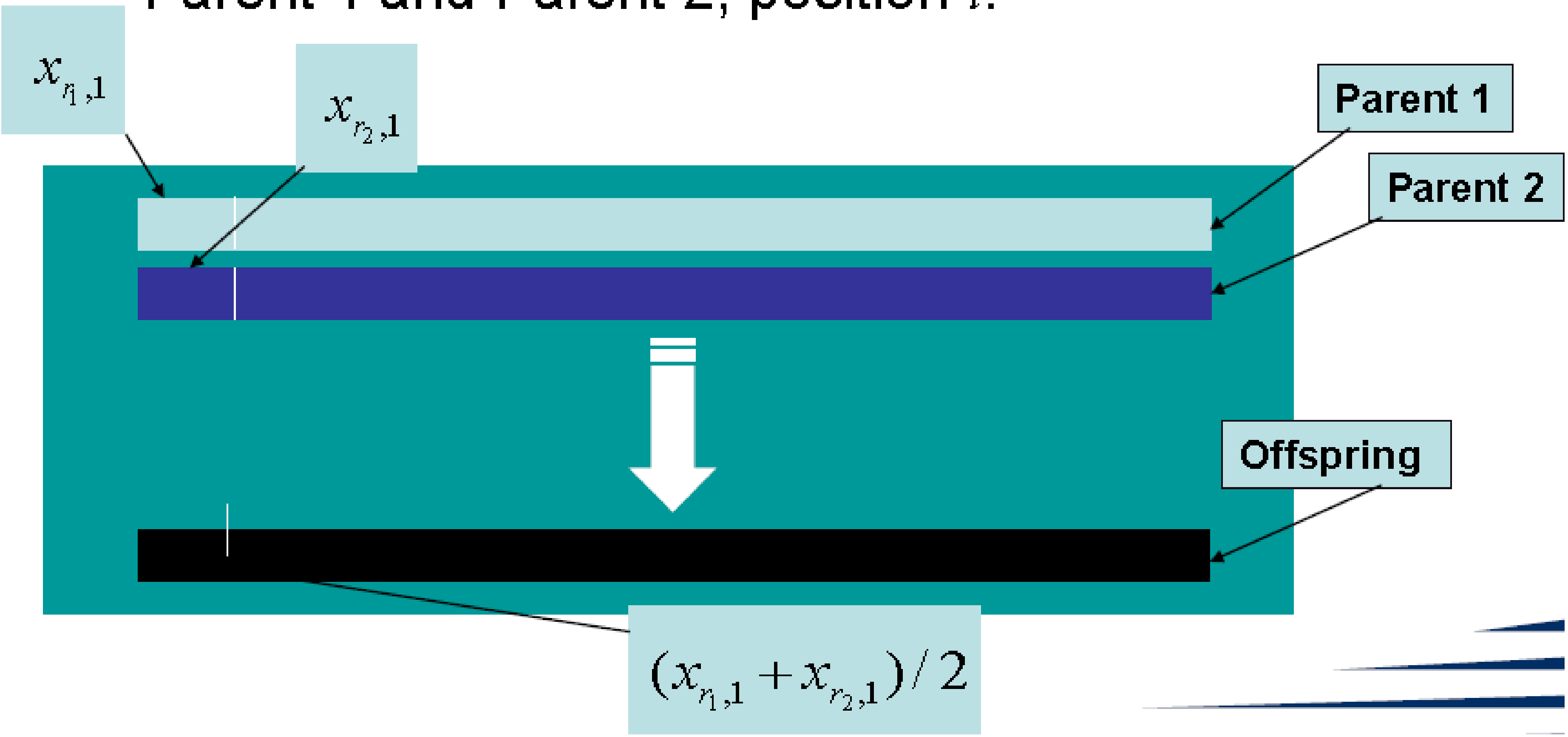


UFMG

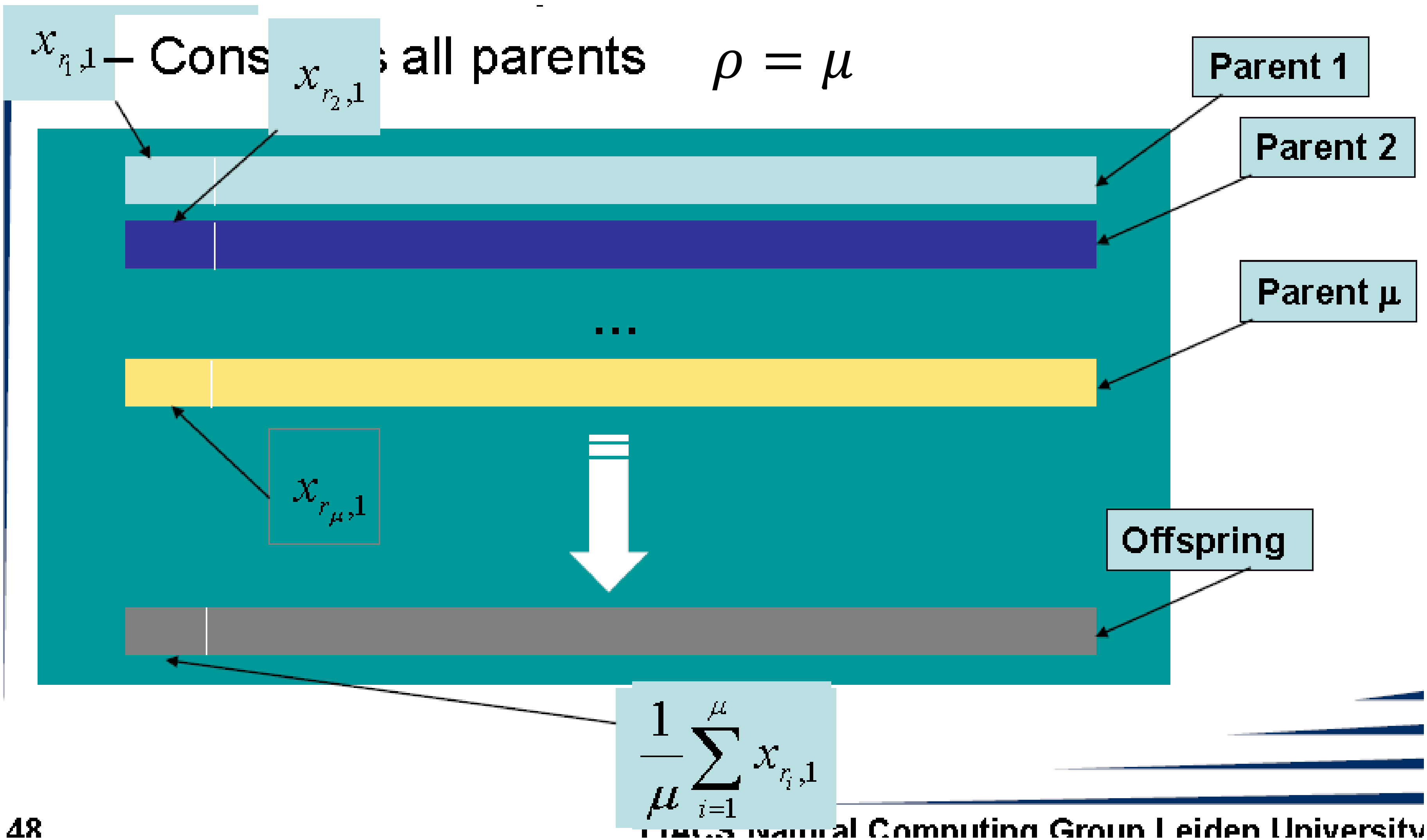
UNIVERSIDADE FEDERAL DE MINAS GERAIS

Recombinação Interm. Local

– Variable at position i is arithmetic mean of Parent 1 and Parent 2, position i.



Recombinação Interm. Global



UFMG
UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Recombinação

Considerações:

- EEs frequentemente usam recombinação global.
- Recombinação discreta é recomendada para as variáveis de otimização:
  - Preserva diversidade na população.
- Recombinação intermediária é recomendada para os parâmetros de otimização:
  - Assegura uma autoadaptação de parâmetros mais cautelosa.

Seleção

- Seleção para reprodução: seleção com distribuição uniforme de probabilidades, ou seja, não polarizada.

- Seleção para sobrevivência: Aplicado aos λ filhos criados após recombinação e mutação. Os μ pais da próxima geração são selecionados:

- (1) entre os λ filhos;
- (2) ou entre a união dos μ pais e dos λ filhos.

