Avaliação do Desempenho de Optimizadores Estocásticos

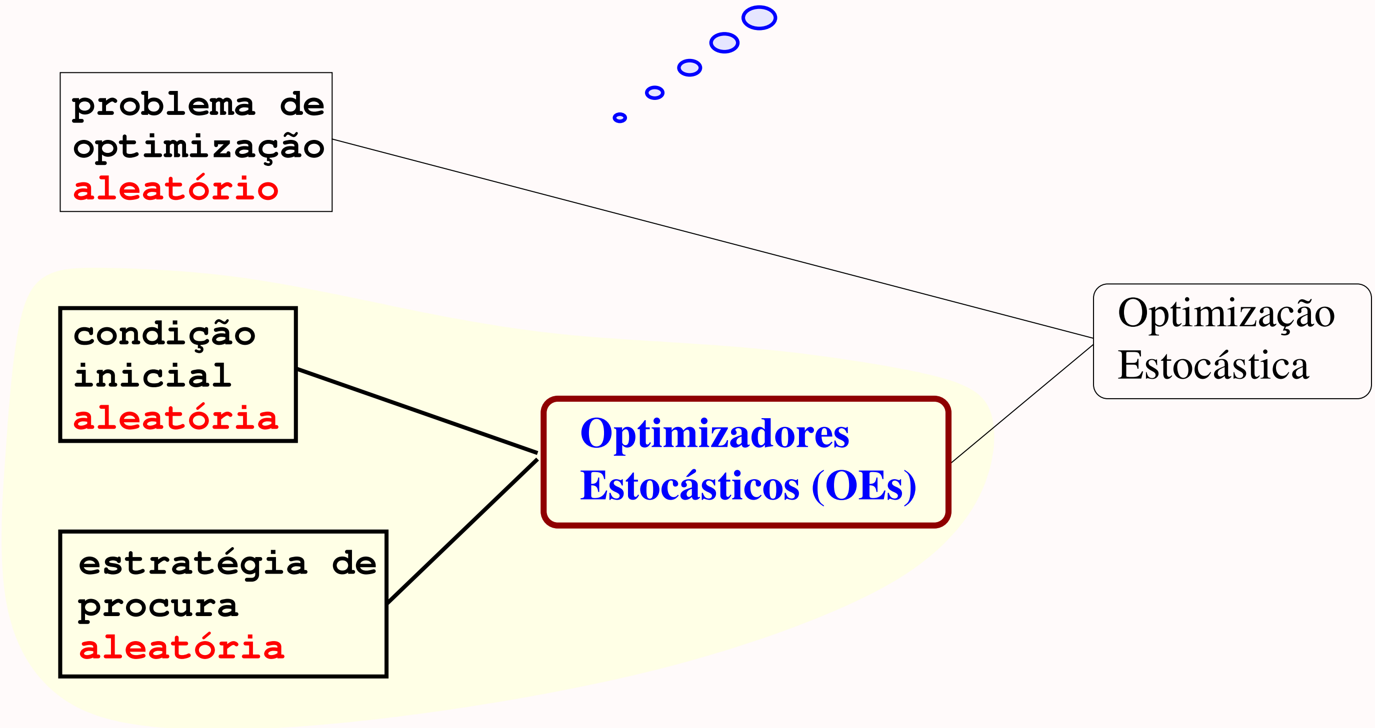
Viviane Grunert da Fonseca

CISUC, Centro de Informática e Sistemas da Universidade de Coimbra

vivianef@dei.uc.pt

4ª Escola Luso-Brasileira de Computação Evolutiva,
Coimbra, 13 de Julho de 2013

Avaliação do Desempenho de **Optimizadores Estocásticos**



Exemplos: algoritmos genéticos, estratégias evolutivas, algoritmos de colónias de formigas, simulated annealing, etc.

Avaliação do $\boxed{\textbf{Desempenho}}$ de Optimizadores Estocásticos

→ O **desempenho** de OEs tem que ver com

- a qualidade dos seus resultados, e
- o tempo que é necessário para produzir esses resultados.

(nº de avaliações da função, tempo de CPU, tempo decorrido, etc.)

→ Há grande interesse na avaliação do desempenho de OEs. Por exemplo, é comum querer escolher o "melhor" optimizador de entre várias alternativas.

→ O que significa $\boxed{\textbf{bom desempenho}}$ de um OE?

- produzir resultados em pouco tempo,
- resultados de boa qualidade, ou seja, resultados "próximos" do(s) valor(es) óptimo(s) da função objectivo.

Avaliação do **Desempenho** de Optimizadores Estocásticos

Questões:

(a) Não se conhece o óptimo no espaço de decisão, nem o(s) valor(es) óptimo(s) da função objectivo!

$\rightsquigarrow$ Problemas de minimização: resultados melhores $\equiv$ valores menores

(b) O desempenho do mesmo OE pode variar em função ...

- da instância do problema de optimização,
- da condição inicial escolhida,
- dos valores dos parâmetros do OE,
- etc.

$\rightsquigarrow$ Delineamento experimental

Avaliação do $\boxed{\textbf{Desempenho}}$ de Optimizadores Estocásticos

Para a avaliação do desempenho consideramos **dois cenários**:

$\boxed{\textbf{Cenário 1}}$ (critério de paragem $\rightleftarrows$ qualidade dos resultados):

Desempenho $\equiv$ Tempo de Execução

onde o tempo de execução é aleatório e é representado por ...

* uma variável aleatória X em $\mathbb{R}^{\geq 0}$
* com distribuição de probabilidade $F_X = P(X \leq x)$.

$\Longrightarrow$ Estudar distribuições **univariadas**.

Avaliação do **Desempenho** de Optimizadores Estocásticos

Cenário 2 (critério de paragem $\rightleftarrows$ tempo de execução):

Desempenho $\equiv$ **Qualidade dos Resultados**

aleatórios !!

$\Longrightarrow$ Estudar distribuições de probabilidades

. . . **univariadas** $\longleftrightarrow$ optimização mono-objectivo

. . . **multivariadas** ou **ainda mais complexas**

$\longleftrightarrow$ optimização multi-objectivo

$\rightsquigarrow$ Tema da presente aula!

Estrutura da aula

- Optimizadores mono-objectivo:
  - *Distribuições de resultados e desempenho*
- Optimizadores multi-objectivo:
  - *Distribuições de resultados*
  - *A abordagem da função de aproveitamento*
  - *A abordagem dos indicadores de qualidade*
- Notas finais

Optimizadores mono-objectivo

⚠ Vamos considerar os resultados de OEs produzidos no <u>espaço dos objectivos</u> após um <u>determinado tempo de execução</u>, i.e. Cenário 2.

