

# Minicurso de Verão da FGV/2025: Fundamentos da Otimização Multiobjetivo: Métodos, Teoria e Aplicações

Jefferson D.G. Melo

IME/UFG

14 de janeiro de 2025

## Descrição do minicurso

- Otimização multiobjetiva
- Método gradiente multiobjetivo
- Método do ponto proximal multiobjetivo
- Método de Newton multiobjetivo
- Método de Frank-Wolfe e método gradiente-proximal multiobjetivos
- Discussão da pesquisa brasileira sobre o tema e problemas em aberto

## Descrição da primeira nota de aula

- O problema de otimização multiobjetiva e solução Pareto
- Condição de otimalidade-Pareto crítico
- Revisão do método gradiente clássico
- Direção de descida multiobjetiva e busca de Armijo
- Método gradiente multiobjetivo
- Propriedades básicas e convergência do método gradiente multiobjetivo
- Análise de convergência do método gradiente multiobjetivo no caso convexo

## O problema de otimização multiobjetivo e solução Pareto ótima

- Estamos interessados no seguinte problema de otimização multiobjetivo:

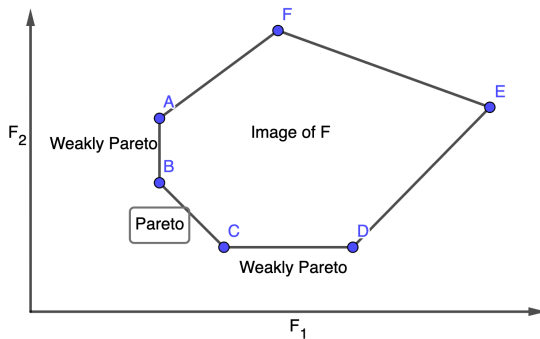
$$\min_{x \in \mathbb{R}^n} F(x), \quad (1)$$

sendo  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  uma função multiobjetiva diferenciável.

- Um ponto  $x^* \in \mathbb{R}^n$  é uma solução Pareto fraca (fracamente Pareto eficiente) para o problema (1) se **não existe**  $x \in \mathbb{R}^n$  tal que  $F(x) \prec F(x^*)$  (ou, equivalentemente,  $F_i(x) < F_i(x^*)$ , para todo  $i = 1, \dots, m$ ).
- Um ponto  $x^* \in \mathbb{R}^n$  é uma solução Pareto (Pareto eficiente) para o problema (1) se **não existe**  $x \in \mathbb{R}^n$  tal que  $F(x) \preceq F(x^*)$  e  $F(x) \neq F(x^*)$  (ou seja, se  $F_i(x) < F_i(x^*)$ , para algum  $i \in \{1, \dots, m\}$ , então  $F_j(x) > F_j(x^*)$ ) para algum  $j \in \{1, \dots, m\}$ .
- Note que utilizamos a ordem parcial usual em  $\mathbb{R}^m$ : se  $a, b \in \mathbb{R}^m$ , temos

$$a \preceq b \iff a_i \leq b_i \quad \text{e} \quad a \prec b \iff a_i < b_i, \quad \forall i \in \{1, \dots, m\}. \quad (2)$$

## Ilustração de soluções Pareto e Pareto fraco



## Exemplo 1: Planejamento econômico sustentável

O governo deseja desenvolver um plano de produção industrial que equilibre dois objetivos conflitantes:

- **Maximizar o crescimento econômico**, representado pela produção total de bens industriais.
- **Minimizar o impacto ambiental**, representado pelas emissões de carbono geradas pela produção.

Defina:

- $x_1$ : quantidade de bens industriais do tipo 1 produzidos.
- $x_2$ : quantidade de bens industriais do tipo 2 produzidos.

As funções objetivo são:

$$f_1(x_1, x_2) = ax_1 + bx_2 \quad (\text{crescimento econômico em unidades monetárias}),$$

$$f_2(x_1, x_2) = cx_1 + dx_2 \quad (\text{emissões de carbono, em toneladas, a minimizar}).$$

## Continuação

Note que para este problema, pode existir algumas restrições:

- recursos disponíveis para produção são limitados;
- demanda mínima deve ser atendida;
- não pode haver produção negativa.

Interpretação:

O governo deseja **maximizar**  $f_1$  enquanto **minimiza**  $f_2$ . No entanto, como esses objetivos são conflitantes (mais produção gera mais emissões), não há uma solução única que otimize ambos. Em vez disso, busca-se o **conjunto de soluções de Pareto**, onde melhorar um objetivo implica piorar o outro.

## Problema tri-objetivo: Planejamento econômico sustentável com bem-estar social

O Prefeito de uma cidade está planejando o desenvolvimento econômico de suas regiões urbanas e busca equilibrar três objetivos conflitantes:

- **Maximizar a produção econômica:** Representa o aumento do PIB regional baseado na produção de bens e serviços.
- **Minimizar os custos ambientais:** Relaciona-se com a redução das emissões de carbono e o uso de recursos naturais.
- **Maximizar o bem-estar social:** Avalia o impacto sobre a qualidade de vida, incluindo a geração de empregos e a saúde da população.



