Notas de estudo

Resumo Da Teoria de Otimização

João Matheus Lira Luiz

Vitória da Conquista, 2024

Conteúdo

1 Resultados Importantes

Lema 1.1. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável em $\mathbb{R}^n$, com derivada Lipschitz-contínua no $\mathbb{R}^n$ com módulo $L > 0$. Então, temos a desigualdade*

$$|f(x+d) - f(x) - \langle f'(x), d\rangle| \leq \frac{L\|d\|^2}{2}, \tag{1}$$

onde $d \in \mathbb{R}^n$.

Demonstração. Considere a função $\psi : [0,1] \to \mathbb{R}$ dada por

$$\psi(t) = f(x+td) - f(x) - t\langle f'(x), d\rangle.$$

Note que, pela regra da cadeia, temos

$$\begin{aligned}\psi'(t) &= \langle f'(x+td), d\rangle - \langle f'(x), d\rangle \\ &= \langle f'(x+td) - f'(x), d\rangle.\end{aligned}$$

Além disso, observe que

$$\begin{aligned}f(x+d) - f(x) - \langle f'(x), d\rangle &= \psi(1) - \psi(0) \\ &= \int_0^1 \psi'(t)\,dt \\ &= \int_0^1 \langle f'(x+td) - f'(x), d\rangle\,dt.\end{aligned}$$

Agora, tomando o valor absoluto de ambos os lados, obtemos

$$\begin{aligned}|f(x+d) - f(x) - \langle f'(x), d\rangle| &= \left|\int_0^1 \langle f'(x+td) - f'(x), d\rangle\,dt\right| \\ &\leq \int_0^1 \|f'(x+td) - f'(x)\| \cdot \|d\|\,dt.\end{aligned}$$

Como a derivada de f é Lipschitz-contínua, sabemos que

$$\|f'(x+td) - f'(x)\| \leq L\|td\| = Lt\|d\|.$$

Logo, temos

$$\begin{aligned}\int_0^1 \|f'(x+td) - f'(x)\| \cdot \|d\|\,dt &\leq \int_0^1 Lt\|d\|^2\,dt \\ &= L\|d\|^2 \int_0^1 t\,dt \\ &= L\|d\|^2 \cdot \frac{1}{2}.\end{aligned}$$

Portanto, concluímos que

$$|f(x+d) - f(x) - \langle f'(x), d\rangle| \leq \frac{L\|d\|^2}{2}.$$

□ □

2 Existência de Soluções

Sejam dados um conjunto $D \subset \mathbb{R}^n$ e uma função $f : \Omega \to \mathbb{R}$, onde $D \subset \Omega$. O objetivo é encontrar um minimizador de f no conjunto D:

$$\min f(x) \text{ sujeito a } x \in D \tag{2}$$

- **Conjunto viável:** o conjunto D.
- **Pontos viáveis:** os pontos de D.
- **Função objetivo:** a função f.

Definição 2.1. *Dizemos que $\bar{x} \in D$ é:*

(a) ***Minimizador global*** *de $f : \Omega \to \mathbb{R}$, se $f(\bar{x}) \leq f(x), \forall x \in D$.*

(b) ***Minimizador local*** *de $f : \Omega \to \mathbb{R}$, se existe uma vizinhança U de $\bar{x}$ tal que $f(\bar{x}) \leq f(x), \forall x \in D \cap U$.*

Se para todo $x \neq \bar{x}$ as desigualdades são estritas, $\bar{x}$ será chamado ***minimizador estrito****.*

Definição 2.2. *Dizemos que $\bar{v} \in \mathbb{R}$, definido por $\bar{v} = \inf\{f(x); x \in D\}$, é o* ***valor ótimo*** *do problema.*

Exemplo 2.1. *Seja $f : \Omega \to \mathbb{R}$ definida por $f(x) = (x+1)^2$, com $D = \Omega = \mathbb{R}$. Temos que:*

- *f é a função objetivo,*
- *$\bar{x} = -1$ é o minimizador global e estrito de f,*
- *$\bar{v} = 0$ é o valor ótimo do problema.*

- Uma função pode admitir vários minimizadores globais, mas o valor ótimo do problema é único.
- Todo minimizador global é um minimizador local, mas a recíproca não é verdadeira.
- Quando $\bar{v} = \infty$ ou $\bar{v} = -\infty$, o problema não possui solução global.

2.1 Condições para Existência de Solução Global

Teorema 2.1 (Teorema de Weierstrass)**.** *Sejam $D \subset \mathbb{R}^n$ um conjunto compacto não vazio, e $f : D \to \mathbb{R}$ uma função contínua. Então o problema de minimizar f tem solução global.*

Demonstração. Como a imagem de um conjunto compacto por uma função contínua é um conjunto compacto, temos que $\{v \in \mathbb{R}; v = f(x),\ \text{para } x \in D\}$ é um conjunto compacto. Em particular, é limitado inferiormente, ou seja, $-\infty < \bar{v} = \inf f(x)$.

Pela definição de ínfimo, $\forall k \in \mathbb{N}, \exists x^k \in D; \bar{v} \leq f(x^k) \leq \bar{v} + \frac{1}{k}$. Passando ao limite quando $k \to \infty$, temos:

$$\lim_{k\to\infty} \bar{v} \leq \lim_{k\to\infty} f(x^k) \leq \lim_{k\to\infty} \left(\bar{v} + \frac{1}{k}\right)$$
$$\bar{v} \leq \lim_{k\to\infty} f(x^k) \leq \bar{v}$$

Como $\{x^k\} \subset D$, que é compacto, segue que $\{x^k\}$ é limitada. Pelo teorema de Bolzano-Weierstrass, possui uma subsequência $\{x^{k_j}\}$ que converge para um ponto $\bar{x} \in D$. Portanto, $\lim_{j\to\infty} x^{k_j} = \bar{x}$ e pela continuidade de f, se conclui que, $\lim_{j\to\infty} f(x^{k_j}) = f(\bar{x})$.

Temos também que, $\lim_{j\to\infty} f(x^{k_j}) = \lim_{k\to\infty} f(x^k)$, pelo fato da $\{f(x^{k_j})\}$ ser uma subsequência $\{f(x^k)\}$ e que $\{f(x^k)\}$ converge par $\bar{v}$, logo, temos que:

$$\lim_{j\to\infty} f(x^{k_j}) = \lim_{k\to\infty} f(x^k) = f(\bar{x}) = \bar{v}$$

Ou seja, f assume o valor mínimo em $\bar{x}$. $\square$

Definição 2.3. *O conjunto de nível da função $f : D \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por:*

$$\mathcal{L}f_D(c) = \{x \in D; f(x) \leq c\}$$

Corolário 2.1. *Sejam $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ contínua no compacto D. Suponhamos que exista $c \in \mathbb{R}$ tal que $\mathcal{L}f_D(c) \neq \varnothing$*

Demonstração. Pelo teorema de Weierstrass, o problema de min $f(x)$ sujeito a $x \in \mathcal{L}f_D(c)$ tem solução global, digamos $\bar{x}$, pelo fato de $\mathcal{L}f_D(c)$ ser um conjunto compacto, ou uma união e compactos. Para todo $x \in D - \{\mathcal{L}f_D(c)\}$, temos $f(x) > c \geq f(\bar{x})$, o que mostra que $\bar{x}$ é um , minimizador global de f e não só de $\mathcal{L}f_D(c)$. $\square$

Definição 2.4. *Dizemos que a função* $f : D \to \mathbb{R}$ *é coerciva no conjunto* D*, quando para cada sequência* $\{x^k\} \subset D$ *tal que ou* $||x^k|| \to \infty$ *ou* $\{x^k\} \to x \in clD \setminus D$ *quando* $k \to \infty$ *tem-se que* $\lim_{k\to\infty} f(x^k) = +\infty$

- Quando o conjunto D é fechado, mas ilimitado, coercividade de f em D significa que $x^k \in D$ e $||x^k|| \to \infty$ implicam $\lim_{k\to\infty} f(x^k) = +\infty$.
- Quando D é limitado, mas não fechado, coercividade de f em D significa que $x^k \in D$ e $\{x^k\} \to x \in clD \setminus D$ tem-se $\lim_{k\to\infty} f(x^k) = +\infty$.
- Quando D é compacto não a sequências em D como a propridedades necessárias.

