Curso de Verão da FGV/2025
Fundamentos da Otimização Multiobjetivo
Método de Newton Multiobjetivo

Jefferson D.G. Melo

IME/UFG

16 de janeiro de 2025

Descrição dessa nota de aula

- Hipóteses sobre o problema de otimização multiobjetivo
- Revisão do método de Newton para problema de otimização escalar
- Método de Newton multiobjetivo
- Análise de convergência do método de Newton multiobjetivo

O Problema de otimização multiobjetivo e solução Pareto ótima

- Nesta parte estamos interessados no seguinte problema de otimização multiobjetivo:

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

  sendo $F : \mathbb{R}^n \to \mathbb{R}^m$ uma função multiobjetiva de classe $\mathcal{C}^2$. Assumiremos também que $\nabla^2 F_j(x)$, $j = 1, \ldots, m$, sejam definidas positivas para x na vizinhança de uma solução Pareto. Note que sob essa hipótese, as soluções Pareto e Pareto fracas coincidem.

- Um ponto $x^* \in \mathbb{R}^n$ é uma solução Pareto fraca de (1) se **não existe** $x \in \mathbb{R}^n$ tal que $F(x) \prec F(x^*)$.

- Um ponto $x^* \in \mathbb{R}^n$ é uma solução Pareto de (1) se **não existe** $x \in \mathbb{R}^n$ tal que $F(x) \preceq F(x^*)$ e $F(x) \neq F(x^*)$.

- Utilizamos a ordem parcial usual em $\mathbb{R}^m$: se $a, b \in \mathbb{R}^m$, temos

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

Ilustração de uma solução Pareto (fraca)

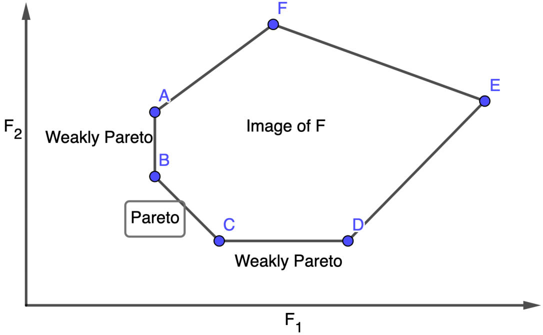
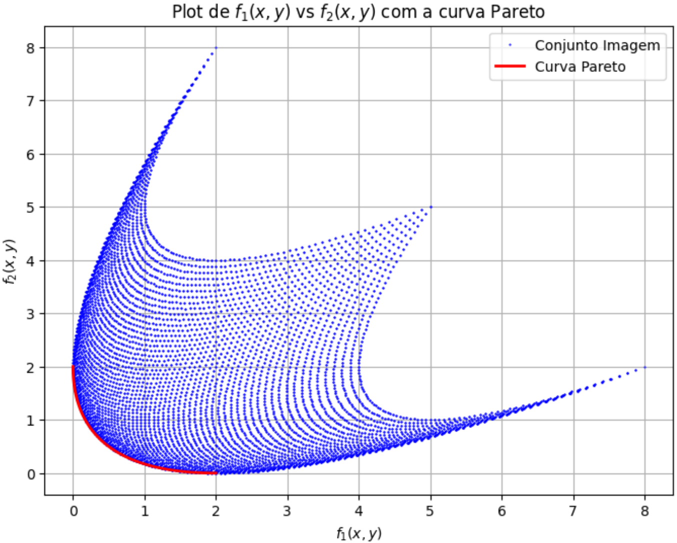


Ilustração de curva Pareto



Método de Newton para encontrar zeros de uma função

Seja $g : \mathbb{R}^n \to \mathbb{R}^n$ uma função diferenciável, e considere o problema de obter um zero de g, isto é, um ponto x^* tal que

$$g(x^*) = 0.$$

O método de Newton para encontrar um zero de g é um procedimento iterativo que a partir de uma aproximação inicial x_0, gera uma sequência $\{x^k\}$ cujas iterações são definidas por:

$$x^{k+1} = x^k - [g'(x^k)]^{-1} g(x^k),$$

onde $g'(x^k)$ é a matriz Jacobiana de g avaliada em x^k.