⚠ No entanto, a discussão que se segue também se aplica a qualquer distribuição de tempo de execução, i.e. Cenário 1.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

- O valor mínimo da função objectivo é um escalar $x^* \in \mathbb{R}$.
- Cada execução do OE produz um resultado aleatório que é representado por uma variável aleatória X em $\mathbb{R}$ com realizações $z \in [x^*, \infty)$.
- A distribuição de resultados é uma distribuição *univariada*, discreta ou contínua, com suporte em $\mathbb{R}^{\geq x^*}$ que pode ser caracterizada pela **função de distribuição**

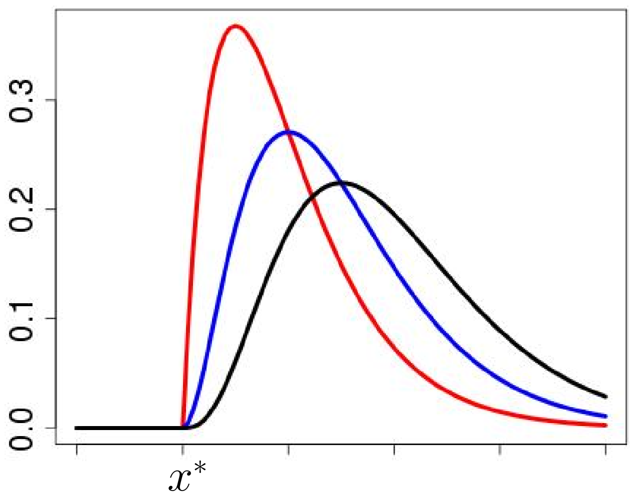
$$F_X(z) = P(X \leq z)$$

onde $F_X(z) = 0$ se $z \in (-\infty, x^*)$.

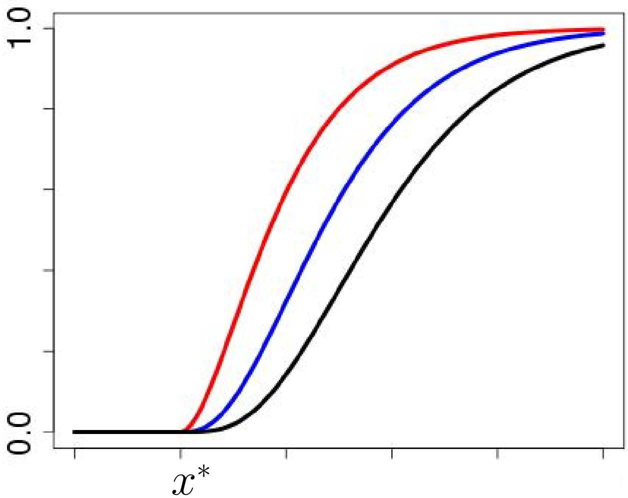
Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

- Os resultados são de <u>melhor</u> qualidade quando a sua distribuição se <u>concentra mais perto</u> do valor mínimo $x^* \in \mathbb{R}$ da função objectivo ...



função de densidade



função de distribuição

$\rightsquigarrow$ Os resultados com a distribuição a vermelho são de <u>melhor</u> qualidade.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

. . . , ou seja, quanto . . .

(a) <u>menor</u> no sentido estocástico for a variável aleatória X, ou ainda, quanto <u>maior</u> $F_X(z)$ para todo o $z \geq x^*$.

$$\left[\text{ caso “ideal”: } F_X(z) = \boldsymbol{I}\{z \geq x^*\} \right]$$

(b) <u>menor</u> o valor de uma **medida de localização**, tal como:

- *média*: $\mu = \mathrm{E}(X) \geq x^*$
- *mediana*: $x_{0.5} \geq x^*$ onde $P(X \leq x_{0.5}) = 0.5$
- *α-quantil*: $x_\alpha \geq x^*$ onde $P(X \leq x_\alpha) = \alpha$

$$\left[\text{ casos “ideais”: } \mu = x_{0.5} = x_\alpha = x^* \right]$$

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

(c) menor o valor de uma **medida de dispersão**, tal como:

- *variância*: $\sigma^2 = \mathrm{Var}(X) \geq 0$
- *amplitude interquartil*: $x_{0.75} - x_{0.25} \geq 0$

$$\left[\text{ casos “ideais”: } \sigma^2 = x_{0.75} - x_{0.25} = 0 \ \right]$$

⚠ Não faz sentido considerar o critério (c) sem ter em conta o critério (b).

⚠ A média é o *1º momento não-centrado* de uma distribuição e a variância é o *2º momento centrado* de uma distribuição.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

Problema

$F_X(\cdot)$ e as medidas de localização/dispersão são desconhecidas!

$\boxed{\textbf{Análise inferencial – Estimação}}$

A função $F_X(\cdot)$ e as medidas de localização/dispersão podem ser estimadas a partir de ...

– uma *amostra aleatória (simples)* $X_1, X_2, \ldots X_n$, i.e.

– n variáveis aleatórias independentes com a mesma distribuição, que representam os resultados de n execuções independentes do optimizador.

Optimizadores mono-objectivo:
Distribuições de resultados e desempenho

Consideram-se, por exemplo, os seguintes estimadores:

- *função de distribuição empírica*:

$$F_n(z) = \frac{1}{n} \cdot \sum_{i=1}^{n} \boldsymbol{I}\{X_i \leq z\}$$

- *média amostral*: $\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^{n} X_i$
- *mediana amostral*:

$$Q_{0.5} = \begin{cases} \frac{1}{2} \cdot \left[X_{(\frac{n}{2})} + X_{(\frac{n+2}{2})}\right] & \text{se } n \text{ é par} \\ X_{(\frac{n+1}{2})} & \text{se } n \text{ é ímpar,} \end{cases}$$

onde $X_{(1)}, \ldots, X_{(n)}$ são as variáveis aleatórias da amostra ordenada.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

- *α-quantil amostral*:

$$Q_\alpha = \begin{cases} \frac{1}{2} \cdot \left[X_{(n\cdot\alpha)} + X_{(n\cdot\alpha+1)}\right] & \text{se } n\cdot\alpha \in \mathbb{N} \\ X_{(\lceil n\cdot\alpha\rceil)} & \text{se } n\cdot\alpha \notin \mathbb{N} \end{cases}$$

onde $X_{(1)}, \ldots, X_{(n)}$ são as variáveis aleatórias da amostra ordenada.

- *variância amostral*: $S^2 = \dfrac{1}{n-1} \cdot \displaystyle\sum_{i=1}^{n} (X_i - \bar{X})^2$

- *amplitude interquartil amostral*: $Q_{0.75} - Q_{0.25}$

⚠ Como a distribuição de resultados deve ser **enviesada para a direita**, pode ser preferível considerar a <u>mediana</u> amostral em vez da média amostral e a <u>amplitude interquartil</u> amostral em vez da variância amostral.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

Testes de hipóteses

Frequentemente, o objectivo é ...

$\boxed{\text{comparar o desempenho de OEs}}$

e, <u>com “sorte”</u> ...

$\boxed{\text{afirmar com certo grau de confiança}}$

que

- o desempenho de dois ou mais OEs é diferente *(nalgum sentido)*.
- um OE tem melhor desempenho que outro OE *(nalgum sentido)*.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

Para esse efeito, considere-se um dos seguintes problemas de teste:

$\boldsymbol{H_0}$ **:** O desempenho de dois ou mais OEs é igual (nalgum sentido).

v.s.

$\boldsymbol{H_1}$ **:** *O desempenho de dois ou mais OEs é diferente (nalgum sentido).*

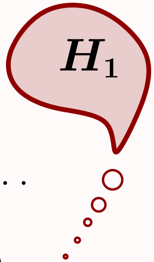
$\boldsymbol{H_0}$ **:** O desempenho de OE-A não é melhor que o de OE-B (nalgum sentido).

v.s.

$\boldsymbol{H_1}$ **:** *O desempenho de OE-A é melhor que o de OE-B (nalgum sentido).*

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho



"Com sorte", há evidência estatisticamente significativa de que ...