Exemplo 2.2. *Seja* $f : D \to \mathbb{R}, x = (x_1, x_2) \mapsto f(x) = x_1^2 + x_2 + 1/x_1 + x_2$*, onde* $D = \{x \in \mathbb{R}^2; x_1 + x_2 > 0\}$

- D *é ilimitado.*
- *Ao tomar uma sequência* $\{x^k\} \in D$ *de modo que* $||x^k|| \to \infty$*, temos* $\sqrt{x_{k1}^2 + x_{k2}^2} \to \infty$
- *Ou seja,* $x_{k_1} \to \infty$ *ou* $x_{k_2} \to \infty$
- *Portanto* $f(x^k) = x_{k_1}^2 + x_{k_2} + \frac{1}{x_{k_1} + x_{k_2}} \to \infty$

Corolário 2.2. *Sejam* $D \subset \mathbb{R}^n$ *e* $f : D \to \mathbb{R}$ *uma função contínua coerciva em* $D \neq \varnothing$*. Então o problema de minimizar* f *em* D *possui uma solução global*

Demonstração. Definimos $c = f(x)$, com qualquer $x \in D$. Com isso, o conjunto $\mathcal{L}f_D(c)$ é não vazio. Provaremos que ele é compacto.

Suponha que $\mathcal{L}f_D(c)$ seja ilimitada. Então existe uma sequência $\{x^k\} \subset \mathcal{L}f_D(c)$ tal que $||x^k|| \to \infty$. Por outro lado, $f(x^k) \leq c$ para todo $k \in \mathbb{N}$ pela definição de conjunto de nível, essa contradição implica que C é limitada

Suponhamos que $\mathcal{L}f_D(c)$ não seja fechado, mas limitado. Então existe $\{x^k\} \subset \mathcal{L}f_D(c)$ tal que $\{x^k\} \to x \in cl\mathcal{L}f_D(c) \setminus \mathcal{L}f_D(c)$.

Como $\mathcal{L}f_D(c) = \{x \in D; f(x) \leq c\}$ e $\{x^k\} \subset \mathcal{L}f_D(c)$, converge para $x \in cl\mathcal{L}f_D(c) \setminus \mathcal{L}f_D(c)$, pela coercividade de f em D, temos que $\lim_{k\to\infty} f(x^k) = +\infty$. o que contradiz o fato de que $f(x) \leq c$

Concluimos que $\mathcal{L}f_D(c)$ é não vazio e compacto, logo possui um minimizador global $\square$

3 Condições de Otimalidade

Quando $D = \mathbb{R}^n$, dizemos que o problema de minimização é irrestrito, e quando $D \neq \mathbb{R}^n$, falamos de problemas com restrições.

- Condições de otimalidade para o problema de minimização irrestrito:

$$\text{minimizar} f(x), \text{ sujeito a } x \in \mathbb{R}^n$$

- Para um problema de minimização restrita, temos

$$\text{minimizar} f(x), \text{ sujeito a } x \in D$$

  Onde $D \subset \mathbb{R}^n$, e $\bar{x} \in \text{int}D$

Teorema 3.1. *Condições de otimalidade no caso irrestrito*

(a) Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$ seja diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Suponhamos também que $\bar{x}$ seja um minimizador local do problema. Então temos:

$$f'(\bar{x}) = 0 \tag{1.5}$$

.

(b) Se f é duas vezes diferenciável em $\bar{x}$, então, além de (1.5), tem-se que a matriz hessiana é semidefinida positiva, isto é:

$$\langle f''(\bar{x})d, d\rangle \geq 0 \quad \forall d \in \mathbb{R}^n \tag{1.6}$$

.

(c) Suponhamos que $f : \mathbb{R}^n \to \mathbb{R}$ seja duas vezes diferenciável no ponto $\bar{x} \in \mathbb{R}^n$. Se $\bar{x}$ satisfaz (1.5) e se a matriz hessiana de f em $\bar{x}$ é definida positiva, isto é, se existe $\gamma > 0$ tal que:

$$\langle f''(\bar{x})d, d\rangle > 0, \quad \forall d \in \mathbb{R}^n \setminus \{0\} \tag{1.7}$$

,

então f satisfaz em torno de $\bar{x}$ a condição de crescimento quadrático: existem uma vizinhançau U de $\bar{x}$ e um número $\beta > 0$ tais que

$$f(x) - f(\bar{x}) \geq \beta ||x - \bar{x}||^2, \quad \forall x \in U \tag{1.8}$$

Em particular, $\bar{x}$ é minimizador local estrito do problema. Reciprocamente, se vale (1.8), então $\bar{x}$ satisfaz as condições (1.5) e (1.6)

Demonstração. **(a):** Seja $d \in \mathbb{R}^n$ um vetor arbitrário, porém fixo. Pela definição de minimizador local, existe $\epsilon > 0$ tal que:

$$f(\bar{x}) \leq f(\bar{x} + td), \quad \forall t \in (0, \epsilon).$$

Como f é diferenciável em $\bar{x}$, podemos usar a expansão de Taylor para obter:

$$f(\bar{x} + td) = f(\bar{x}) + t\langle f'(\bar{x}), d\rangle + o(t).$$

Substituindo na desigualdade, obtemos:

$$f(\bar{x}) \leq f(\bar{x}) + t\langle f'(\bar{x}), d\rangle + o(t).$$

Cancelando $f(\bar{x})$ de ambos os lados, temos:

$$0 \leq t\langle f'(\bar{x}), d\rangle + o(t).$$

Dividindo ambos os lados por t e tomando o limite quando $t \to 0^+$, obtemos:

$$0 \leq \langle f'(\bar{x}), d\rangle.$$

Como $d \in \mathbb{R}^n$ é arbitrário, escolhemos $d = -f'(\bar{x})$. Assim, temos:

$$0 \leq \langle f'(\bar{x}), -f'(\bar{x})\rangle = -||f'(\bar{x})||^2.$$

Como $-||f'(\bar{x})||^2 \leq 0$, isso implica que:

$$||f'(\bar{x})||^2 = 0.$$

Logo, concluímos que:

$$f'(\bar{x}) = 0.$$

(b): Seja $d \in \mathbb{R}^n$ arbitrário, porém fixo. Se $\bar{x}$ é minimizador local do problema, então, para todo $t > 0$ suficientemente pequeno, temos:

$$0 \leq f(\bar{x} + td) - f(\bar{x}).$$

Pela expansão de Taylor de segunda ordem, temos:

$$f(\bar{x} + td) = f(\bar{x}) + t\langle f'(\bar{x}), d\rangle + \frac{t^2}{2}\langle f''(\bar{x})d, d\rangle + o(t^2).$$