O método de Newton pode ser interpretado como a busca pelo zero da aproximação linear de g no ponto atual x^k. A aproximação linear de g em x^k é dada por:

$$L_g(x) := g(x^k) + g'(x^k)(x - x^k).$$

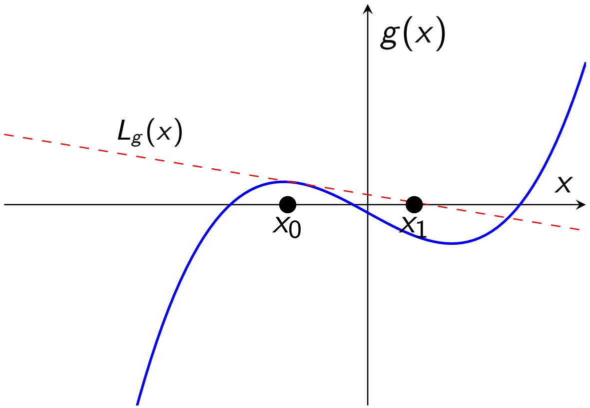
Definimos x^{k+1} como o zero dessa aproximação. Para isso, resolvemos a equação $L_g(x) = 0$:

$$g(x^k) + g'(x^k)(x - x^k) = 0,$$

o que resulta na iteração:

$$x^{k+1} = x^k - [g'(x^k)]^{-1} g(x^k).$$

Método de Newton - Ilustração geométrica de uma iteração



Como $L_g(x) = g(x_0) + g'(x_0)(x - x_0)$, temos que $L_g(x_1) = 0 \longleftrightarrow x_1 = x_0 - \frac{g(x_0)}{g'(x_0)}$.

Principal resultado de convergência do método de Newton

Hipóteses:

- Seja x^* um zero da função $g : \mathbb{R}^n \to \mathbb{R}^n$, i.e., $g(x^*) = 0$;
- Assuma que a derivada de g seja Lipschitz contínua na vizinhança de x^*, i.e., existe uma constante $L > 0$ tal que:

$$|g'(x) - g'(y)| \leq L|x - y|, \quad \forall\, x, y \text{ em uma vizinhança de } x^*.$$

- Assuma que a aplicação $g'(x)$ seja inversível em x^*;

Conclusão:

Se x^0 for suficientemente próximo de x^*, então a sequência $\{x^k\}$ converge quadraticamente para x^*; isto é, existe $c > 0$ tal que

$$\|x^{k+1} - x^*\| \leq c\|x^k - x^*\|^2, \qquad \forall k \geq 0.$$

Zeros de uma função e pontos críticos de um problema escalar

Considere o seguinte problema de otimização escalar:

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

onde $f : \mathbb{R}^n \to \mathbb{R}$ é uma função de classe $\mathcal{C}^2$.

Uma condição necessária para que um ponto $x^* \in \mathbb{R}^n$ seja uma solução ótima do problema acima é:

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

Logo para resolver este problema, podemos considerar o método de Newton para achar zero da função $g(x) = \nabla f(x)$.

Método de Newton para minimização de uma função

Considere o problema de minimizar uma função $f : \mathbb{R}^n \to \mathbb{R}$, onde f é de classe $\mathcal{C}^2$. O método de Newton para minimizar f corresponde ao método de Newton para obter zero da função $g(x) = \nabla f(x)$ e portanto utiliza-se informação de segunda ordem, isto é, da Hessiana $\nabla^2 f(x)$. As iterações do método de Newton são definidas por:

$$x^{k+1} = x^k - [\nabla^2 f(x^k)]^{-1} \nabla f(x^k),$$

onde $\nabla f(x^k)$ é o gradiente de f e $\nabla^2 f(x^k)$ é a matriz Hessiana de f, ambos avaliadas em x^k. Essa iteração corresponde a primeiro obter a aproximação quadrática de f em torno de x^k:

$$q_f(x) = f(x^k) + \nabla f(x^k)(x - x^k) + \frac{1}{2}\langle \nabla^2 f(x^k)(x - x^k), x - x^k \rangle,$$

em seguida, minimizar q_f: $(\nabla q_f(x^{k+1}) = 0)$, i.e., $\nabla f(x^k) + \nabla^2 f(x^k)(x - x^k) = 0$.