## Um exemplo ilustrativo

A seguir, apresentamos um exemplo de uma função quadrática para ilustrar a impossibilidade de encontrar um otimizador para ambas as componentes em um problema multiobjetivo:

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} = \begin{bmatrix} (x_1 - 1)^2 + x_2^2 \\ x_1^2 + (x_2 - 1)^2 \end{bmatrix},$$

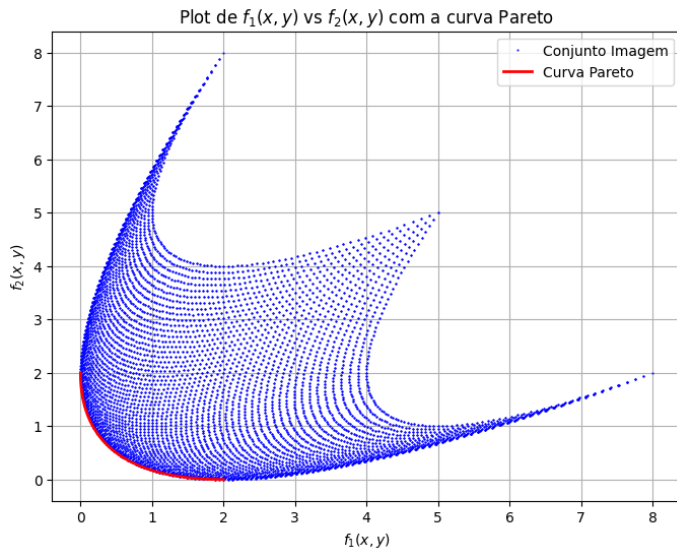
onde  $x = (x_1, x_2) \in \mathbb{R}^2$ .

Note que:

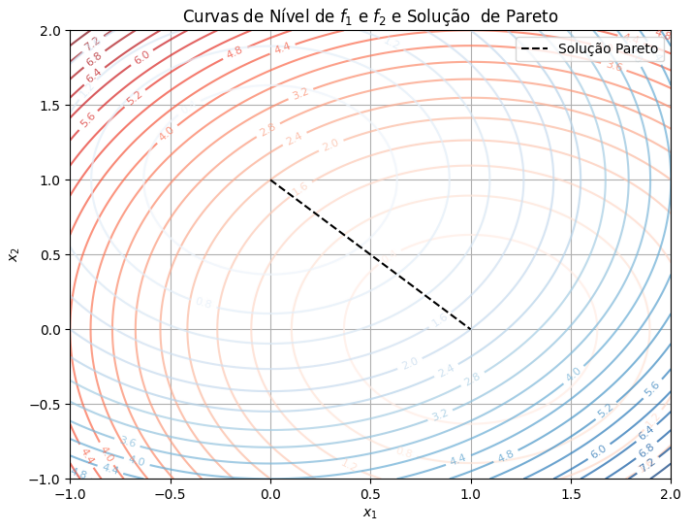
- O minimizador de  $f_1(x)$  é  $x^* = (1, 0)$ , com  $f_1(x^*) = 0$ .
- O minimizador de  $f_2(x)$  é  $x^{**} = (0, 1)$ , com  $f_2(x^{**}) = 0$ .

Como não existe  $x \in \mathbb{R}^2$  que minimize simultaneamente  $f_1(x)$  e  $f_2(x)$ , evidencia-se que os objetivos entram em conflito, necessitando de um conceito mais amplo de solução, como as soluções de Pareto ou Pareto fracas.

## Ilustração da curva Pareto do exemplo anterior



## Curvas de nível de $f_1$ e $f_2$ com as soluções Pareto.



## Problema de otimização escalar e condição de otimalidade

Considere o seguinte problema de otimização escalar:

$$\min_{x \in \mathbb{R}^n} f(x),$$

onde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  é uma função diferenciável.

A condição de otimalidade de primeira ordem para um ponto  $x^* \in \mathbb{R}^n$  ser uma solução ótima local é:

$$\nabla f(x^*) = 0,$$

onde  $\nabla f(x^*)$  é o gradiente de  $f$  avaliado em  $x^*$ .

Esta condição implica que o vetor gradiente, cuja direção oposta aponta na direção de maior decrescimento de  $f$ , deve ser nulo em  $x^*$ , indicando um ponto crítico.

## Condição de otimalidade-otimização multiobjetivo

Agora, considere o seguinte conjunto

$$\text{Im } J_F(x) := \{J_F(x)v \mid v \in \mathbb{R}^n\} = \{(\langle \nabla F_1(x), v \rangle, \dots, \langle \nabla F_m(x), v \rangle) \mid v \in \mathbb{R}^n\}.$$