Subtraindo $f(\bar{x})$ de ambos os lados, obtemos:

$$f(\bar{x} + td) - f(\bar{x}) = t\langle f'(\bar{x}), d\rangle + \frac{t^2}{2}\langle f''(\bar{x})d, d\rangle + o(t^2).$$

Pelo resultado anterior, sabemos que $\langle f'(\bar{x}), d\rangle = 0$. Portanto:

$$0 \leq f(\bar{x} + td) - f(\bar{x}) = \frac{t^2}{2}\langle f''(\bar{x})d, d\rangle + o(t^2).$$

Dividindo ambos os lados por t^2 e tomando o limite quando $t \to 0^+$, obtemos:

$$0 \leq \frac{1}{2}\langle f''(\bar{x})d, d\rangle.$$

Logo, concluímos que:

$$\langle f''(\bar{x})d, d\rangle \geq 0.$$

(c): Suponhamos que as condições (1.5) e (1.7) sejam válidas, mas que (1.8) não se verifique. Então, existe uma sequência $\{x^k\} \subset \mathbb{R}^n \setminus \{\bar{x}\}$ tal que $\{x^k\} \to \bar{x}$ e

$$f(x^k) - f(\bar{x}) < \frac{1}{k}||x^k - \bar{x}||^2, \quad \forall k.$$

Como a sequência $\{(x^k - \bar{x})/||x^k - \bar{x}||\}$ é limitada ($\frac{b}{||b||} \Rightarrow ||\frac{b}{||b||}|| = 1$), pelo Teorema de Bolzano-Weierstrass, possui pontos de acumulação. Escolhendo uma subsequência conveniente, podemos admitir que $\{(x^k - \bar{x})/||x^k - \bar{x}||\} \to d \in \mathbb{R}^n \setminus \{0\}$.

Assim, temos:

$$\begin{aligned}
\frac{1}{k}||x^k - \bar{x}||^2 &> f(x^k) - f(\bar{x}) \\
&= \langle f'(x^k), x^k - \bar{x}\rangle + \frac{1}{2}\langle f''(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(||x^k - \bar{x}||^2) \\
&= \frac{1}{2}\langle f''(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(||x^k - \bar{x}||^2).
\end{aligned}$$

Dividindo ambos os lados por $||x^k - \bar{x}||^2 > 0$ e tomando o limite quando $k \to \infty$, obtemos

$$0 \geq \langle f''(\bar{x})d, d\rangle,$$

o que contradiz a condição (1.7).

Agora, suponha que a condição de crescimento quadrático (1.8) seja válida. Como ela implica que $\bar{x}$ é um minimizador, obtemos também (1.5).

Seja $d \in \mathbb{R}^n$ um vetor arbitrário, mas fixo. Para todo $t \in \mathbb{R}$ suficientemente pequeno, temos $\bar{x} + td \in U$. Utilizando (1.8) e (1.5), obtemos:

$$\begin{aligned}
\beta t^2||d||^2 &= \beta||td||^2 \\
&= \beta||(\bar{x} + td) - \bar{x}||^2 \\
&\leq f(\bar{x} + td) - f(\bar{x}) \\
&= t\langle f'(\bar{x}), d\rangle + \frac{t^2}{2}\langle f''(\bar{x})d, d\rangle + o(t^2) \\
&= \frac{t^2}{2}\langle f''(\bar{x})d, d\rangle + o(t^2).
\end{aligned}$$

Dividindo por $t^2 > 0$ e tomando o limite quando $t \to 0$, obtemos:

$$\beta||d||^2 \leq \frac{1}{2}\langle f''(\bar{x})d, d\rangle.$$

Como $d \in \mathbb{R}^n$ é arbitrário, concluímos que $f''(\bar{x})$ é definida positiva.

□

Observação: Note que (1.5) e (1.6) não são suficientes para que um ponto $\bar{x}$ seja minimizador. Um contra exemplo, é a função dada por $f : \mathbb{R} \to \mathbb{R}, f(x) = x^3$, que satisfaz essas condições no ponto $\bar{x}$, mas não é o minimizador local da função.

4 Convexidade

Definição 4.1. *Um conjunto $D \subset \mathbb{R}^n$ é chamado conjunto convexo se para quaisquer $x, y \in D$ e $\alpha \in [0,1]$, tem-se que $\alpha x + (1-\alpha)y \in D$*

Uma projeção (Ortogonal) do ponto $x \in \mathbb{R}^n$ sobre um conjunto $D \subset \mathbb{R}^n$ é um ponto de D que está mais próximo de x. Em outras palavras, uma solução (global) do problema

$$\text{Min } ||x-y|| \text{ Sujeito a } y \in D \tag{3}$$

Teorema 4.1. *Teorema da Projeção*

(a) Seja $D \subset \mathbb{R}^n$ um conjunto fechado. Então, $\forall x \in \mathbb{R}^n$, uma projeção de x sobre D existe.

(b) Se, além de fechado, D é convexo, então $\forall x \in \mathbb{R}^n$, a projeção de x sobre D, denotada por $P_D(x)$, é única. Além disso, $\bar{x} = P_D(x)$ se, e somente se,

$$\bar{x} \in D, \quad \langle x - \bar{x}, y - \bar{x} \rangle \leq 0, \quad \forall y \in D. \tag{4}$$

(c) Também tem-se que o operador projeção é monótono e não-expansivo (Lipschitz-contínuo com L = 1), ou seja

$$\begin{gathered}\langle x-y, P_D(x) - P_D(y)\rangle \geq ||P_D(x) - P_D(y)||^2 \geq 0 \\ ||P_D(x) - P_D(y)|| \leq ||x-y||, \quad \forall x, y \in \mathbb{R}^n.\end{gathered}$$

Demonstração. **(a):** Fixemos $y \in \mathbb{R}^n$ arbitrário. Seja $f : \mathbb{R}^n \to \mathbb{R}_+, f(x) = ||x-y||$. Por ser norma, f é uma função contínua e

$$\mathcal{L}f_D(c) = D \cap B(y,c)$$

Onde

$$B(y,c) = \{x \in \mathbb{R}^n; ||x-y|| < c\} = \{x \in \mathbb{R}^n; f(x) \leq c\}$$

O conjunto de nível $\mathcal{L}f_D(c)$ é a intersecção do conjunto fechado D com o conjunto compacto $B(y,c)$. Por tanto $\mathcal{L}f_D(c)$ é compacto. Além disso para $c > 0$ suficientemente grande, $B(y,c)$ a contém pontos de D. Logo $\mathcal{L}f_D(c) \neq \varnothing$. A afirmação segue pelo corolário 2

(b): Sejam $\bar{x}$ solução do problema e $y \in D$ qualquer. Como $\bar{x} \in D$ e D é convexo, $(1-\alpha)\bar{x} + \alpha y = x(\alpha) \in D, \forall \alpha \in (0,1]$. Temos então que $||x - \bar{x}|| \leq ||x - x(\alpha)||$, pois $\bar{x}$ é minimizador,

Portanto

$$\begin{aligned}
||x - \bar{x}|| \leq ||x - x(\alpha)|| \Rightarrow ||x - \bar{x}||^2 \leq ||x - x(\alpha)||^2 & \\
0 &\geq ||x - \bar{x}||^2 - ||x - x(\alpha)||^2 \\
&= ||x - \bar{x}||^2 - ||x - \bar{x} + \bar{x} - x(\alpha)||^2 \\
&= ||x - \bar{x}||^2 - ||(x - \bar{x}) - (x(\alpha) - \bar{x})||^2 \\
&= ||x - \bar{x}||^2 - \left(||x - \bar{x}||^2 - 2\langle x - \bar{x}, x(\alpha) - \bar{x}\rangle + ||x(\alpha) - \bar{x}||^2\right) \\
&= 2\langle x - \bar{x}, x(\alpha) - \bar{x}\rangle - ||x(\alpha) - \bar{x}||^2
\end{aligned}$$

Como

$$x(\alpha) = (1-\alpha)\bar{x} + y\alpha = \bar{x} - \alpha\bar{x} + \alpha y \Rightarrow x(\alpha) - \bar{x} = \alpha(y - \bar{x})$$

Temos assim que

$$\begin{aligned}
0 &\geq 2\langle x - \bar{x}, \alpha(y - \bar{x})\rangle - ||\alpha(y - \bar{x})||^2 \\
&= 2\alpha\langle x - \bar{x}, y - \bar{x}\rangle - \alpha^2||y - \bar{x}||^2
\end{aligned}$$

Dividindo Por 2α em ambos os lados, e passando o limite quan $\alpha \to 0$, se conclui

$$0 \geq \langle x - \bar{x}, y - \bar{x}\rangle$$

Sendo $y \in D$ arbitrário.