Propriedades de convergência do método de Newton

1. Convergência Quadrática:
Seja f de classe $\mathcal{C}^2$ com Hessiana Lipschitz contínua. Seja x^* um ponto crítico de f, $\nabla f(x^*) = 0$, tal que $\nabla^2 f(x^*)$ seja definida positiva. Se x_0 for suficientemente próximo de x^*, então a sequência $\{x^k\}$ gerada pelo método de Newton converge quadraticamente para x^*.

2. Robustez da Convergência:
A rápida convergência depende da qualidade da aproximação inicial. Para x_0 distante do ponto de mínimo, o método pode não convergir.

3. Custos Computacionais:
Cada iteração exige a resolução de um sistema linear envolvendo a Hessiana da função, tornando o método computacionalmente caro em alta dimensão. Variações do método, como os métodos de Quase-Newton, abordam essa questão.

Direção de Newton multiobjetiva

Dado $x \in \mathbb{R}^n$, a direção de Newton multiobjetiva $s(x)$ é definida como sendo a única solução do subproblema

$$s(x) = \arg\min_{y \in \mathbb{R}^n} \max_{j=1,\ldots,m} \left\{ \langle \nabla F_j(x), y \rangle + \frac{1}{2} \langle \nabla^2 F_j(x) y, y \rangle \right\}. \tag{2}$$

Consideraremos o passo puro, para o qual $x^{k+1} = x^k + s^k$, onde $s^k = s(x^k)$.
Note que o subproblema acima tem uma única solução porque $\nabla^2 F_j(x)$ é assumida ser definida positiva. O valor ótimo de (2) é denotado por $\theta(x)$, a qual é uma função contínua e tem as seguintes propriedades [Exercício]:

1. Para todo $x \in \mathbb{R}^n$, $\theta(x) \leq 0$.
2. As seguintes condições são equivalentes:
   (a) O ponto x não é Pareto crítico.
   (b) $\theta(x) < 0$.
   (c) A direção de Newton $s(x)$ é não nula.

Reformulação do subproblema

Note que o subproblema[1]

$$\theta(x) = \min_{y\in\mathbb{R}^n} \max_{j=1,\ldots,m} \left\{ \langle \nabla F_j(x), y\rangle + \frac{1}{2}\langle \nabla^2 F_j(x) y, y\rangle \right\} \tag{3}$$

é não suave e pode ser reformulado como o seguinte problema quadrático convexo:

$$\min \; g(t,y) := t \tag{4}$$

$$s.a. \; \langle \nabla F_j(x), y\rangle + \frac{1}{2}\langle \nabla^2 F_j(x) y, y\rangle \leq t, \quad \forall j = 1, \ldots, m. \tag{5}$$

[1] J. Fliege, L.M.G. Drummond, and B. F. Svaiter. Newton's Method for Multiobjective Optimization. SIAM J. Optim. Vol. 20, 602–626. 2009

Lagrangiana do subproblema e condições de KKT

Note que a Lagrangiana do subproblema de Newton é definida por

$$L(t, y, \lambda) = t + \sum_{j=1}^{m} \lambda_j \left(\langle \nabla F_j(x), y \rangle + \frac{1}{2} \langle \nabla^2 F_j(x) y, y \rangle - t \right), \tag{6}$$

e portanto as condições de KKT se tornam:

$$\sum_{j=1}^{m} \lambda_j = 1, \qquad \sum_{j=1}^{m} \lambda_j \left(\nabla F_j(x) + \nabla^2 F_j(x) y \right) = 0, \tag{7}$$

$$\lambda_j \geq 0, \quad \langle \nabla F_j(x), y \rangle + \frac{1}{2} \langle \nabla^2 F_j(x) y, y \rangle - t \leq 0, \quad \forall j = 1, \ldots, m, \tag{8}$$

$$\lambda_j \left(\langle \nabla F_j(x), y \rangle + \frac{1}{2} \langle \nabla^2 F_j(x) y, y \rangle - t \right) = 0, \quad \forall j = 1, \ldots, m. \tag{9}$$

Observe que as restrições satisfazem as condiçõe de regularidade de Slater, o ponto $(1, 0)$ pertence ao interior do conjunto restrição.