A condição clássica de otimalidade do caso escalar, “gradiente igual a zero”, pode ser estendida para o contexto multiobjetivo utilizando o conjunto acima da seguinte forma: um ponto  $x^* \in \mathbb{R}^n$  é um Pareto crítico para o problema (1) se

$$\text{Im } J_F(x^*) \cap (-\mathbb{R}_{++})^m = \emptyset. \quad (3)$$

Equivalentemente, para todo  $v \in \mathbb{R}^n$  existe  $j \in \{1, \dots, m\}$  tal que

$$\langle \nabla F_j(x^*), v \rangle \geq 0.$$

Logo, se  $x^*$  não for um Pareto crítico, então existe uma direção  $v \in \mathbb{R}^n$  tal que

$$\langle \nabla F_i(x^*), v \rangle < 0, \quad \forall i = 1, \dots, m.$$

## Relação entre criticalidade e solução Pareto (fraca)

- 1) Se um ponto  $x^*$  é uma solução Pareto fraca, então  $x^*$  é um Pareto crítico. Em geral, o recíproco é falso.
- 2) Se  $F$  é uma função convexa, o recíproco é válido, i.e., todo ponto Pareto crítico é solução Pareto fraca.
- 3) Todo minimizador (ou ponto crítico) de uma função escalarizada  $f(x) = \sum_{j=1}^m \lambda_j F_j(x)$  é um Pareto fraco (Pareto crítico) de  $F$ .
- 4) Se  $F$  é uma função convexa, vale a volta de 3), isto é, toda solução Pareto fraca de  $F$  é minimizador de alguma função escalarizada.

## Justificativas dos fatos anteriores

**Justificativa de (1):** Se  $x^*$  não é Pareto crítico, existe  $v \in \mathbb{R}^n$  tal que  $J_F(x^*)v \prec 0$ . Logo, pela expansão de Taylor de primeira ordem, temos que

$$F(x^* + tv) - F(x^*) = tJ_F(x^*)v + o(t) = t \left( J_F(x^*)v + \frac{o(t)}{t} \right)$$

Como  $\lim_{t \rightarrow 0} o(t)/t = 0$ , temos que para  $t > 0$  suficientemente pequeno, vale que  $F(x^* + tv) - F(x^*) \prec 0$ , ou seja  $F(x^* + tv) \prec F(x^*)$ , implicando que  $x^*$  não é Pareto fraco.

**Justificativa de (2):** Como  $F$  é convexa, temos que para todo  $y \in \mathbb{R}^n$ :

$F_j(y) - F_j(x^*) \geq \langle \nabla F_j(x^*), y - x^* \rangle$ , e portanto

$$\max_{j=1,\dots,m} \{F_j(y) - F_j(x^*)\} \geq \max_{j=1,\dots,m} \{\langle \nabla F_j(x^*), y - x^* \rangle\} \geq 0,$$

onde a última desigualdade usamos o fato que  $x^*$  é Pareto crítico. Concluimos então que  $x^*$  é um ponto Pareto fraco.

## Justificativa do fato (3)

Assuma que  $x^*$  seja um minimizador da função escalarizada  $f(x) = \sum_{j=1}^m \lambda_j F_j(x)$ , para algum  $\lambda \in \mathbb{R}_+^m$  tal que  $\sum_{j=1}^m \lambda_j = 1$ . Se  $x^*$  não for Pareto fraco, então existe  $y \in \mathbb{R}^n$  tal que  $F_j(y) < F_j(x^*)$  para todo  $j = 1, \dots, m$ . Portanto, temos que

$$f(y) = \sum_{j=1}^m \lambda_j F_j(y) < \sum_{j=1}^m \lambda_j F_j(x^*) = f(x^*),$$

contrariando o fato que  $x^*$  é minimizador de  $f$ .

Similarmente, assuma que  $x^*$  seja um ponto crítico da função escalarizada  $f(x) = \sum_{j=1}^m \lambda_j F_j(x)$ , para algum  $\lambda \in \mathbb{R}_+^m$  tal que  $\sum_{j=1}^m \lambda_j = 1$ . Logo,  $\sum_{j=1}^m \lambda_j \nabla F_j(x^*) = 0$ . Se  $x^*$  não for Pareto crítico de  $F$ , existiria  $v \in \mathbb{R}^n$  tal que  $\langle \nabla F_j(x^*), v \rangle < 0$  para todo  $j = 1, \dots, m$ . Portanto, obteríamos um absurdo, pois

$$0 = \left\langle \sum_{j=1}^m \lambda_j \nabla F_j(x^*), v \right\rangle = \sum_{j=1}^m \lambda_j \langle \nabla F_j(x^*), v \rangle < 0,$$

onde para a última desigualdade usamos que  $\lambda_j \geq 0$  e  $\lambda \neq 0$ .

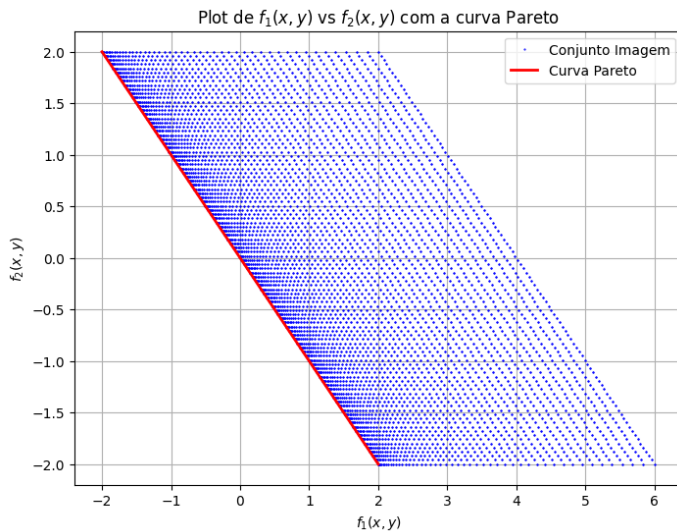


## Exemplo ilustrando problema com escalarização “a priori”

Obs.: O fato (4) pode ser provado utilizando caracterização do subdiferencial de uma função convexa dado pelo máximo de funções afins.