Suponhamos agora que um certo $\bar{x}$ satisfaça (2). Então para todo $y \in \mathbb{R}$.

$$\begin{aligned}
0 &\geq \langle x-\bar{x}, y-\bar{x}\rangle \\
&= \langle x, y\rangle + \langle x, -\bar{x}\rangle + \langle -\bar{x}, y\rangle + \langle -\bar{x}, -\bar{x}\rangle \\
&= \frac{1}{2}\Big(2\langle x, y\rangle + 2\langle x, -\bar{x}\rangle + 2\langle -\bar{x}, y\rangle + 2\langle -\bar{x}, -\bar{x}\rangle\Big) \\
&= \frac{1}{2}\Big(2\langle x, y\rangle + 2\langle -\bar{x}, x\rangle + 2\langle -\bar{x}, y\rangle + 2\langle -\bar{x}, -\bar{x}\rangle + \langle x, x\rangle - \langle x, x\rangle + \langle y, y\rangle - \langle y, y\rangle\Big) \\
&= \frac{1}{2}\Big((\langle x, x\rangle + 2\langle -\bar{x}, x\rangle + \langle -\bar{x}, -\bar{x}\rangle) + (\langle y, y\rangle + 2\langle y, -\bar{x}\rangle + \langle -\bar{x}, -\bar{x}\rangle) - (\langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle)\Big) \\
&= \frac{1}{2}\Big(||x-\bar{x}||^2 + ||y-\bar{x}||^2 - ||x-y||^2\Big) \\
&\geq \frac{1}{2}\Big(||x-\bar{x}||^2 - ||x-y||^2\Big).
\end{aligned}$$

Concluindo assim, temos

$$\begin{aligned}
0 &\geq ||x-\bar{x}||^2 - ||x-y||^2 \\
||x-y||^2 &\geq ||x-\bar{x}||^2 \\
||x-y|| &\geq ||x-\bar{x}||
\end{aligned}$$

Temos então, $||x-\bar{x}|| \leq ||x-y||, \forall y \in D$, isto é, $\bar{x}$ é uma projeção de x sobre D.

Finalmente, mostrando que a projeção é única. Seja x' alguma outra solução do problema. Usando (3) para $\bar{x}$ com $y = x' \in D$, temos

$$\begin{aligned}
0 &\geq \langle x-\bar{x}, x'-\bar{x}\rangle + \langle x-x', \bar{x}-x'\rangle \\
&= \langle x-\bar{x}, x'-\bar{x}\rangle - \langle x-x', x'-\bar{x}\rangle \\
&= \langle x-\bar{x}-(x-x'), x'-\bar{x}\rangle \\
&= \langle x'-\bar{x}, x'-\bar{x}\rangle \\
&= ||x'-\bar{x}||^2
\end{aligned}$$

Temos assim $0 \geq ||x'-\bar{x}|| \geq 0$, que implica $x' = \bar{x}$. Logo a projeção é única.

(c): Provando agora que o operador projeção é monótono e não-exansivo.

Como $P_D(x)$ e $P_D(y) \in D$, Usando (3) duas vezes (para x e y), obtemos:

$$\langle x - P_D(x), P_D(y) - P_D(x)\rangle \leq 0$$
$$\langle y - P_D(y), P_D(x) - P_D(y)\rangle \leq 0$$

Somando as duas desigualdades temos:

$$\begin{aligned}
0 &\geq \langle x - P_D(x), P_D(y) - P_D(x)\rangle + \langle y - P_D(y), P_D(x) - P_D(y)\rangle \\
&= \langle P_D(x) - x, P_D(x) - P_D(y)\rangle + \langle y - P_D(y), P_D(x) - P_D(y)\rangle \\
&= \langle P_D(x) - x + y - P_D(y), P_D(x) - P_D(y)\rangle \\
&= \langle [P_D(x) - P_D(y)] + y - x, P_D(x) - P_D(y)\rangle \\
&= \langle P_D(x) - P_D(y), P_D(x) - P_D(y)\rangle + \langle y - x, P_D(x) - P_D(y)\rangle \\
&= ||P_D(x) - P_D(y)||^2 - \langle x - y, P_D(x) - P_D(y)\rangle
\end{aligned}$$

Portanto

$$\langle x - y, P_D(x) - P_D(y)\rangle \geq ||P_D(x) - P_D(y)||^2 \geq 0$$

O que prova a primeira afirmação

Provando que o operador projeção é não expansivo. Basta notar que por Cauchy-Schwarz temos:

$$\begin{aligned}
||P_D(x) - P_D(y)|| \cdot ||x - y|| &\geq |\langle P_D(x) - P_D(y), y - x\rangle \\
&= | - \langle x - y, P_D(x) - P_D(y)\rangle| \\
&= \langle x - y, P_D(x) - P_D(y)\rangle \geq ||P_D(x) - P_D(y)||^2
\end{aligned}$$

Se $P_D(x) = P_D(y)$, vale igualdade, do contrário, temos

$$\langle x - y, P_D(x) - P_D(y)\rangle \geq ||P_D(x) - P_D(y)||^2$$

e dividindo por $||P_D(x) - P_D(y)||^2$ em ambos os lados temos

$$||y - x|| = ||x - y|| \geq ||P_D(x) - P_D(y)||$$

□

Definição 4.2. *Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Diz-se que a função $f : D \to \mathbb{R}^n$ é* *convexa em D quando para quaisquer $x, y \in D$ e $\alpha \in [0,1]$ tem-se:*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)y$$

- A função f diz-se estritamente convexa quando a desigualdade acima é estrita

- A função f diz-se fortemente convexa com $\gamma > 0$, quando para quaisquer $x, y \in D$ e $\alpha \in [0,1]$, tem-se que

$$f(\alpha x + (1-\alpha y)) \leq \alpha f(x) + (1-\alpha)f(y) - \gamma\alpha(1-\alpha)||x-y||^2$$

- Uma função função fortemente convexa é estritamente convexa

- Uma função estritamente convexa é convexa

Exemplo 4.1. *A função $f : \mathbb{R} \to \mathbb{R}, f(x) = x^2$ é fortemente convexa:*

Demonstração.