Relação envolvendo a direção de Newton

Segue da segunda relação em (7) que

$$\sum_{j=1}^{m} \lambda_j \left(\nabla F_j(x) + \nabla^2 F_j(x) s(x)\right) = 0$$

Portanto, considerando

$$H_\lambda := \sum_{j=1}^{m} \lambda_j \nabla^2 F_j(x), \quad v_\lambda := \sum_{j=1}^{m} \lambda_j \nabla F_j(x).$$

temos que a direção de Newton multiobjetiva satisfaz $H_\lambda s(x) = -v_\lambda$. isto é, $s(x) = -H_\lambda^{-1} v_\lambda$. Veja que se $j = 1$ (caso escalar), então $s(x) = -(\nabla^2 F)^{-1} \nabla F(x)$, coincidindo com a direção de Newton clássica.

Relação importante: zero gap de dualidade

A existência dos multiplicadores de Lagrange no sistema KKT implica que não há brecha de dualidade entre os problemas primal e dual, e, portanto,

$$\theta(x) = \sup_{\lambda \succeq 0} \inf_{s \in \mathbb{R}^n} L((t,s),\lambda) = \sup_{\substack{\lambda \succeq 0 \\ \sum_j \lambda_j = 1}} \inf_{s \in \mathbb{R}^n} L((t,s),\lambda) \tag{10}$$

$$= \sup_{\substack{\lambda \succeq 0 \\ \sum_j \lambda_j = 1}} \inf_{s \in \mathbb{R}^n} \sum_{j=1}^{m} \lambda_j \left(\nabla F_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_j(x) s \right) \tag{11}$$

$$\geq \inf_{s \in \mathbb{R}^n} \sum_{j=1}^{m} \lambda_j \left(\nabla F_j(x)^\top s + \frac{1}{2} s^\top \nabla^2 F_j(x) s \right), \tag{12}$$

onde a última desigualdade vale para todo $\lambda \succeq 0 : \sum_j \lambda_j = 1$.

Estimativa entre $\theta(x)$ e a direção de Newton

Agora defina:

$$H_\lambda := \sum_{j=1}^{m} \lambda_j \nabla^2 F_j(x), \quad v_\lambda := \sum_{j=1}^{m} \lambda_j \nabla F_j(x).$$

Logo se $\nabla^2 F_j(x) \succeq aI$, então $H_\lambda \succeq aI$ e a desigualdade anterior implica

$$\theta(x) \geq \inf_{s \in \mathbb{R}^n} \{\langle v_\lambda, s\rangle + \frac{a}{2}\|s\|^2\} = -\frac{1}{2a}\|v_\lambda\|^2$$

para qualquer $\lambda \succeq 0$ e $\sum_{j=1}^{m} \lambda_j = 1$. Ou seja,

$$|\theta(x)| \leq \frac{1}{2a}\|v_\lambda\|^2, \qquad \forall \lambda \succeq 0, \sum_{j=1}^{m} \lambda_j = 1. \tag{13}$$

Controle do passo de Newton

Considere $x \in \mathbb{R}^n$ e $0 < a \leq b$ tais que $aI \preceq \nabla^2 F_j(x) \preceq bI$, para todo $j = 1, \ldots, m$. Então

$$\theta(x) = -\frac{1}{2}\langle H_{\lambda(x)} s(x), s(x)\rangle, \qquad \frac{a}{2}\|s(x)\|^2 \leq |\theta(x)| \leq \frac{b}{2}\|s(x)\|^2.$$

Prova. Das condições de KKT para o subproblema de Newton e da hipótese acima, temos que

$$H_\lambda := \sum_{j=1}^{m} \lambda_j \nabla^2 F_j(x), \quad v_\lambda := \sum_{j=1}^{m} \lambda_j \nabla F_j(x) \tag{14}$$

satisfaz $aI \preceq H_\lambda \preceq bI$ e $v_{\lambda(x)} = -H_{\lambda(x)} s(x)$. Além disso, temos que

$$\theta(x) = \langle v_{\lambda(x)}, s(x)\rangle + \frac{1}{2}\langle H_{\lambda(x)} s(x), s(x)\rangle = -\frac{1}{2}\langle H_{\lambda(x)} s(x), s(x)\rangle.$$

Logo, a conclusão segue do fato que $\theta(x) \leq 0$.