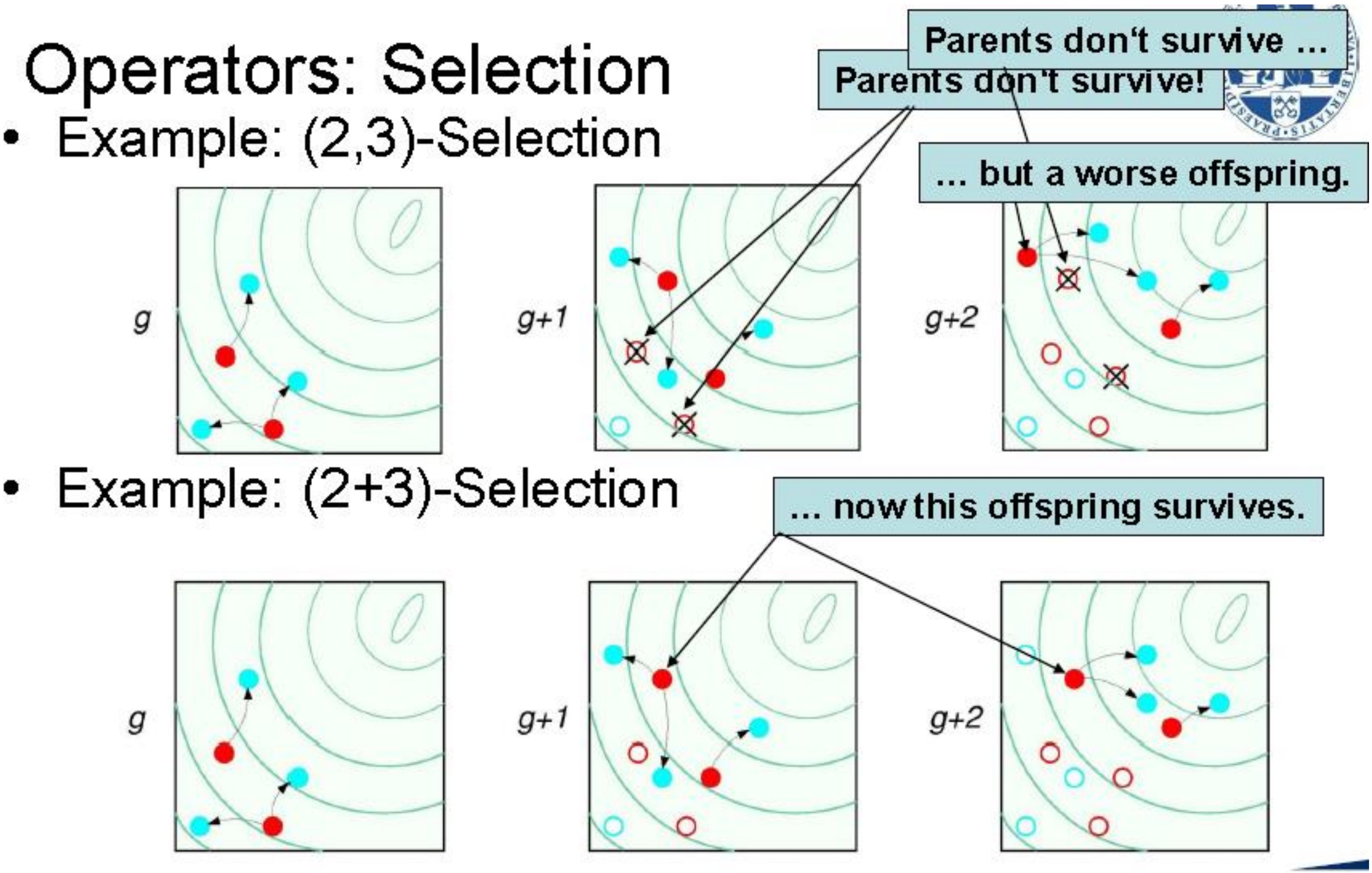
Seleção

- Seleção (μ+λ):
  - μ pais produzem λ filhos por recombinação + mutação
  - os melhores μ indivíduos do total (μ+λ) são selecionados para sobreviver
    - seleção determinística baseada em *fitness (ranking).*
  - garante que não haja deterioração:
    - o melhor indivíduo na geração *t* sempre estará na geração *t+1* (Estratégia Elitista).

Seleção

- Seleção (μ,λ):
  - μ pais produzem $\lambda >> \mu$ filhos por
    - recombinação + mutação
  - os melhores μ indivíduos de λ descendentes são selecionados para sobreviver:
    - seleção determinística baseada em *fitness.*
  - não garante comportamento monotônico do *fitness:*
    - o melhor valor de *fitness* na geração t+1 pode ser pior que na geração t;

Exemplo Ilustrativo



Seleção

- em geral, (μ,λ) é preferida sobre $(\mu+\lambda)$ devido as seguintes razões:
  - (μ,λ) descarta todos os pais e por isso ela é mais propensa a escapar de ótimos locais. Isso é vantajoso em superfícies multimodais;
  - Se a funções objetivo é dinâmica (muda no tempo), $(\mu+\lambda)$ preserva soluções desatualizadas, não sendo capaz de acompanhar a dinâmica da superfície;
  - $(\mu+\lambda)$ tende a dificultar o mecanismo de coevolução, pois parâmetros mal adaptados podem sobreviver por um número grande de gerações quando um indivíduo possui boas variáveis de decisão e parâmetros de estratégia ruins.

UFMG

UNIVERSIDADE FEDERAL
DE MINAS GERAIS

Seleção

- A relação λ/μ controla a pressão seletiva, e quanto maior a relação, maior a pressão seletiva para sobrevivência.

- Em geral recomenda-se $\lambda/\mu = 7$ (e.g., $\mu = 15$ e $\lambda = 100$).

Pseudocódigo

```
Criar uma população inicial X[t];
Inicializar os parâmetros de estratégia;

Enquanto critério de parada não satisfeito faça:
        para i=1,...,λ faça
                Escolha ρ pais aleatoriamente;
                Faça recombinação para gerar z;
                Faça mutação nos parâmetros da estratégia;
                z ← z + mutação;
                Avalie f(z);
        fimpara
        Aplique seleção ES(μ/ρ,κ,λ)
        t ← t+1;
Fimenquanto
```

Referências

- **Leitura Recomendada**:
  - Capítulo 4 do Livro:

  A.E. EIBEN, J.E. SMITH, Introduction to Evolutionary Computing (Natural Computing Series), Springer.