- o desempenho de dois ou mais OEs é diferente *(nalgum sentido)*.
- O desempenho de OE-A é melhor que o de OE-B *(nalgum sentido)*.

Isso é: $P(\text{decisão "rejeitar } H_0\text{" é errada}) \leq \alpha, \quad \alpha \in \{0.01, 0.05\}$

"Com menos sorte", *não se pode rejeitar* H_0 !

... **não** há tal evidência estatística.

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

Testes de hipóteses para OEs *"independentes"*:

Teste	**compara**	**nº de OEs**
Kolmogorov-Smirnov	função de distribuição	dois ou mais
Mann-Whitney	mediana	dois
Kruskall-Wallis	mediana	três ou mais
Brown-Forsythe	variância	dois ou mais

Testes de hipóteses para OEs *"dependentes"*:

Teste	**compara**	**nº de OEs**
Wilcoxon	mediana	dois
Friedman	mediana	três ou mais

Optimizadores mono-objectivo:

Distribuições de resultados e desempenho

- **Comparações múltiplas** com $\boxed{\text{três ou mais OEs}}$,

  - *"entre todos os pares"*,

  - *"com um controlo"* (OE de referência),

  - *"com o melhor"*,

  requerem testes múltiplos com ajustamento de α.

- Considerar diferentes ...

  - instâncias do problema de optimização

  - parâmetros do optimizador,

  - critérios de paragem

  requer **métodos de delineamento experimental**.

Optimizadores multi-objectivo

⚠ Vamos considerar os resultados de OEs produzidos no espaço dos objectivos após um determinado tempo de execução, i.e. Cenário 2.

Optimizadores multi-objectivo: Distribuições de resultados

- Os valores mínimos (não-dominados) da função objectivo dão origem a uma *fronteira óptima (de Pareto)* $\mathcal{X}^*$ em $\mathbb{R}^d$.
- Em cada execução, o optimizador produz múltiplos resultados aleatórios em $\mathbb{R}^d$ que se representam por um **conjunto aleatório de pontos não-dominados** (conjunto APN)

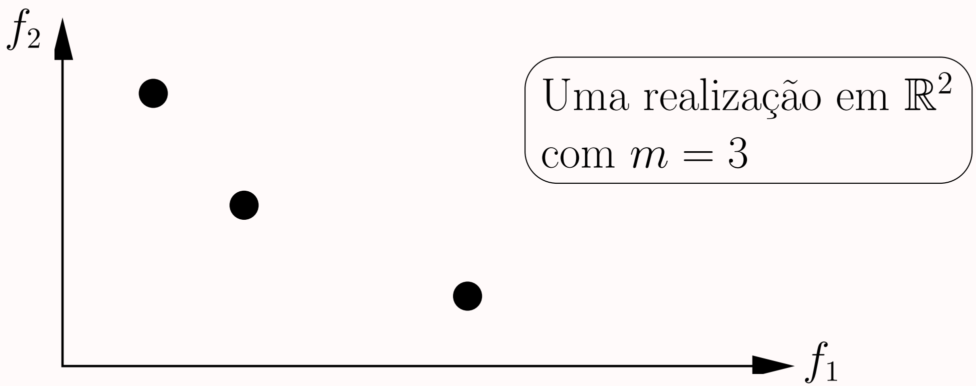
$$\mathcal{X} = \{X_1, \dots, X_M \in \mathbb{R}^d : \; P(X_i \leq X_j) = 0, \; i \neq j\},$$

sendo

– $X_1, X_2, \dots$ vectores aleatórios em $\mathbb{R}^d$ e

– M uma variável aleatória em $\mathbb{N}$.

Optimizadores multi-objectivo: Distribuições de resultados



Características de um conjunto APN $\mathcal{X}$:

$\rightarrow$ Os pontos (vectores) $x_1, x_2, \ldots x_m$ de uma realização de $\mathcal{X}$ são não-dominados no sentido de Pareto.

$\rightarrow$ Os vectores aleatórios $X_1, X_2, \ldots X_M$ são dependentes.

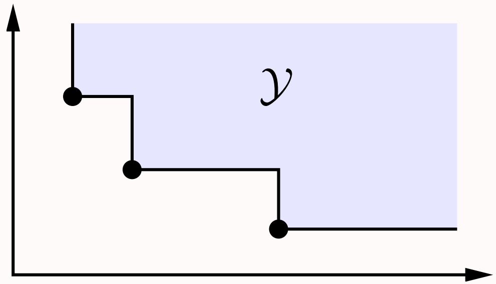
$\rightarrow$ $\mathcal{X}$ é fechado e é "não-estacionário".

Optimizadores multi-objectivo: Distribuições de resultados

Representações alternativas de um conjunto APN $\mathcal{X}$:

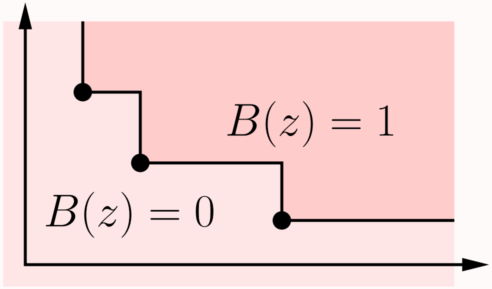
→ o conjunto aleatório fechado
(*"conjunto atingido"*)

$$\mathcal{Y} = \{z \in \mathbb{R}^d \mid X_1 \leq z \vee \ldots \vee X_M \leq z\}$$



→ o campo aleatório binário

$$\{B(z), z \in \mathbb{R}^d\} = \{\boldsymbol{I}\{z \in \mathcal{Y}\}, z \in \mathbb{R}^d\}$$



Optimizadores multi-objectivo: Distribuições de resultados

- A distribuição de resultados é bastante complexa pois compreende
  - as distribuições multivariadas (idênticas) de $X_1, X_2, \ldots$
  - a distribuição univariada discreta de M, e
  - a dependência entre $X_1, X_2, \ldots$.

Dado $M \leq m^*$, ou seja, dado que o optimizador não produz mais que m^* resultados por execução, a distribuição de $\mathcal{X}$ pode ser caracterizada pela **função de aproveitamento de ordem $\boldsymbol{m^*}$**

$$\boxed{\alpha_{\mathcal{X}}^{(m^*)}(z_1, \ldots, z_{m^*}) = P\Big(\mathcal{X} \trianglelefteq z_1 \ \wedge \ \ldots \ \wedge \ \mathcal{X} \trianglelefteq z_{m^*}\Big)}$$

"atinge" / "domina fracamente"

onde

$$[\mathcal{X} \trianglelefteq z] \quad \text{significa} \ \big[X_1 \leq z \ \vee \ X_2 \leq z \ \vee \ldots \vee \ X_M \leq z\big].$$

Optimizadores multi-objectivo: Distribuições de resultados

Por palavras, $\alpha_{\mathcal{X}}^{(m^*)}(z_1, \ldots, z_{m^*})$ é ...

"a probabilidade de atingir todas as metas $z_1, \ldots, z_{m^*}$ numa única execução do optimizador".