Comparação entre $s(x)$ e v_λ:

Mais ainda, sob a hipótese anterior temos que v_λ satisfaz

$$\|s(x)\| \leq \frac{1}{a}\|v_{\lambda(x)}\|.$$

De fato, como da condição de KKT temos que $H_{\lambda(x)}s(x) = -v_\lambda$, utilizando a desigualdade de Cauchy-Schwarz e a hipótese considerada, temos que

$$a\|s(x)\|^2 \leq \langle H_{\lambda(x)}s(x), s(x)\rangle = -\langle v_{\lambda(x)}, s(x)\rangle \leq \|v_{\lambda(x)}\|\|s(x)\|.$$

De onde temos que $\|s(x)\| \leq \frac{1}{a}\|v_{\lambda(x)}\|$.

Para todo k vale a seguinte propriedade $\|s^{k+1}\| \leq \frac{L}{2a}\|s^k\|^2$

Sabemos que se $\nabla^2 F_j$ for L-Lipschitz contínua para $j = 1, \ldots, m$, então

$$\|\nabla F_j(y) - \nabla F_j(x) - \nabla^2 F_j(x)(y-x)\| \leq \frac{L}{2}\|y-x\|^2. \tag{15}$$

Tomando $x := x^k$ e $y =: x^{k+1} = x^k + s^k$, temos

$$\|\nabla F_j(x^{k+1}) - \nabla F_j(x^k) - \nabla^2 F_j(x^k)s^k\| \leq \frac{L}{2}\|s^k\|^2.$$

Definindo $\hat{v}^{k+1} = \sum_{j=1}^m \lambda_j^k \nabla F_j(x^{k+1})$, temos do KKT e da des. de Cauchy-Schwarz que

$$\|\hat{v}^{k+1}\| = \|\sum_{j=1}^m \lambda_j^k \left(\nabla F_j(x^{k+1}) - \nabla F_j(x^k) - \nabla^2 F_j(x^k)s^k\right)\| \tag{16}$$

$$\leq \sum_{j=1}^m \lambda_j^k \|\nabla F_j(x^{k+1}) - \nabla F_j(x^k) - \nabla^2 F_j(x^k)s^k\| \leq \frac{L}{2}\sum_{j=1}^m \lambda_j^k \|s^k\|^2 = \frac{L}{2}\|s^k\|^2. \tag{17}$$

Portanto o resultado segue por usar que $(a/2)\|s^{k+1}\|^2 \leq |\theta(x^{k+1})| \leq (1/2a)\|\hat{v}^{k+1}\|^2$.

Convergência de s^k

Assuma que $\|s^0\| \leq \frac{a}{L}$. Então $\|s^{k-1}\| \leq a/(2^{k-1}L)$ para todo $k \geq 1$. Claramente essa desigualdade vale para $k = 1$. Assuma que vale para algum $k \geq 1$. Temos pela estimativa do slide anterior que

$$\|s^k\| \leq \frac{L}{2a}\|s^{k-1}\|^2 \leq \frac{L}{2a}\|s^{k-1}\|\frac{a}{2^{k-1}L} = \frac{1}{2^k}\|s^{k-1}\| \leq \frac{1}{2^k}\frac{a}{L}.$$

Em particular, $\{s^k\}$ converge a zero e temos que $\sum_k \|x^{k+1} - x^k\| = \sum_k \|s^k\| < +\infty$. Este resultado implica que a sequência $\{x^k\}$ converge para algum $\bar{x}$. Além disso, como $\{s^k\}$ converge a zero e $|\theta(x^k)| \leq b\|s^k\|$, concluimos que $\theta(\bar{x}) = \lim_k \theta(x^k) = 0$ e portanto, $\bar{x}$ é um Pareto fraco de F. Note que se x^0 estiver suficientemente próximo de um Pareto fraco, então $\theta(x^0) \approx 0$, e como consequência a desigualdade $(a/2)\|s^0\| \leq |\theta(x^0)|$ implica que $\|s^0\|$ vai ser suficientemente pequeno a ponto de cumprir a condição exigida, isto é, $\|s^0\| \leq a/L$.