$$\begin{aligned}
f(\alpha x + (1-\alpha)y) &= \alpha^2 x^2 + y^2(1 - 2\alpha + \alpha^2) + 2xy\alpha(1-\alpha) \\
&= \alpha^2 x^2 + y^2 - 2y^2\alpha + y^2\alpha^2 + 2xy\alpha(1-\alpha) \\
&= \alpha^2 x^2 + y^2(1-\alpha) - y^2\alpha + y^2\alpha^2 + 2xy\alpha(1-\alpha) \\
&= \alpha^2 x^2 + y^2(1-\alpha) + y^2(-\alpha) + y^2\alpha^2 + 2xy\alpha(1-\alpha) \\
&= \alpha^2 x^2 + y^2(1-\alpha) + y^2(-\alpha) + y^2\alpha^2 + 2xy\alpha(1-\alpha) \\
&= \alpha^2 x^2 + y^2(1-\alpha) - \alpha(1-\alpha)y^2 + 2xy\alpha(1-\alpha) \\
&= \alpha^2 x^2 + y^2(1-\alpha) - \alpha(1-\alpha)(y^2 - 2xy)
\end{aligned}$$

Note agora que como

$$(x-y)^2 = (y^2 - 2xy + x^2) \geq y^2 - 2xy$$

Tomando agora $\gamma > 0$ suficientemente pequeno, temos

$$\begin{aligned}
f(\alpha x + (1-\alpha)y) &= \alpha^2 x^2 + y^2(1-\alpha) - \alpha(1-\alpha)(y^2 - 2xy) \\
&= \alpha^2 f(x) + (1-\alpha)f(x) - \alpha(1-\alpha)(y^2 - 2xy) \\
&\leq \alpha^2 f(x) + (1-\alpha)f(x) - \gamma\alpha(1-\alpha)(x-y)^2
\end{aligned}$$

Logo f é fortemente convexa $\square$

Definição 4.3. *Dizemos que*

$$\textit{min } f(x) \textit{ sujeito a } x \in D$$

é um problema de minimização convexa quando $D \subset \mathbb{R}^n$ é um conjunto convexo e $f : D \to \mathbb{R}$ é uma função convexa no conjunto D

Teorema 4.2. *Minimização convexa Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \to \mathbb{R}^n$ uma função convexa em D. Então todo minimizador loval do problema é tmabém um minimizador global. Além disso, o conjunto de minimizadores é convexo. Se f é estritamente convexa, não pode haver mais de um minimizador.*

Demonstração. Suponhamos que $\bar{x} \in D$ seja um minimizador local que não é global. Então existe $y \in DX$ tal que $f(y) < f(\bar{x})$. Definindo $x(\alpha) = \alpha y + (1 - \alpha)\bar{x}$. Pela convexidade de D, $x(\alpha) \in D, \forall \alpha \in [0, 1]$. Agora pela convexidade de f, para todo α, tem-se

$$\begin{aligned} f(x(\alpha)) &= f(\alpha y + (1 - \alpha)\bar{x}) \\ &\leq \alpha f(y) + (1 - \alpha) f(\bar{x}) \\ &= \alpha f(y) + f(\bar{x}) - \alpha f(\bar{x}) \\ &= f(\bar{x}) + \alpha(f(y) - f(\bar{x})) \\ &< f(\bar{x}) \end{aligned}$$

Portanto

$$f(x(\alpha)) \leq f(\bar{x}) + \alpha(f(y) - f(\bar{x})) < f(\bar{x})$$

Tomando $\alpha > 0$ suficientemente pequeno, podemos garantir ue o ponto $x(\alpha)$ é arbitrariamente próximo a $\bar{x}$, e ainda temos que $f(x(\alpha)) < f(\bar{x})$ e $\bar{x}$ é minimizador local do problema. Absurdo, pois assim $\bar{x}$ não é minimizador local don problema. Logo qualquer solução local do problema é também uma solução global.

Seja agora $S \subset D$ o conjunto dos minimizadores (globais) e $\bar{v} \in \mathbb{R}$ o valor ótimo do problema ($f(x) = \bar{v}, \forall x \in S$). Para quaisquer $x, \bar{x} \in S$ e $\alpha \in [0, 1]$, pela convexidade de f, obtemos

$$\begin{aligned} f(\alpha x + (1 - \alpha)\bar{x}) &\leq \alpha f(x) + (1 - \alpha) f(\bar{x}) \\ &= \alpha \bar{v} + \bar{v} - \alpha \bar{v} = \bar{v} \end{aligned}$$

Como também $\alpha x + (1 - \alpha)\bar{x} \in D$, concluímos que $f(\alpha x + (1 - \alpha)\bar{x} = \bar{v})$, isto é, $\alpha x + (1 - \alpha)\bar{x} \in S$, ou seja, S é convexo.

Suponhamos agora que f seja estritamente convexa e que existem $x, \bar{x} \in S$ com $x \neq \bar{x}$. Pela convexidade estrita de f tem-se que

$$f(\alpha x + (1-\alpha)\bar{x}) < \alpha f(x) + (1-\alpha) f(\bar{x}) = \bar{v}$$

Como o conjunto S é convexo, temos então, uma solução do problema $\alpha x + (1-\alpha)\bar{x}$, portanto:

$$f(\alpha x + (1-\alpha)\bar{x}) < \bar{v} \tag{5}$$

Temos assim que o valor da função objetivo menor que o valor ótimo do problema $\bar{v}$, que é uma contradição. Logo o minimizador é único.

Definição 4.4. *Se $D \subset \mathbb{R}^n$ é um conjunto convexo, dizemos que $f : D \to \mathbb{R}$ é uma função côncava em D, quando a função $-f$ é convexa em D*

Teorema 4.3. *Continuidade de funções convexas Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto, e $f : \Omega \to \mathbb{R}$ uma função convexa em Ω. Então f é Lipschitz-Contínua em Ω. Em particular f é contínua*

□

5 Classificação dos métodos e noções de convergência

Seja o problema de:

$$\min f(x) \quad \text{sujeito a} \quad x \in D$$

Onde $D \in \mathbb{R}^n$.

Todo método para resolução numérica do problema faz uso de cálculos dos valores da função, das restrições e de suas derivadas.

- **Método Passivo**: Quando os pontos de avaliação destes valores são escolhidos de acordo com uma estratégia definida a priori, sem levar em conta as informações obtida ao longo do processo iterativo

- **Método Sequencial**:Cada ponto é definido de acordo com a informação obtida nos pontos anteriores. Um método sequencial gera uma sequência de pontos no $\mathbb{R}^n$, como $x^1, x^2, \ldots, x^k, \ldots$, que serão chamados de aproximações à solução do problema.

Com frequência, algoritmos são descritos em formato de um esquema iterativo, como:

$$x^{k+1} = \Psi_k(x^k), \quad k = 0, 1, \dots$$

Onde Ψ_k é um operador cujos valores estão contidos em $\mathbb{R}^n$.

Naturalmente, supomos que na iteração $k+1$, Ψ_k seja conhecida. Se o esquema iterativo for fixado, a escolha do ponto inicial x^0 define uma trajetória do método se, e somente se,

$$\Psi_{k-1}(\Psi_{k-2}(\dots(\Psi_0(x^0))\dots)).$$

Algumas nomenclaturas importantes:

- **Ordem:** A ordem do método é a máxima ordem das derivadas da função objetivo e das restrições que estão sendo utilizadas.

- **Métodos com Convergência Assintótica:** Não há garantia de que algum x^k da trajetória seja uma solução do problema. A tarefa é provar que, para algum k suficientemente grande, x^k é uma boa aproximação da solução do problema.

5.1 Tipos de Convergência

(1) Se $\{x^k\} \to \bar{x}$, onde $\bar{x}$ é uma solução do problema, falamos de **convergência em relação às variáveis do problema** a partir de x^0.