- Os resultados do optimizador são de melhor qualidade quando a sua distribuição se concentra mais perto da fronteira óptima $\mathcal{X}^*$ em $\mathbb{R}^d$,

... , ou seja, quanto ...

  menor no sentido estocástico for o conjunto APN $\mathcal{X}$, ou ainda, quanto maior $\alpha_{\mathcal{X}}^{(m^*)}(z_1, \ldots, z_{m^*})$ para todo o vector $z_i \in \mathbb{R}^d$, $\mathcal{X}^* \trianglelefteq z_i$.

Optimizadores multi-objectivo: Distribuições de resultados

A complexidade da distribuição dos resultados ...

– torna difícil a sua representação gráfica.

$\rightsquigarrow$ No caso $d = 2$ a função $\alpha_{\mathcal{X}}^{(2)}(z_1, z_2)$ já depende de 2×2 variáveis!

– deu origem a muitos critérios de desempenho para optimizadores multi-objectivo!

$\boxed{\text{Duas principais abordagens}}$

1. **Abordagem da função de aproveitamento** que descreve a distribuição do conjunto APN $\mathcal{X}$ directamente através de uma hierarquia de funções encaixadas, cada vez mais informativas.
2. **Abordagem dos indicadores de qualidade** que transforma o conjunto $\mathcal{X}$ numa variável aleatória em $\mathbb{R}$ e estuda a respectiva distribuição univariada, (geralmente) em termos da sua média.

A abordagem da função de aproveitamento

Optimizadores multi-objectivo: A abordagem da função de aproveitamento

- O desempenho de um optimizador multi-objectivo é avaliado na sua totalidade através de $\alpha_{\mathcal{X}}^{(m^*)}(z_1, \ldots, z_{m^*})$, onde m^* é o número máximo de resultados X_i por execução.

- Uma descrição de tipo "média" da distribuição de $\mathcal{X}$ é possível com a **função de aproveitamento (de ordem 1)**

$$\alpha_{\mathcal{X}}^{(1)}(z) = \alpha_{\mathcal{X}}(z) = P\big(\mathcal{X} \trianglelefteq z\big)$$
$$= P\big(X_1 \leq z \ \vee \ X_2 \leq z \ \vee \ldots \vee \ X_M \leq z\big),$$

ou seja, com a

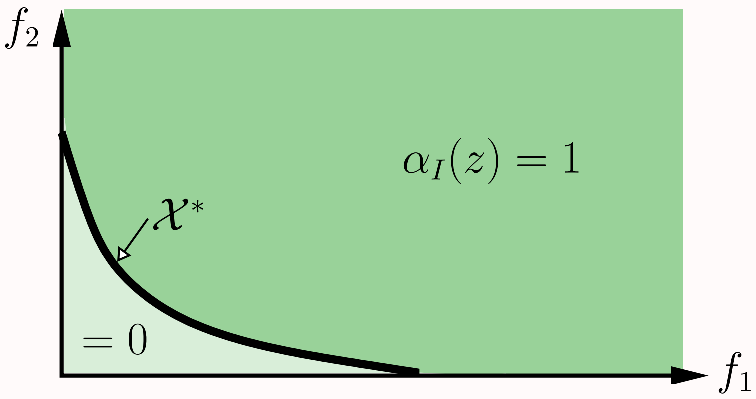
"*probabilidade* de atingir cada meta $z \in \mathbb{R}^d$ numa único execução do optimizador".

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

⚠ Tal como a média μ, a função $\alpha_{\mathcal{X}}(\cdot)$ <u>não</u> descreve a distribuição de $\mathcal{X}$ <u>na sua totalidade</u> — a não ser que $\mathcal{X} = \{X\}$. Nesse caso, $\alpha_{\mathcal{X}}(\cdot)$ reduz-se a uma função de distribuição $F_X(\cdot)$!

→ O caso ideal quanto ao desempenho em termos de $\alpha_{\mathcal{X}}(\cdot)$ é a *função de aproveitamento ideal*: $\alpha_I(z) = \boldsymbol{I}\{\mathcal{X}^* \trianglelefteq z\}$.

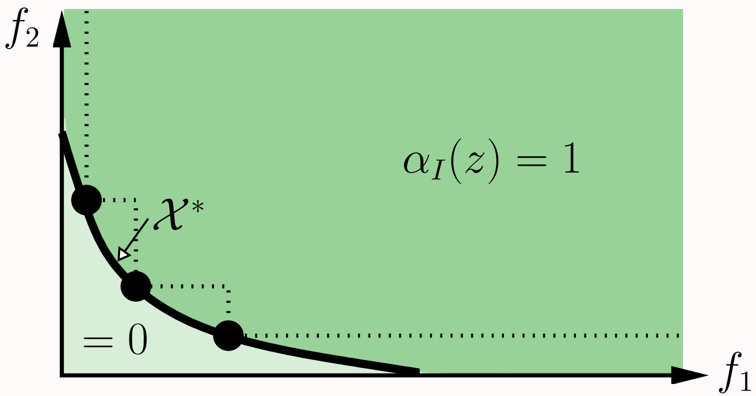


Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

⚠ Tal como a média μ, a função $\alpha_{\mathcal{X}}(\cdot)$ não descreve a distribuição de $\mathcal{X}$ na sua totalidade — a não ser que $\mathcal{X} = \{X\}$. Nesse caso, $\alpha_{\mathcal{X}}(\cdot)$ reduz-se a uma função de distribuição $F_X(\cdot)$!

$\rightarrow$ O caso ideal quanto ao desempenho em termos de $\alpha_{\mathcal{X}}(\cdot)$ é a *função de aproveitamento ideal*: $\alpha_I(z) = \boldsymbol{I}\{\mathcal{X}^* \trianglelefteq z\}$.

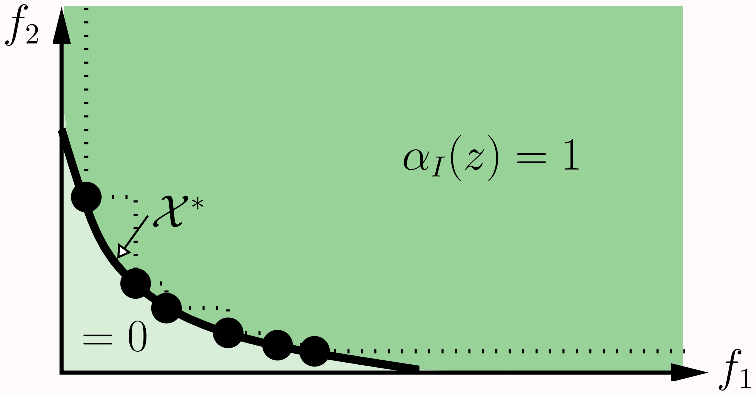


Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

⚠ Tal como a média μ, a função $\alpha_{\mathcal{X}}(\cdot)$ <u>não</u> descreve a distribuição de $\mathcal{X}$ <u>na sua totalidade</u> — a não ser que $\mathcal{X} = \{X\}$. Nesse caso, $\alpha_{\mathcal{X}}(\cdot)$ reduz-se a uma função de distribuição $F_X(\cdot)$!

$\rightarrow$ O caso ideal quanto ao desempenho em termos de $\alpha_{\mathcal{X}}(\cdot)$ é a *função de aproveitamento ideal*: $\alpha_I(z) = \boldsymbol{I}\{\mathcal{X}^* \trianglelefteq z\}$.