Convergência Quadrática

Para todo $i > k > k_0$ temos

$$\begin{aligned}
\|x^i - x^{k+1}\| &\leq \sum_{j=k+1}^{i-1} \|x^{j+1} - x^j\| \leq \sum_{j=k+1}^{i-1} \|s^j\| \leq \sum_{j=k+1}^{i-1} \frac{L}{a}\|s^{j-1}\|^2 \\
&= \frac{L}{a}\left\{\|s^k\|^2 + \|s^{k+1}\|^2 + \|s^{k+2}\|^2 + \|s^{k+3}\|^2 \ldots\right\} \\
&\leq \frac{L}{a}\left\{\|s^k\|^2 + \frac{L^2}{a^2}\|s^k\|^4 + \frac{L^2}{a^2}\|s^{k+1}\|^4 + \frac{L^2}{a^2}\|s^{k+2}\|^4 \ldots\right\} \\
&\leq \frac{L}{a}\left\{\|s^k\|^2 + \frac{L^2}{a^2}\|s^k\|^4 + \frac{L^6}{a^6}\|s^k\|^8 + \frac{L^{14}}{a^{14}}\|s^k\|^{16} \ldots\right\} \\
&= \frac{L}{a}\|s^k\|^2\left\{1 + \frac{L^2}{a^2}\|s^k\|^2 + \frac{L^6}{a^6}\|s^k\|^6 + \frac{L^{14}}{a^{14}}\|s^k\|^{14} \ldots\right\} \\
&\leq \frac{L}{a}\|s^k\|^2 \frac{1}{1 - [L^2/a^2]\|s^k\|^2}
\end{aligned} \tag{18}$$

As desigualdades acima implicam $\|x^i - x^{k+1}\| \leq C\|s^k\|^2$ para todo $i > k$. Note que isso implica que $\{x^i\}$ é limitada e portanto ao longo de uma subsequência converge para um certo $\bar{x}$. Por outro lado,

$$\|x^k - \bar{x}\| \geq \|x^k - x^{k+1}\| - \|x^{k+1} - \bar{x}\| \geq \|s^k\| - C\|s^k\|^2 = \|s^k\|(1 - C\|s^k\|), \tag{19}$$

implicando que

$$\|s^k\| \leq \frac{1}{1 - C\|s^k\|}\|x^k - \bar{x}\|.$$

Combinando essa desigualdade com $\|x^{k+1} - \bar{x}\| \leq C\|s^k\|^2$, temos

$$\|x^{k+1} - \bar{x}\| \leq \frac{C}{(1 - C\|s^k\|)^2}\|x^k - \bar{x}\|^2 \leq \tilde{C}\|x^k - \bar{x}\|^2,$$

implicando a convergência quadrática. Note que para última desigualdade usamos o fato que para k suficientemente grande, temos que $\|s^k\|$ é suficientemente pequeno.

Ilustração Numérica

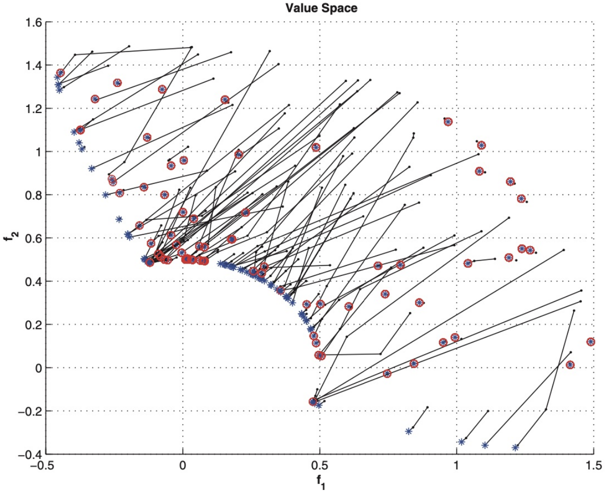


Ilustração Numérica