Agora, ilustraremos o problema em considerar escalarização a priori. Para isto, considere a função convexa bi-objetiva  $F(x, y) = (x^2 - y, y)$ . Note que a Imagem de  $F$  é dada por  $Im F = F(\mathbb{R}^2) = \{(u, v) : u = x^2 - y, v = y, \text{ com } x, y \in \mathbb{R}\}$ . Logo,  $u + v = x^2 \geq 0$  e todo  $u, v$  tal que  $u + v \geq 0$  está na  $Im F$ , basta tomar  $y = v$  e  $x = \sqrt{u + v}$ , de onde teremos  $x^2 = u + v$ , i.e.,  $u = x^2 - v = x^2 - y$ . Portanto  $Im F = \{(u, v) : u + v \geq 0\}$ . As soluções Pareto (fracos) correspondem aos pontos em que  $u + v = 0$ , i.e.,  $(x^2 - y) + y = x^2 = 0$ . Logo os pontos Pareto formam o conjunto  $\{0\} \times \mathbb{R}$ . Mais ainda, embora qualquer mínimo da função escalarizada  $f(z) = aF_1(z) + bF_2(z)$ , com  $(a, b) \in \mathbb{R}_+^2$  e  $a + b = 1$ , seja um Pareto fraco de  $F$ , esse mínimo pode não existir. De fato,  $f(x, y) = ax^2 + (b - a)y$ . Logo, essa função possui minimizador se e somente se  $a = b = 1/2$ , os quais correspondem ao conjunto de solução Pareto de  $F$ , i.e.,  $\{0\} \times \mathbb{R}$ .

## Ilustração da curva Pareto do exemplo anterior



## Direção de descida para o método do gradiente

Note que um ponto  $x^*$  é Pareto crítico se

$$\max_{i=1,\dots,m} \langle \nabla F_i(x^*), v \rangle \geq 0, \quad \forall v \in \mathbb{R}^n. \quad (4)$$

Portanto, se  $x$  não for um Pareto crítico, então existe uma direção de descida  $v \in \mathbb{R}^n$  para todas as funções componentes  $F_j$ . A idéia principal dos métodos de descida é verificar de alguma forma se a condição (4) é satisfeita; caso contrário, calcular uma direção  $v$  e um tamanho de passo  $\alpha$  tal que

$$F_j(x + \alpha v) < F_j(x), \quad \forall j \in \{1, \dots, m\}. \quad (5)$$

## Método do gradiente clássico

O **método do gradiente** busca minimizar uma função escalar  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  iterativamente, movendo-se na direção oposta ao gradiente da função, i.e., gerando uma sequência de pontos  $\{x^k\}$  a partir de um ponto inicial  $x^0$ , por meio da seguinte formulação básica:

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k)$$

- $x^k \in \mathbb{R}^n$  é o ponto da  $k$ -ésima iteração,
- $\nabla f(\cdot)$  é o gradiente da função,
- $\alpha_k$  é o tamanho do passo.

Note que a direção  $v^k := -\nabla f(x^k)$  é solução do problema:

$$\arg \min_{v \in \mathbb{R}^n} \left\{ \langle \nabla f(x^k), v \rangle + \frac{1}{2} |v|^2 \right\} \quad (6)$$

## Escolha do tamanho do passo

A escolha do passo  $\alpha_k$  é crucial para a convergência. Algumas abordagens incluem:

- **Tamanho fixo:**  $\alpha_k = \alpha$ , constante.
- **Regra de Armijo:** Ajuste adaptativo de  $\alpha_k := 2^{-\ell_k}$  com base na redução da função:

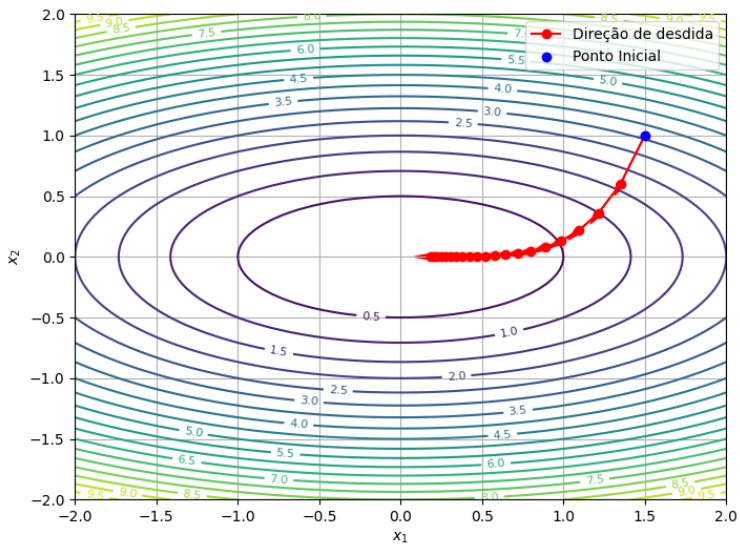
$$f(x^k - 2^{-\ell} \nabla f(x^k)) \leq f(x^k) - \delta 2^{-\ell} \|\nabla f(x^k)\|^2,$$

onde  $\delta \in (0, 1)$  e  $\ell_k$  é o menor número natural  $\ell$  satisfazendo a desigualdade acima.