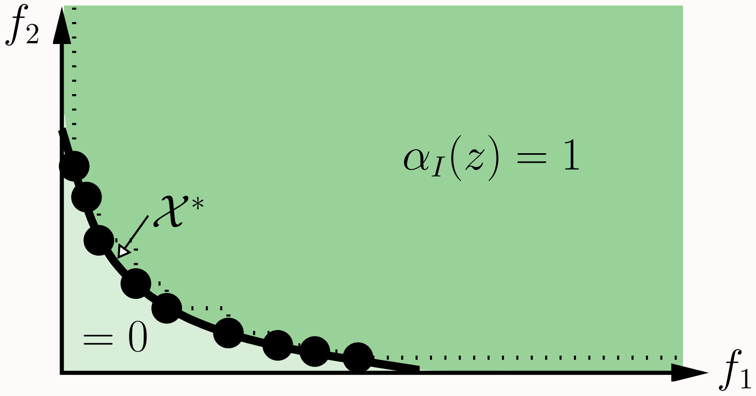


Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

⚠ Tal como a média μ, a função $\alpha_{\mathcal{X}}(\cdot)$ <u>não</u> descreve a distribuição de $\mathcal{X}$ <u>na sua totalidade</u> — a não ser que $\mathcal{X} = \{X\}$. Nesse caso, $\alpha_{\mathcal{X}}(\cdot)$ reduz-se a uma função de distribuição $F_X(\cdot)$!

→ O caso ideal quanto ao desempenho em termos de $\alpha_{\mathcal{X}}(\cdot)$ é a *função de aproveitamento ideal*: $\alpha_I(z) = \boldsymbol{I}\{\mathcal{X}^* \trianglelefteq z\}$.



Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

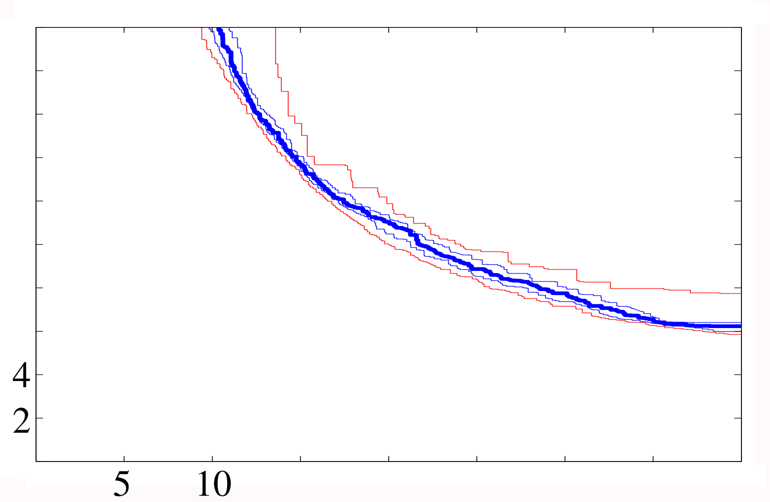
→ $\boxed{\text{Estimação}}$

A função $\alpha_{\mathcal{X}}(\cdot)$, que é <u>desconhecida</u>, pode ser estimada através da *função de aproveitamento empírica*

$$\alpha_n(z) = \frac{1}{n} \cdot \sum_{i=1}^{n} \boldsymbol{I}\{\mathcal{X}_i \trianglelefteq z\}$$

<u>Ilustração</u> de um caso bi-objectivo ($n = 21$):

Curvas de nível ϵ, 0.25, 0.5, 0.75, e $1 - \epsilon$



Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

→ $\boxed{\text{Teste de hipóteses}}$

Um teste de tipo "*Kolmogorov-Smirnov*" permite comparar duas funções de aproveitamento de dois optimizadores A e B.

– Problema de teste (bilateral):

$$H_0 : \; \alpha_{\mathcal{X}_A}(z) = \alpha_{\mathcal{X}_B}(z) \quad \text{para todo o } z \in \mathbb{R}^d$$

v.s.

$$H_1 : \; \alpha_{\mathcal{X}_A}(z) \neq \alpha_{\mathcal{X}_B}(z) \quad \text{para pelo menos um } z \in \mathbb{R}^d,$$

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

– Regra de decisão do teste bilateral: Para o nível de significância $\alpha \in (0,1)$, rejeitar H_0 se

$$D_{n,m} = \sup_{z \in \mathbb{R}^d} \left| \alpha_n^A(z) - \alpha_m^B(z) \right| > d_{n;m;1-\alpha},$$

onde $d_{n;m;1-\alpha}$ pode ser aproximado através de simulação (*teste de permutações*).

⚠ Também é possível formular testes unilaterais.

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

- Uma descrição de <u>tipo “mediana”</u> da distribuição de $\mathcal{X}$ é possível através do conjunto

$$V_{0.5} = \{z \in \mathbb{R}^d \mid \alpha_{\mathcal{X}}(z) \geq 0.5\}$$

(*mediana Vorob’ev* do conjunto atingido)

$$\left[\text{caso ideal: } V_{0.5} = \{z \in \mathbb{R}^d \mid \mathcal{X}^* \trianglelefteq z\} \right]$$

<u>Não</u> proporciona mais informação do que a função de aproveitamento $\alpha_{\mathcal{X}}(\cdot)$.

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

- A descrição de tipo "variância" com a *função de variância*

$$\mathrm{Var}_{\mathcal{X}}(z) = \alpha_{\mathcal{X}}(z) - [\alpha_{\mathcal{X}}(z)]^2$$

também não fornece informação adicional, porque é determinada exclusivamente pela função de aproveitamento $\alpha_{\mathcal{X}}(\cdot)$.

- Mais informação sobre a distribuição de $\mathcal{X}$ e, portanto, sobre o desempenho de um optimizador multi-objectivo
  - diz respeito à dependência entre $X_1, X_2, \ldots X_M$, e
  - é capturada pelas funções de aproveitamento de ordem ≥ 2.

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

<u>Dúvida</u>

– Existem <u>muitas</u> funções de aproveitamento de ordem ≥ 2, (no total: $m^* - 1$)

– e a sua complexidade aumenta gradualmente! ...

➽ ... $\boxed{\text{Até que ordem vale a pena considerá-las?}}$

<u>Resposta</u>

Não é claro!

<u>No entanto</u>, é certamente interessante considerar a chamada *função de covariância.*

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

- Uma descrição de tipo “covariância” da distribuição do conjunto APN $\mathcal{X}$, baseada na função de aproveitamento de **2ª ordem**, é possível através da *função de covariância*

$$\mathrm{Cov}_{\mathcal{X}}(z_1, z_2) = \alpha^{(2)}_{\mathcal{X}}(z_1, z_2) - \alpha_{\mathcal{X}}(z_1) \cdot \alpha_{\mathcal{X}}(z_2)$$

  A função mostra em que regiões do espaço dos objectivos **duas metas** têm tendência a ser ...

  – atingidas conjuntamente, na mesma execução do optimizador.