(2) Quando não é possível provar convergência a uma solução específica, mas apenas garantir que **Dist.**$(x^k, S) \to 0$, onde S é o conjunto de soluções do problema, falamos de **convergência ao conjunto de soluções**.

(3) Se **Dist.**$(x^k, D) \to 0$ e $f(x^k) \to \bar{v}$, falamos de **convergência em relação à função objetivo**.

(4) No caso de problemas irrestritos, se $\{f'(x^k)\} \to 0$, falamos de **convergência em relação ao gradiente**.

(5) Quando $U_0 = \mathbb{R}^n$ e há convergência a partir de qualquer ponto $x^0 \in \mathbb{R}^n$, falamos de **convergência global**.

(6) Se algum tipo de convergência é garantido apenas para pontos suficientemente próximos à solução do problema, falamos de **convergência local do método**.

5.2 Taxas de convergência. Regras de parada

Suponhamos que um método dado seja convergente a uma solução $\bar{x}$, isto é, $\{x^k\} \to \bar{x}(k \to \infty)$. Resultados sobre a taxa de convergência fornecem uma garantia de velocidade na redução da distância $||x^{k+1} - \bar{x}||$ em relação a $||x^k - \bar{x}||$, quando k é suficientemente grande.

(1) Supomos que $x^0 \in \mathbb{R}^n$ é suficientemente próximo de $\bar{x}$, que $x^k \neq \bar{x}$ $\forall k \in \mathbb{N}$, e que todos os números constantes mencionados a seguir não dependem de x^0. Se existe $q \in (0,1)$ tal que:

$$\lim_{k\to\infty} \sup \frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} = q,$$

dizemos que o método possui **convergência linear**

(2) Convergência mais lenta que a linear é chamada de **Sublinear**

(3) Um exemplo que garante apenas taxa sublinear é a seguinte

$$\lim_{k\to\infty} \sup k||x^k - \bar{x}|| \leq C,$$

Onde $C > 0$ é uma constante. Neste caso falamos de **Convergência Aritmética**

(4) Quando:

$$\lim_{k\to\infty} \sup \frac{||x^k - \bar{x}||}{q^k} \leq C$$

Onde $q \in (0,1)$, falamos de **Convergência Geométrica**.

(5) Quando

$$\lim_{k\to\infty} \sup \frac{||x^k - \bar{x}||}{||x^k - \bar{x}||} = 0$$

falamos de **Convergência Superlinear**

(6) Um caso especial de convergência superlinear é a **Convergência quadrática**

$$\lim_{k\to\infty} \sup \frac{||x^k - \bar{x}||}{||x^k - \bar{x}||^2} \leq C$$

Temos também que a convergência quadrática é mais rápida que a convergência superlinear que é mais rápida que a convergência linear

6 Otimização irrestrita

6.1 Métodos de descida

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função dada. Uma das estratégias para resolver o problema irrestrito

$$\min f(x), x \in \mathbb{R}^n \tag{6}$$

é o seguinte, dada uma aproximação $x^k \in \mathbb{R}^n$ da solução do problema, encontramos um ponto $x^{k+1} \in \mathbb{R}^n$ tal que

$$f(x^{k+1}) < f(x^k)$$

Uma estratégia básica para isso é tomar uma direção $d^k \in \mathbb{R}^n$ tal que f é decrescente, e computar um comprimento de passo $\alpha_k > 0$ que fornece um valor de f menor que o ponto x^k, isto é,

$$f(x^k + \alpha_k d^k) < f(x^k)$$

Assim obtento o ponto $x^{k+1} = x^k + \alpha_k d^k$. Repetindo o processo, tomando $x^k = x^{k+1}$, construimos o esquema iterativo dos métodos de descida.

6.2 Esquema geral dos métodos de descida

Definição 6.1. *Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida da função $f : \mathbb{R}^n \to \mathbb{R}$ no ponto $x \in \mathbb{R}^n$, se existe $\epsilon > 0$ tal que*

$$f(x + td) < f(x), \forall t \in (o, \epsilon]$$

Denotamos por $\mathcal{D}_f(x)$ o conjunto de todas a direções de descidas da função f no ponto x

Lema 6.1. ***(Direções de descidas)**: Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $x \in \mathbb{R}^n$. Então:*

(a) Para todo $d \in \mathcal{D}_f(x)$, tem-se $\langle f'(x), d\rangle \leq 0$

(b) Se $d \in \mathbb{R}^n$ satisfaz $\langle f'(x), d\rangle < 0$, tem-se $d \in \mathcal{D}_f(x)$

Demonstração. Seja $d \in \mathcal{D}_f(x)$. Para todo $t > 0$ suficientemente pequeno, pela diferenciabilidade de f em x, temos:

$$\begin{aligned} -f(x+td) &= f(x) + \langle f'(x), td \rangle + o(t) \\ &= f(x) + t\left(\langle f'(x), d \rangle + \frac{o(t)}{t}\right). \end{aligned}$$

Pelo fato de d ser uma direção de descida, segue que

$$0 \geq f(x+td) - f(x) = t\left(\langle f'(x), d \rangle + \frac{o(t)}{t}\right).$$

Portanto:

$$0 \geq t\left(\langle f'(x), d \rangle + \frac{o(t)}{t}\right).$$

Dividindo por t e passando ao limite quando $t \to 0$, concluímos que:

$$0 \geq \langle f'(x), d \rangle.$$

o que prova o ítem (a)

Suponha agora que $\langle f'(x), d \rangle < 0$. Temos que:

$$f(x+td) - f(x) = t\left(\langle f'(x), d \rangle + \frac{o(t)}{t}\right).$$

Note que para $t > 0$ suficientemente pequeno, vale:

$$\langle f'(x), d \rangle + \frac{o(t)}{t} \leq \langle f'(x), d \rangle < 0,$$

Já que $\frac{o(t)}{t} \to 0$ e $\langle f'(x), d \rangle < 0$. Sendo assim:

$$\begin{aligned} f(x+td) - f(x) =& t\left(\langle f'(x), d \rangle + \frac{o(t)}{t}\right) < 0 \\ &\implies f(x+td) - f(x) < 0 \\ &\implies f(x+td) < f(x). \end{aligned}$$

O que prova o ítem (b)

□

O esquema geral dos métodos de descida consiste no seguinte

$$x^{k+1} = x^k + \alpha_k d^k, d^k \in \mathcal{D}_f(x^k), k = 1, 2, \ldots$$

Onde os valores do comprimento de passo $\alpha_k > 0$ são calculados para obter pelo menos a condição de que $f(x^{k+1}) < f(x^k)$, ou seja $\{f(x^k)\}$ é decrescente

Por $d^k \in \mathcal{D}_f(x^k)$, temos que $\alpha_k > 0$ sempre existe, portanto, obtemos um decréscimo dos membros de $\{x^k\}$ quando olhamos para as curvas de nível de f.

6.3 Buscas lineares

Apresentando principais regras de busca linear, supondo um pomto x^k dado e uma direção $d \in \mathcal{D}_f(x)$ já escolhida

6.3.1 Regra de Minimização Unidimensional

O objetivo desta regra é minimizar a função objetivo ao longo da semi-reta $x^k + \alpha d^k$, com $\alpha \geq 0$. Nessa regra, o comprimento de passo α_k é dado pela condição

$$f(x_k + \alpha_k d^k) = \min f(x^k + \alpha d^k)$$

Isto é, α_k é a solução do problema

$$\min \varphi_k(\alpha), \alpha \in \mathbb{R}_+ \tag{7}$$

onde $\varphi_k : \mathbb{R}_+ \to \mathbb{R}, \quad \varphi(\alpha) = f(x^k + \alpha d^k)$.