- **Busca Linear Exata:** Cálculo do tamanho do passo minimizando a função ao longo da direção de descida.

A escolha adequada de  $\alpha_k$  melhora a eficiência e a estabilidade do método.

## Trajetória do método do gradiente



## Idéia do método do gradiente multiobjetivo

O método do gradiente multiobjetivo, proposto por Fliege e Svaiter <sup>1</sup>, gera uma sequência  $\{x^k\}$  por meio do procedimento:

$$x^{k+1} = x^k + \alpha_k v^k,$$

onde  $v^k$  é a única solução do seguinte subproblema fortemente convexo (não suave):

$$v^k := \arg \min_{v \in \mathbb{R}^n} \left\{ \max_{j=1, \dots, m} \langle \nabla F_j(x^k), v \rangle + \frac{1}{2} |v|^2 \right\}, \quad (7)$$

e o tamanho do passo  $\alpha_k$  é escolhido por uma regra do tipo Armijo.

Note que a direção de descida multiobjetiva  $v^k$  generaliza a direção de máxima descida, no contexto escalar, dada pela direção oposta ao gradiente :  $v^k = -\nabla f(x^k)$ .

---

<sup>1</sup>Fliege, J., Svaiter, B.F.: [Steepest descent methods for multicriteria optimization](#).  
Math. Methods Oper. Res. 51(3):479-494, 2000.

## Subproblema da direção de descida multiobjetiva

Note que o problema (7) para obter a direção multiobjetiva generaliza (6) e pode ser reformulado como o seguinte problema de otimização quadrática com restrições lineares:

$$\min_{(v,t) \in \mathbb{R}^n \times \mathbb{R}} \left\{ t + \frac{1}{2}|v|^2 : \langle \nabla F_j(x^k), v \rangle \leq t, \quad \forall j = 1, \dots, m \right\}. \quad (8)$$

Note que a função Lagrangiana para este subproblema é:

$$L_{x^k}(t, v, \lambda) = t + \frac{1}{2}|v|^2 + \sum_{j=1}^m \lambda_j (\langle \nabla F_j(x^k), v \rangle - t),$$

e portanto as condições de KKT são:

$$1 - \sum_{j=1}^m \lambda_j = 0, \quad v + \sum_{j=1}^m \lambda_j \nabla F_j(x^k) = 0, \quad (9)$$

$$\lambda_j \geq 0, \quad \langle \nabla F_j(x^k), v \rangle \leq t, \quad \lambda_j (\langle \nabla F_j(x^k), v \rangle - t) = 0, \quad \forall j = 1, \dots, m \quad (10)$$



Observe que o KKT anterior tem solução. Por exemplo, o problema é convexo e satisfaz Slater ((1, 0) é ponto de Slater.)

A função  $\theta(x)$  definida por

$$\theta(x) := \min_{v \in \mathbb{R}^n} \left\{ \max_{j=1, \dots, m} \langle \nabla F_j(x), v \rangle + \frac{1}{2} |v|^2 \right\}. \quad (11)$$

possui um papel importante, tanto na caracterização da direção  $v^k := v(x^k)$  quanto na análise de convergência do método gradiente multiobjetivo.

Temos as seguintes propriedades:

- $\theta(x) \leq 0$  e  $\theta(x) = 0$  se e somente se  $v(x) = 0$ , este caso acontece se e somente se  $x$  é um ponto Pareto crítico de  $F$ .
- $\theta(x)$  e a solução  $v(x)$  correspondente são funções contínuas.

## Justificativa das propriedades de $\theta(x)$

Da definição de  $\theta(x)$ , temos que

$$\theta(x) \leq \max_{j=1,\dots,m} \langle \nabla F_j(x), 0 \rangle + \frac{1}{2}|0|^2 = 0$$

Note também que

$$\theta(x) = \max_{j=1,\dots,m} \langle \nabla F_j(x), v(x) \rangle + \frac{1}{2}|v(x)|^2.$$

Em particular, segue trivialmente que se  $v(x) = 0$ , então  $\theta(x) = 0$ . Assuma agora que  $\theta(x) = 0$ . Então,

$$0 = \theta(x) \leq \max_{j=1,\dots,m} \langle \nabla F_j(x), sv(x) \rangle + \frac{1}{2}|sv(x)|^2, \quad \forall s \in \mathbb{R}.$$

Dividindo a expressão acima por  $s > 0$  e passando o limite com  $s \rightarrow 0$ , obtemos que  $\max_{j=1,\dots,m} \langle \nabla F_j(x), v(x) \rangle \geq 0$ . Voltando na definição de  $\theta(x)$ , concluimos que  $0 = \theta(x) \geq (1/2)\|v(x)\|^2$  e portanto que  $v(x) = 0$ .