  $\rightsquigarrow$ **covariância positiva**

  – atingidas em alternativa uma à outra, na mesma execução do optimizador. $\rightsquigarrow$ **covariância negativa**

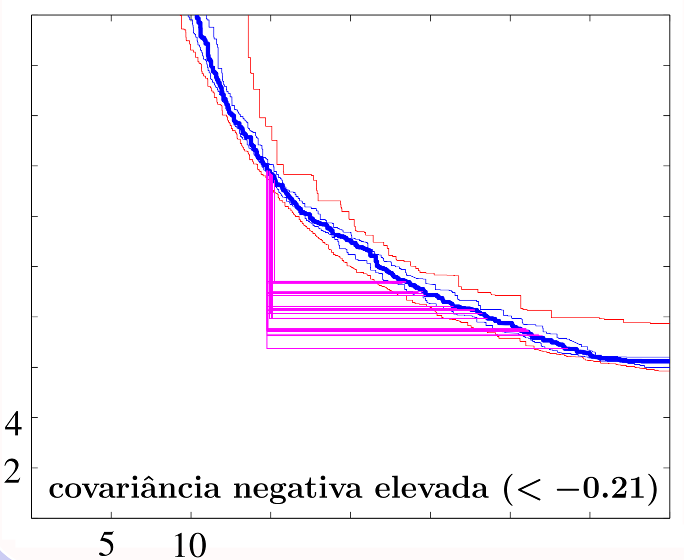
Optimizadores multi-objectivo:
A abordagem da função de aproveitamento

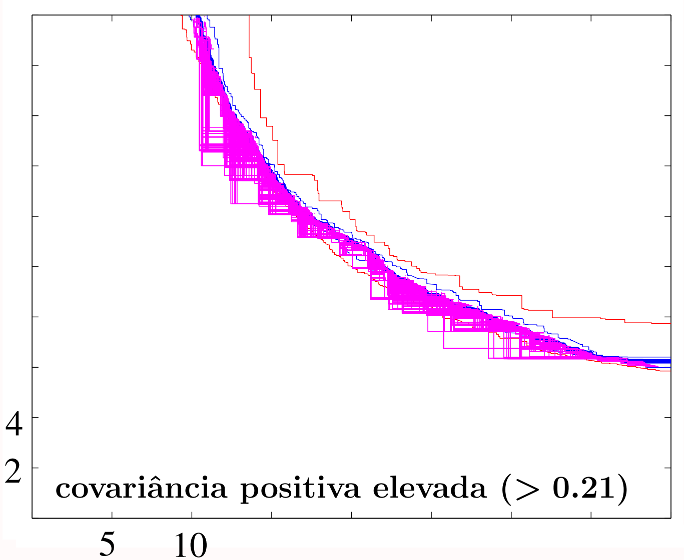
A função $\mathrm{Cov}_{\mathcal{X}}(\cdot,\cdot)$ tem o valor **Zero**, quando ...

– as duas metas podem ser atingidas independentemente.

– uma das duas metas nunca é atingida (ou ambas são sempre).

Ilustração da função de covariância empírica:





Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

→ Casos "não-ideais":

– Covariância **positiva** entre metas **distantes**.

$\Longrightarrow$ Em cada execução, o optimizador aproxima toda a fronteira $\mathcal{X}^*$ de maneira equilibrada, umas vezes melhor que outras.

– Covariância **negativa** entre metas **distantes**.

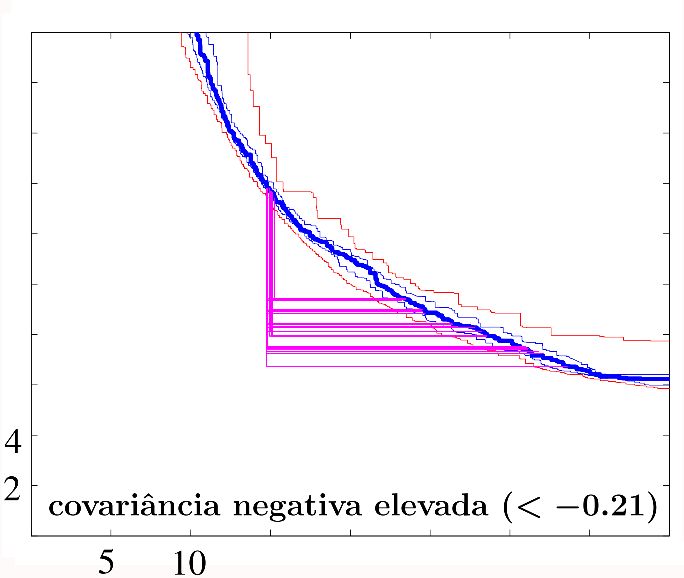
$\Longrightarrow$ Em cada execução, o optimizador aproxima apenas uma região (pequena) da fronteira $\mathcal{X}^*$, diferente em cada execução.

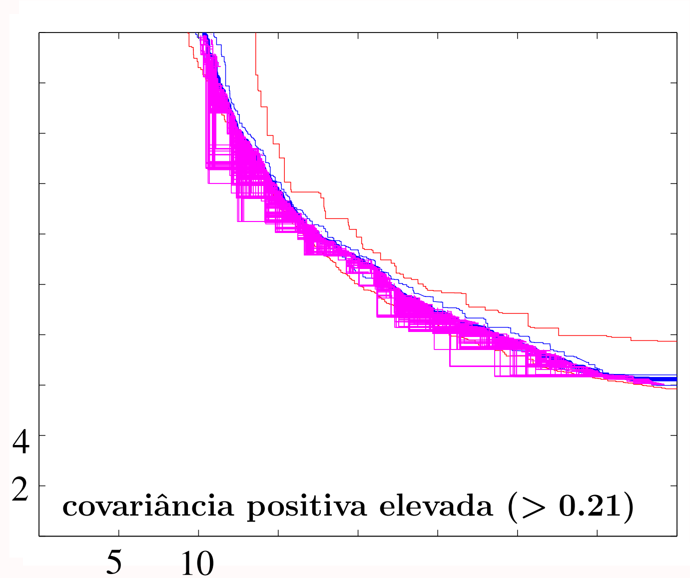
⚠ A covariância é sempre positiva para duas metas próximas e tem o valor da variância para duas metas iguais.

Optimizadores multi-objectivo:

A abordagem da função de aproveitamento

Então . . .





– Existem valores da função de covariância negativos elevados entre algumas metas distantes.

– Não existem valores da função de covariância positivos elevados entre metas distantes.

A abordagem dos indicadores de qualidade

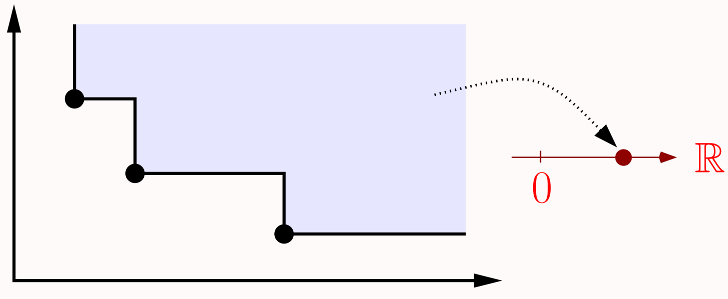
Optimizadores multi-objectivo: A abordagem dos indicadores de qualidade

Para evitar a complexidade da distribuição de um conjunto APN $\mathcal{X}$ de resultados no espaço dos objectivos $\mathbb{R}^d$ ...

– transforma-se $\mathcal{X}$ numa variável aleatória em $\mathbb{R}$, com respeito a um conjunto de referência Z_{ref}, i.e.

$$\mathcal{X} \mapsto I_{Z_{ref}}(\mathcal{X}) \in \mathbb{R}$$

– considera-se a respectiva **distribuição univariada**, geralmente, em termos da sua média.