Note que, para $\alpha > 0$ suficientemente pequeno, temos que $\varphi(0) > \varphi(\alpha)$, pois, como $d^k \in \mathcal{D}_f(x^k)$, segue que $f(x^k) > f(x^k + \alpha d^k)$. Portanto, se α_k é o mínimo de $\varphi(\alpha)$, então $\varphi(\alpha_k) < \varphi(0)$. Assim, temos um problema de minimização unidimensional.

Se f é diferenciável, então $\varphi_k(\alpha)$ também é diferenciável. Com isso, se α_k é o minimizador de $\varphi_k(\alpha)$, segue, por 1.5, que:

$$0 = \varphi_k'(\alpha_k) = \langle f'(x^k + \alpha_k d^k), d^k \rangle = \langle f'(x^{k+1}), d^k \rangle$$

Do ponto de vista geométrico, o gradiente do ponto x^{k+1} é ortogonal à direção d^k, e x^{k+1} é o ponto de interseção da semi-reta a partir de x^k com a curva de nível de f que passa pelo ponto x^{k+1}. Este valor α_k é o melhor possível no sentido de que, no ponto correspondente, obtemos o menor valor de f entre todos os pontos da forma $x^k + \alpha d^k$.

6.3.2 Regra de Armijo

Enquanto a minimização unidimensional busca encontrar o valor exato do comprimento de passo α_k que minimiza a função objetivo ao longo de uma direção, a Regra de Armijo adota uma abordagem mais flexível. Ela garante que o comprimento de passo escolhido produza uma redução suficiente na função objetivo, evitando tanto passos excessivamente grandes, que podem ultrapassar o mínimo ou aumentar o valor da função, quanto passos muito pequenos, que podem comprometer a eficiência do algoritmo ao torná-lo lento.

Ideia da Regra de Armijo

- Considere o gráfico de $\varphi_k(\alpha) - \varphi_k(0)$
- Note que $\varphi_k(0) - \varphi_k(0) = 0$
- Como $d^k \in \mathcal{D}_f(x^k)$, para $\alpha > 0$ suficientemente pequeno, temos que $\varphi_k(\alpha) - \varphi_k(0) < 0$
- Tome a aproximação linear de $\varphi_k(\alpha) - \varphi_k(0)$ quando $\alpha = 0$ que é dada por $\alpha\varphi_k'(0)$
- Note que, para $\alpha > 0$ suficientemente pequeno temos que $\alpha\varphi_k'(0) < \varphi_k(\alpha) - \varphi_k(0)$
- Tome uma reta, com inclinação negativa, mas acima de $\alpha\varphi_k'(0)$, a partir da mesma
- Seja $\sigma\alpha\varphi_k'(0)$, com $\sigma \in (0, 1)$ essa reta (σ escolhido arbitrariamente)
- Como $\alpha\varphi_k'(0) < 0$ e $\sigma > 0$, segue que $\sigma\alpha\varphi_k'(0) < 0$ (Isso para α suficientemente pequeno)
- Tome um $\hat{\alpha} > 0$, e um parâmetro $\theta \in (0, 1)$ de modo que $\theta\hat{\alpha} < \hat{\alpha}$
- Note que como $\theta^{i_k}\hat{\alpha} \to 0$, quando $k \to \infty$ existe $\varphi_k(\theta^{i_k}\hat{\alpha}) - \varphi_k(0) < \sigma\theta^{i_k}\hat{\alpha}\varphi_k'(0)$
- O primeiro $\theta^{i_k}\hat{\alpha}$ que satisfaz a condição acima será escolhido como α_k
- Tomando $\alpha = \theta^{i_k}\hat{\alpha}$ temos que

$$\varphi_k(\alpha) - \varphi(0) = f(x^k + \theta^{i_k}\hat{\alpha}d^k) - f(x^k) = f(x^{k+1}) - f(x^k) < 0$$

Em resumo, a *regra de Armijo* consiste em computar um comprimento de passo que resulta em um decréscimo suficiente da função f em relação ao valor $f(x^k)$, que garanta convergência, isto é

$$f(x^k + \alpha d^k) \leq f(x^k) + \sigma\alpha\langle f'(x^k), d^k\rangle \tag{8}$$

em que f é diferenciável no ponto x^k, $\alpha > 0$ e $\sigma \in (0, 1)$.

O valor $\alpha\langle f'(x^k), d^k\rangle$ representa o valor de redução de f, previsto pela aproximação linear para α na direção d^k

Lema 6.2. ***(A regra de Armijo está bem definida)***

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto x^k. Suponhamos que $d^k \in \mathbb{R}^n$ satisfaça $\langle f'(x^k), d^k\rangle < 0$. Então a desigualdade (8) é satisfeita para todo $\alpha > 0$ suficientemente pequeno. Em particular, Armijo está bem definida e termina com $\alpha_k > 0$

Demonstração. Como f é diferenciável, para $\alpha > 0$ suficientemente pequeno, temos

$$\begin{aligned}
f(x^k + \alpha d^k) - f(x^k) &= \langle f'(x^k), \alpha d^k\rangle + o(\alpha) \\
&= \alpha\langle f'(x^k), d^k\rangle + \sigma\alpha\langle f'(x^k), d^k\rangle - \sigma\alpha\langle f'(x^k), d^k\rangle + o(\alpha) \\
&= \sigma\alpha\langle f'(x^k), d^k\rangle + (1-\sigma)\alpha\langle f'(x^k), d^k\rangle + o(\alpha) \\
&= \sigma\alpha\langle f'(x^k), d^k\rangle + \alpha((1-\sigma)\langle f'(x^k), d^k\rangle + \frac{o(\alpha)}{\alpha}) \\
&\leq \sigma\alpha\langle f'(x^k), d^k\rangle
\end{aligned}$$

Pois

$$(1-\sigma)\langle f'(x^k), d^k\rangle + \frac{o(\alpha)}{\alpha} \leq \frac{1-\sigma}{2}\langle f'(x^k), d^k\rangle < 0$$

Que vale para $\alpha > 0$ suficientemente pequeno. Portanto $f(x^k + d^k) - f(x^k) \leq \sigma\alpha\langle f'(x^k), d^k\rangle$, logo vale a desigualdade de armijo. □

Lema 6.3. ***(Cota inferior para o comprimento de passo dado por Armijo)***

Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no $\mathbb{R}^n$, com derivada Lipschitz-contínua no $\mathbb{R}^n$ com módulo $L > 0$. Se $x^k, d^k \in \mathbb{R}^n$ satisfaz 6.1, então a desigualdade (8)é válida para todo $\alpha \in (0, \bar{\alpha}_k)$, onde

$$\bar{\alpha}_k = \frac{2(\sigma-1)\langle f'(x^k), d^k\rangle}{L||d^k||^2}$$

Demonstração. Pelo lema (1), para todo $\alpha \in \mathbb{R}$, tem-se que

$$\begin{aligned} f(x^k + \alpha d^k) - f(x^k) &\leq \langle f'(x^k), \alpha d^k \rangle + \frac{L}{2} \cdot \alpha^2 ||d||^2 \\ &= \alpha(\langle f'(x^k), d^k \rangle + \frac{L}{2} \cdot \alpha ||d^k||^2) \end{aligned}$$