## Caracterização de Pareto fraco usando $v(x)$

Do fato que  $\theta(x) \leq 0$ , temos que

$$\max_{j=1,\dots,m} \langle \nabla F_j(x), v(x) \rangle \leq -\frac{1}{2}|v(x)|^2.$$

Logo, se  $v(x) \neq 0$ , obtemos que  $\langle \nabla F_j(x), v(x) \rangle < 0$  para todo  $j = 1, \dots, m$ , e portanto  $v(x)$  é uma direção de descida, implicando que  $x$  não é Pareto crítico.

Obs. A desigualdade acima será bastante utilizada.

Agora se  $v(x) = 0$ , temos que  $\theta(x) = 0$  e portanto para todo  $v \in \mathbb{R}^n$  e  $s \in \mathbb{R}$ , temos

$$0 = \theta(x) \leq \max_{j=1,\dots,m} \langle \nabla F_j(x), sv \rangle + \frac{1}{2}|sv|^2, \quad \forall s \in \mathbb{R}.$$

Dividindo a expressão acima por  $s > 0$  e passando o limite com  $s \rightarrow 0$ , obtemos que  $\max_{j=1,\dots,m} \langle \nabla F_j(x), v \rangle \geq 0$ , concluindo que  $x$  é Pareto crítico.

**Exercício** Continuidade das funções  $\theta(x)$  e  $v(x)$  (Dica: utilize o sistema KKT, subsequências adequadas para os multiplicadores  $\lambda^k$  e a descrição de  $v(x^k)$ ).

## Busca de Armijo-multiobjetiva

Seja  $v$  uma direção de descida para  $F$  em  $x$ :

$$\max_{i=1,\dots,m} \langle \nabla F_i(x), v \rangle < 0, \quad \forall v \in \mathbb{R}^n. \quad (12)$$

A busca de Armijo consiste em obter um tamanho de passo  $\alpha_\ell = 2^{-\ell}$  tal que  $\ell$  é o menor número natural satisfazendo:

$$F(x + \alpha_\ell v) \preceq F(x) + \alpha_\ell \delta J_F(x)v \quad (13)$$

para algum  $\delta \in (0, 1)$ .

A desigualdade vetorial acima é equivalente a

$$F_j(x + \alpha_\ell v) \leq F_j(x) + \alpha_\ell \delta \langle \nabla F_j(x), v \rangle, \quad \text{Para todo } j = 1, \dots, m. \quad (14)$$

Note que (12) implica que  $\langle \nabla F_j(x), v \rangle < 0$ , e portanto a busca de Armijo garante que o tamanho de passo  $\alpha := \alpha_\ell$  seja tal que ponto  $x^+ := x + \alpha v$  é de descida para  $F$  em  $x$ , isto é,  $F(x^+) \prec F(x)$ .

## Boa definição da busca de Armijo

Seja  $v$  uma direção de descida para  $F$  em  $x$ :

$$q(v) := \max_{i=1,\dots,m} \langle \nabla F_i(x), v \rangle < 0. \quad (15)$$

Pela fórmula de Taylor, temos

$$F_j(x + \alpha v) = F_j(x) + \alpha \left( \langle \nabla F_j(x), v \rangle + \frac{o_j(\alpha)}{\alpha} \right) \quad (16)$$

Como  $\delta \in (0, 1)$ ,  $q(v) < 0$  e  $\lim_{\alpha \rightarrow 0} \frac{o_j(\alpha)}{\alpha} = 0$ ,  $j = 1, \dots, m$ , podemos tomar  $\bar{\alpha} > 0$  suficientemente pequeno tal que, para todo  $\alpha \in (0, \bar{\alpha})$ , temos

$$\frac{\|o_j(\alpha)\|}{\alpha} \leq (\delta - 1)q(\alpha) \leq (\delta - 1)\langle \nabla F_j(x), v \rangle, \quad \forall j = 1, \dots, m.$$

Logo, utilizando essa desigualdade em (16), temos  $F_j(x + \alpha v) \leq F_j(x) + \alpha\delta\langle \nabla F_j(x), v \rangle$ , para todo  $\alpha \in (0, \bar{\alpha})$ .

## Método do gradiente multiobjetivo

- 0) (Passo Inicial) Escolha  $x^0 \in \mathbb{R}^n$ ,  $\delta \in (0, 1)$  e faça  $k = 0$ ;
- 1) (Direção de Descida) Compute  $v^k$  como sendo a única solução de

$$\min_{v \in \mathbb{R}^n} \left\{ \max_{j=1, \dots, m} \langle \nabla F_j(x^k), v \rangle + \frac{1}{2} |v|^2 \right\}, \quad (17)$$

- 2) (Critério de Parada) Se  $v^k = 0$ , pare;
- 3) (Escolha do Passo) Determine o primeiro inteiro não negativo  $\ell_k := \ell$  satisfazendo

$$F(x^k + 2^{-\ell} v^k) \preceq F(x^k) + 2^{-\ell} \delta J_F(x^k) v^k; \quad (18)$$

- 4) (Atualização), Faça  $\alpha_k = 2^{-\ell_k}$ ,  $x^{k+1} := x^k + \alpha_k v^k$ ,  $k \leftarrow k + 1$  e retorne ao Passo 1.