Há muitos indicadores de qualidade unários na literatura.

Exemplos tratados em seguida:

1. Indicador ϵ unário
2. Indicador da fração atingida *("covered fraction", "coverage")*
3. Indicador (ponderado) de hipervolume

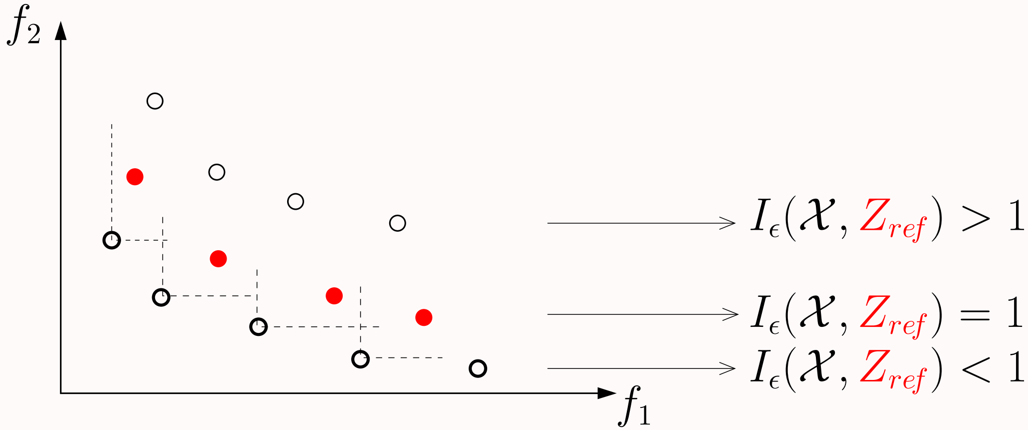
Indicador ϵ unário:

→ **Definição**: Seja $Z_{ref} = \{z_1^{ref}, z_2^{ref}, \ldots, z_k^{ref}\}$ com pontos não-dominados em $\mathbb{R}^d$.

$$I_\epsilon(\mathcal{X}, Z_{ref}) = \inf \left\{ \epsilon \in \mathbb{R}^+ \ : \prod_{i=1}^{k} \boldsymbol{I}\{\mathcal{X} \trianglelefteq \epsilon \cdot z_i^{ref}\} = 1 \right\}$$

→ O indicador tem realizações em $(0, \infty)$, onde um valor menor indica melhor qualidade do conjunto de resultados realizado.

Optimizadores multi-objectivo: Indicadores de qualidade



Ideia:

– Quanto é que Z_{ref} se tem que deslocar para cima para a realização de $\mathcal{X}$ atingir todos os elementos z_i^{ref} ?

– Até onde é que Z_{ref} se pode deslocar para baixo para a realização de $\mathcal{X}$ ainda atingir todos os elementos z_i^{ref} ?

Em termos absolutos, Z_{ref} deverá representar um nível de qualidade minimamente satisfatório para um conjunto de resultados. Assim,

– um valor médio do indicador $\boxed{> 1}$ indicará uma qualidade não-satisfatória.

– um valor médio do indicador $\boxed{\leq 1}$ indicará uma qualidade satisfatória.

??? Mas, por que não usar antes a mediana da distribuição deste indicador, cujo valor é

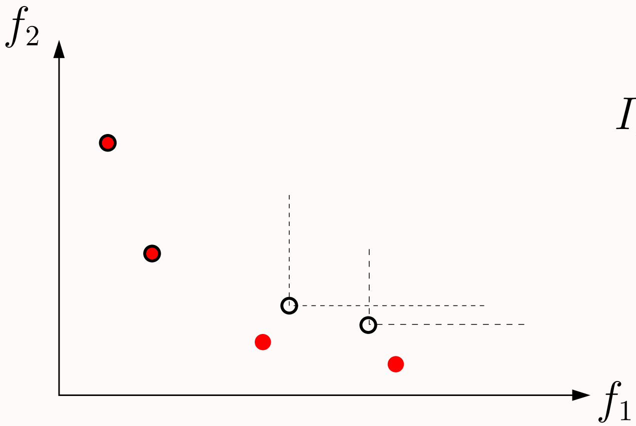
$$\inf \left\{ \epsilon \in \mathbb{R}^+ : \; \alpha_{\mathcal{X}}^{(k)}(\epsilon \cdot z_1^{ref}, \ldots, \epsilon \cdot z_k^{ref}) = 0.5 \right\} \qquad \textbf{?}$$

Optimizadores multi-objectivo: Indicadores de qualidade

Indicador da fração atingida:

→ **Definição original** com $Z_{ref} = \mathcal{X}^*$ um conjunto finito de pontos:

Proporção da fronteira óptima de Pareto $\mathcal{X}^*$ atingida por $\mathcal{X}$

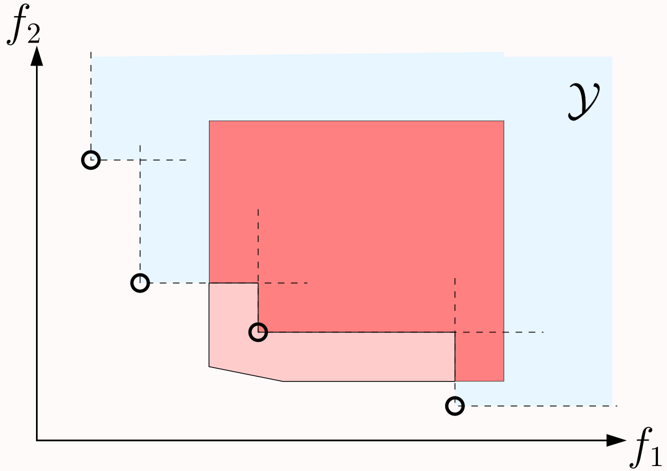


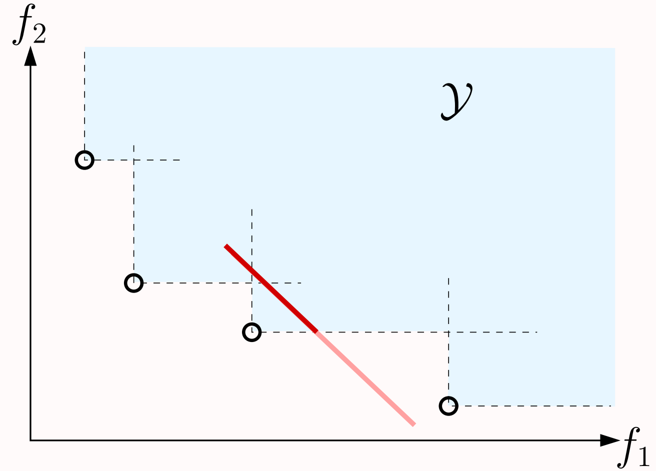
$$I_{FA}(\mathcal{X}, Z_{ref}) = 2/4 = 0.5$$

→ O indicador tem realizações em $[0, 1]$, onde um valor maior indica melhor qualidade do conjunto de resultados realizado.

Optimizadores multi-objectivo: Indicadores de qualidade

$\rightarrow$ **Definição generalizada** com conjuntos $Z_{ref} \in \mathbb{R}^d$ compactos:





O indicador mede ...