## Propriedades básicas

- Se o Método gradiente multiobjetivo parar na  $k$ -ésima iteração, então  $x^k$  é um ponto Pareto crítico;
- Existe  $\lambda^k \in \mathbb{R}_+^m$  tal que  $\sum_{j=1}^m \lambda_j^k = 1$  e  $v^k = -\sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)$ ;
- Se  $x^k$  não é Pareto crítico, então  $v^k$  é uma direção de descida:  $F(x^{k+1}) \prec F(x^k)$ ;
- Se o conjunto de nível  $L_F(x^0) := \{x \in \mathbb{R}^n : F(x) \preceq F(x^0)\}$  é compacto, então  $\{x^k\}$  é limitada;
- Todo ponto de acumulação de  $\{x^k\}$ , caso exista, é um ponto crítico.

*Vamos provar este resultado de convergência logo em seguida.*

## Convergência do método gradiente multiobjetivo

Apresentaremos alguns passos da prova da convergência do método gradiente multiobjetivo:

- Como  $\{F(x^k)\}$  é decrescente (na ordem parcial de  $\mathbb{R}^m$ ), ela converge;
- Assuma, sem perda de generalidade, que  $\{x^k\}$  converge para  $\bar{x}$ ;
- Definindo  $q(v) = \max_{j=1, \dots, m} \langle \nabla F_j(x^k), v \rangle$ , temos de (17) que  $\|v^k\|^2 \leq -2q(v^k)$ ;
- Da busca de Armijo, definição de  $q(v)$  e do fato que  $F(x^k) \succeq F(\bar{x})$ , temos:

$$0 \leq \sum_{k=0}^N -\delta \alpha_k q(v^k) \leq F_j(x^0) - F_j(x^{N+1}) \leq F_j(x^0) - F_j(\bar{x}), \quad \forall j = 1, \dots, m.$$

- Da última desigualdade acima, temos  $\{\alpha_k q(v^k)\} \rightarrow 0$ . Logo, se  $\inf_k \alpha_k > 0$ , então  $\{q(v^k)\}$  converge a zero, o qual implica que  $\{v^k\}$  também converge a zero;
- Como  $v^k = -\sum_{j=1}^m \lambda_j^k \nabla F_j(x^k)$ , com  $\lambda^k \in \mathbb{R}_+^m$  e  $\sum_{j=1}^m \lambda_j^k = 1$ , temos que existe  $\bar{\lambda} \in \mathbb{R}_+^m$  tal que  $\sum_{j=1}^m \bar{\lambda}_j = 1$  e  $\sum_{j=1}^m \bar{\lambda}_j \nabla F_j(\bar{x}) = 0$ . De onde concluímos que  $\bar{x}$  é crítico.



## Continuação da análise de convergência

Assumiremos agora que  $\inf_k \alpha_k = 0$ . Refinando a uma subsequência se necessário, podemos assumir que  $v^k \rightarrow \bar{v}$ ,  $x^k \rightarrow \bar{x}$ ,  $\alpha_k \rightarrow 0$  e que  $\alpha_k < 1$  para todo  $k \geq k_0$ .

Em particular, temos que  $\hat{\alpha}_k := 2\alpha_k$  também converge a zero e da regra de Armijo temos que existe um índice  $j_k$  tal que

$$F_{j_k}(x^k + \hat{\alpha}_k v^k) - F_{j_k}(x^k) > \hat{\alpha}_k \delta \langle \nabla F_{j_k}(x^k), v^k \rangle.$$

Como  $j_k \in \{1, \dots, m\}$ , temos que existe  $\bar{j} \in \{1, \dots, m\}$  tal que  $j_k = \bar{j}$  infinitas vezes (Refinando a uma subsequência, se necessário, assumiremos que  $j_k = \bar{j}$  para todo  $k$ ).

Logo dividindo ambos os lados da desigualdade acima por  $\hat{\alpha}_k$  e utilizando o Teorema do valor médio, temos que

$$\langle \nabla F_{\bar{j}}(x^k + \tilde{\alpha}_k v^k), v^k \rangle > \delta \langle \nabla F_{\bar{j}}(x^k), v^k \rangle,$$

com  $\tilde{\alpha}_k \in (0, \hat{\alpha}_k)$ .

## Continuação da análise de convergência

Portanto, passando o limite na desigualdade acima, obtemos

$$\langle \nabla F_{\bar{j}}(\bar{x}), \bar{v} \rangle \geq \delta \langle \nabla F_{\bar{j}}(\bar{x}), \bar{v} \rangle,$$

equivalentemente,

$$(1 - \delta) \langle \nabla F_{\bar{j}}(\bar{x}), \bar{v} \rangle \geq 0$$

e portanto, da definição de  $q(v)$  e do fato que  $\delta \in (0, 1)$ , concluímos que

$$q(\bar{v}) \geq \langle \nabla F_{\bar{j}}(\bar{x}), \bar{v} \rangle \geq 0.$$

Mas sabemos que  $\|v^k\|^2 \leq -2q(v^k)$ , o qual passando o limite com  $k \rightarrow +\infty$ , implica que  $\|\bar{v}\|^2 \leq -2q(\bar{v}) \leq 0$ . Concluindo que  $\sum_{j=1}^m \bar{\lambda}_j \nabla F_j(\bar{x}) = \bar{v} = 0$  e portanto que  $\bar{x}$  é Pareto crítico.

## Convergência total sob convexidade

Em seguida mostraremos que, sob convexidade e uma hipótese adicional, toda a sequência  $\{x^k\}$  gerada pelo método gradiente multiobjetivo converge para um ponto Pareto fraco.

**Hipótese 1:** A função  $F$  é convexa, i.e., as componentes  $F_j$  são funções convexas.

Para mostrar a convergência total de  $\{x^k\}$ , utilizaremos da propriedade de Fejér convergência, a qual discutimos abaixo.

## Definição e resultado básico de Fejér convergência

**Definition:** Seja  $S$  um subconjunto não vazio de  $\mathbb{R}^n$ . Uma sequência  $\{x^k\}$  em  $\mathbb{R}^n$  é dita quasi-Fejér convergente para  $S$  se, e somente se, para todo  $x \in S$ , existe uma sequência somável  $\{\varepsilon_k\} \subset \mathbb{R}_+$  tal que, para todo inteiro  $k \geq 0$ , temos

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \varepsilon_k.$$

A sequência  $\{x^k\}$  é Fejér convergente para  $S$  se, na definição acima,  $\varepsilon_k = 0$  para todo  $k \geq 0$ . A importância da convergência quasi-Fejér é ilustrada no seguinte lema.

**Lema:** Se  $\{x^k\}$  é quasi-Fejér convergente para um conjunto não vazio  $S$ , então as seguintes afirmações são válidas:

- (i) a sequência  $\{x^k\}$  é limitada;
- (ii) se um ponto de acumulação de  $\{x^k\}$  pertence a  $S$ , então  $\{x^k\}$  converge para um ponto em  $S$ .

## Fejér convergência de $\{x^k\}$ gerada pelo método gradiente multiobjetivo

Seja  $\Omega_k := \{x \in \mathbb{R}^n : F(x) \preceq F(x^k)\}$ . Note que como  $\{F(x^k)\}$  é decrescente, temos que  $\Omega_{k+1} \subset \Omega_k$ . Precisaremos da seguinte hipótese:

**Hipótese 2:** O conjunto  $\Omega := \bigcap_k \Omega_k$  é não vazio.

**Lemma:** A seguinte desigualdade vale:

$$\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \alpha_k \max_{j=1,\dots,m} \{F_j(x) - F_j(x^k)\} + \alpha_k^2 \|v^k\|^2.$$

Como consequência:

**Teorema:** Se valem Hipótese 1 e Hipótese 2, então a sequência  $\{x^k\}$  é quasi-Fejér convergente a um ponto Pareto fraco pertencente ao conjunto  $\Omega$ .

## Prova da Fejér convergência de $\{x^k\}$ .

Lembre que  $x^{k+1} = x^k + \alpha_k v^k$ . Logo, temos

$$\begin{aligned}\|x^{k+1} - x\|^2 &= \|x^k - x\|^2 + 2\alpha_k \langle v^k, x^k - x \rangle + \alpha_k^2 \|v^k\|^2 \\ &= \|x^k - x\|^2 + \alpha_k \sum_{j=1}^m \lambda_j^k \langle \nabla F_j(x^k), x - x^k \rangle + \alpha_k^2 \|v^k\|^2 \\ &\leq \|x^k - x\|^2 + \alpha_k \sum_{j=1}^m \lambda_j^k \left( F_j(x) - F_j(x^k) \right) + \alpha_k \|v^k\|^2\end{aligned}\tag{19}$$

onde usamos a convexidade de  $F$  e que  $\alpha_k^2 \leq \alpha_k \leq 1$ . Note que a busca de Armijo e a definição de  $q(v)$  garantem que

$$F_j(x^{k+1}) \leq F_j(x^k) + \delta \alpha_k \langle \nabla F_j(x^k), v^k \rangle \leq F_j(x^k) + \delta \alpha_k q(v^k),$$

o qual implica que

$$-\alpha_k q(v^k) \leq \frac{F_j(x^k) - F_j(x^{k+1})}{\delta}$$

## Continuação da análise de convergência de $\{x^k\}$ .

Lembre  $\|v^k\|^2 \leq -2q(v^k)$ , logo concluímos que

$$\alpha_k \|v^k\|^2 \leq -2\alpha_k q(v^k) \leq \frac{2[F_j(x^k) - F_j(x^{k+1})]}{\delta},$$

implicando que

$$\sum_{k=0}^N \alpha_k \|v^k\|^2 \leq \frac{F_j(x^0) - F_j(x^{N+1})}{\delta}.$$

Pela Hipótese 2, temos para todo  $\bar{x} \in \Omega$  e  $j = 1, \dots, m$ , que  $F_j(\bar{x}) \leq F_j(x^{N+1})$ . Portanto,

$$\sum_{k=0}^N \alpha_k \|v^k\|^2 \leq \frac{F_j(x^0) - F_j(\bar{x})}{\delta}.$$

Usando  $x = \bar{x}$  em (19), temos que  $\{x^k\}$  é quasi-Fejér convergente para  $\Omega$ . Como  $\{F(x^k)\}$  é decrescente, temos que todo ponto de acumulação de  $\{x^k\}$  está em  $\Omega$ . Logo,  $\{x^k\}$  converge a um ponto de  $\Omega$ . A prova do teorema segue, já que todo ponto de acumulação é Pareto crítico, o qual é Pareto fraco, pois  $F$  é convexa.