– a proporção do conjunto Z_{ref} atingida por $\mathcal{X}$, ou

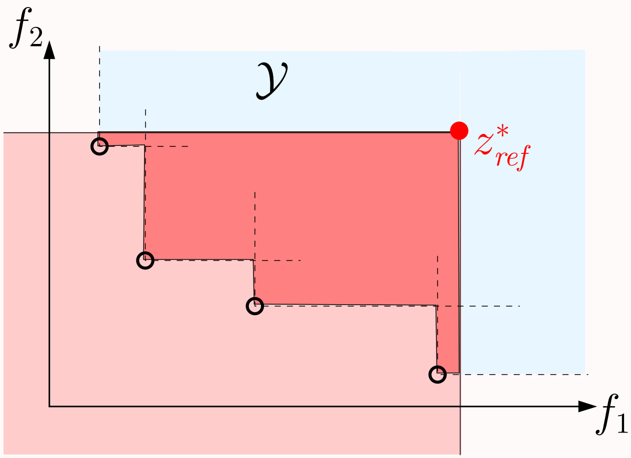
– o $\boxed{\text{tamanho}}$ da interseção do *conjunto atingido*

$$\mathcal{Y} = \{z \in \mathbb{R}^d : \ \mathcal{X} \trianglelefteq z\}$$

com o conjunto Z_{ref}, multiplicado com $1/\text{tamanho}(Z_{ref})$.

<u>**Indicador (ponderado) de hipervolume**</u>:

→ **Definição** com conjuntos <u>fechados</u> $Z^*_{ref} = \{z \in \mathbb{R}^d : \; z \leq z^*_{ref}\}$



Mede ...

o $\boxed{\text{tamanho}}$ da interseção de $\mathcal{Y}$ com Z^*_{ref} <u>ponderado</u> por uma função de ponderação $w(\cdot)$ não-negativa, integrável sobre Z^*_{ref}.

Fórmula **comum** para os últimos dois indicadores:

$$\boxed{\int_{Z_{ref}} \mathbf{I}\{\mathcal{X} \trianglelefteq z\} \cdot w(z)\, \mathcal{H}(dz)}$$

medida Hausdorff adequada

onde para o . . .

→ Indicador generalizado de fração atingida:

$$w(z) = \frac{1}{\text{tamanho}(Z_{ref})} = \frac{1}{\mathcal{H}(Z_{ref})}$$

→ Indicador ponderado de hipervolume:

$$Z_{ref} = Z^{*}_{ref}$$

Optimizadores multi-objectivo: Indicadores de qualidade

Fórmula **comum** para a **média** dos últimos dois indicadores:

$$\boxed{\int_{Z_{ref}} \boxed{\boldsymbol{\alpha_{\mathcal{X}}(z)}} \cdot \boldsymbol{w(z)}\ \boldsymbol{\mathcal{H}(dz)}}$$

função de aproveitamento de ordem 1

!

Mostra a ligação entre a **abordagem dos indicadores de qualidade** e a **abordagem da função de aproveitamento**!

Notas finais

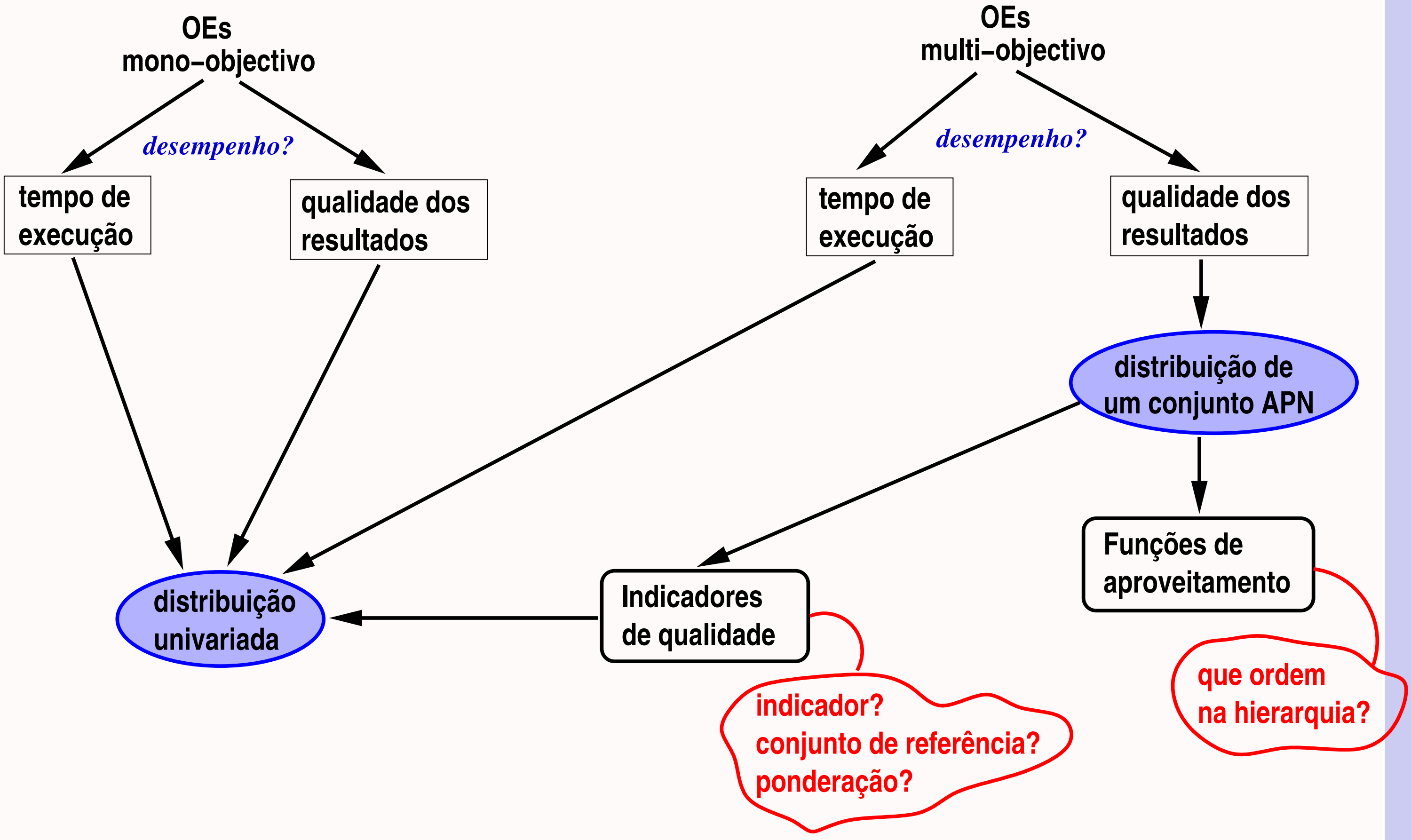
- O desempenho de um OE é associado às

  distribuições de probabilidade

  dos seus resultados e do tempo de execução.

- A complexidade dessas distribuições dificulta a avaliação completa do desempenho de um OE.

- Uma avaliação baseada em informação parcial deve ser planeada com cuidado!


OEs
mono−objectivo
desempenho?
tempo de execução
qualidade dos resultados
OEs
multi−objectivo
desempenho?
tempo de execução
qualidade dos resultados
distribuição de um conjunto APN
Funções de aproveitamento
Indicadores de qualidade
distribuição univariada
indicador?
conjunto de referência?
ponderação?
que ordem
na hierarquia?

Obrigada!