Logo, para $\alpha \in (0, \bar{\alpha}_k)$, segue

$$\begin{aligned} f(x^k + \alpha d^k) - f(x^k) &= \alpha(\langle f'(x^k), d^k \rangle + \frac{L}{2} \cdot \alpha ||d^k||^2) \\ &= \alpha(\langle f'(x^k), d^k \rangle + \frac{L}{2} |d^k||^2 \cdot \frac{2(\sigma - 1)\langle f'(x^k), d^k \rangle}{L||d^k||^2}) \\ &= \alpha(\langle f'(x^k), d^k \rangle + (\sigma - 1)\langle f'(x^k), d^k \rangle) \\ &= \alpha\sigma \langle f'(x^k), d^k \rangle \end{aligned}$$

Onde a última igualdade segue de (6.3) □

Interpretação do lema: No caso me que o gradiente da f é Lipschitz-contínuo, o valor de $f(x^k + \alpha d^k) - f(x^k)$ é menor ou igual a valor da função quadrática $\alpha \langle f'(x^k), d^k \rangle + \alpha^2 \frac{L||d^k||^2}{2}$, que é uma aproximação a taylor de segunda ordem de $\varphi_k(\alpha) - \varphi_k(0)$

Sob as hipóteses do lema anterior, se $\frac{\langle f'(x^k), d^k \rangle}{||d^k||^2} \leq \delta < 0$, onde δ é uma constante que não de pende k, e se os parâmetros α^*, σ, θ são os mesmos a cada iteração, segue que (8) é satisfeita para todo $\alpha \in (0, \bar{\alpha})$, onde

$$\bar{\alpha} = \frac{-2\delta(1 - \sigma)}{L} > 0$$

Como um α_k maior não foi aceito, ou $\alpha_k = \alpha^*$, ou então $\frac{\alpha_k}{\theta} > \bar{\alpha}$. Concluimos então que $\alpha_k \geq \min \{\alpha^*, \theta\bar{\alpha}\} = \breve{\alpha} > 0$

6.3.3 Regra de Goldstein

Nessa regra, o comprimento de passo $\alpha_k > 0$, é calculado para satisfazer às seguintes desigualdades em relação a α

1. $\sigma_1 \alpha \langle f'(x^k), d^k \rangle \leq f(x^k + \alpha d^k) - f(x^k)$
2. $f(x^k + \alpha d^k) - f(x^k) \leq \alpha\sigma_2 \langle f'(x^k), d^k \rangle$ (Armijo)

Onde $0 < \sigma_1 < \sigma_2 < 1$

Note que a primeira desigualdade é facilmente violada para $\alpha > 0$ suficientemente pequeno, por conta disso ela tem como objetivo evitar passos muitos curtos que seria aceito pela regra de Armijo. Essas duas desigualdades pode ser escrita da seguinte forma

$$\sigma_1 \leq \frac{f(x^k + \alpha d^k) - f(x^k)}{\alpha \langle f'(x^k), d^k \rangle} \leq \sigma_2 \tag{9}$$

Ou ainda

$$\sigma_1 \alpha \langle f'(x^k), d^k \rangle \leq f(x^k + \alpha d^k) - f(x^k) \leq \sigma_2 \alpha \langle f'(x^k), d^k \rangle$$

Ideia da Regra de Goldstein

- Considere a função $\varphi_k(\alpha) - \varphi_k(0)$, a desigualdade de Armijo e suas conclusões
- Como $0 < \sigma_1 < \sigma_2 < 0$, temos $\sigma_1 \alpha \langle f'(x^k), d^k \rangle > \sigma_2 \alpha \langle f'(x^k), d^k \rangle$, para $\alpha > 0$
- Com isso, existem valores de α_k tal que $\sigma_1 \alpha_k \langle f'(x^k), d^k \rangle \leq f(x^k + \alpha_k d^k) - f(x^k) \leq \sigma_2 \alpha_k \langle f'(x^k), d^k \rangle$

Tais valores α_k serão aceito.

Teorema 6.1. ***(A regra de Goldstein está bem definida)*** *Sejam $f : \mathbb{R}^n \to \mathbb{R}$ uma função continuamente diferenciável e limitada inferiormente ao longo do conjunto $\{x^k + \alpha d^k; \alpha > 0\}$ e d^k uma direção de descida a partir de x^k. Se $0 < \sigma_1 < \sigma_2 < 1$, então existe um intervalo de comprimentos de passo que satisfazem simultaneamente as condições de Goldstein.*

Demonstração. Por hipótese, temos que f é contínua e limitada inferiormente, então a restrição de f na direção d_k a partir de x_k dada por $\varphi(\alpha) = f(x_k + \alpha d_k)$ é contínua e limitada inferiormente para todo $\alpha > 0$. Além disso, temos que d_k é direção de descida, ou seja, $\langle f'(x^k), d^k \rangle < 0$, o que nos fornece

$$\lim_{\alpha \to \infty} l_1(\alpha) = \lim_{\alpha \to \infty} l_2(\alpha) = -\infty.$$

Note que a função $l_1(\cdot)$ tem a inclinação negativa dada por $\sigma_1 \alpha_k \langle f'(x^k), d^k \rangle$, mas como $\sigma_1 \in (0, 1)$, a reta l_1 fica acima do gráfico de φ para pequenos valores positivos de α (ou para α próximo de 0), e o mesmo ocorre para a reta l_2. Então, existe α' próximo de zero tal que $\varphi(\alpha') = c' < l_2(\alpha')$.

Por outro lado, pela definição de limite infinito no infinito, temos que, dado $c'' < 0$, existe $N > 0$ tal que $\alpha'' > N$ implica $\varphi(\alpha'') \geq c'' > l_2(\alpha'')$. Pelo Teorema do Valor Intermediário, existe $\alpha''' \in (\alpha', \alpha'')$ tal que $\varphi(\alpha''') = c''' = l_2(\alpha''')$, ou seja, φ intersecta l_2.

Utilizando um raciocínio análogo, pode-se mostrar que φ também intersecta l_1. Note também que $\sigma_1 < \sigma_2$, então o coeficiente angular de l_2 é menor do que o coeficiente angular de l_1, ou seja, l_2 decresce mais rapidamente do que l_1, o que implica que $l_2(\alpha) < l_1(\alpha)$, para todo $\alpha > 0$.

Disso decorre que $\varphi(\alpha)$ intersecta $l_2(\alpha)$ antes de intersectar $l_1(\alpha)$. Além disso, $\varphi(\alpha)$ intersecta l_1 e l_2 em no máximo um número finito de vezes, já que $\varphi(\alpha)$ é limitada inferiormente, enquanto l_1 e l_2 não são.

Seja α_1 o comprimento do passo que gera a primeira interseção de φ com l_1 e α_2 o maior comprimento de passo menor do que α_1 onde φ intersecta l_2. O intervalo (α_2, α_1) é formado por comprimentos de passo que satisfazem as condições de Goldstein. O mesmo critério pode ser usado para os outros pontos de interseção de φ com l_1 e l_2. $\square$

6.3.4 Regra de Wolfe

Nessa Regra o comprimento de passo $\alpha_k > 0$ é calculado para satisfazer às seguintes desigualdades em relação a α

1. $f(x^k + \alpha d^k) \leq f(x^k) + \sigma_1 \alpha \langle f'(x^k), d^k \rangle$ (Armijo)
2. $\langle f'(x^k + \alpha d^k), d^k \rangle \geq \sigma_2 \langle f'(x^k), d^k \rangle$

Onde $0 < \sigma_1 < \sigma_2 < 1$.

Novamente a primeira desigualdade é a regra de Armijo. Já a segunda serve como controle para a minimização de f ao longo de $x^k + \alpha d^k$, onde $\alpha \geq 0$. Como visto na minimização unidimensional, se $\alpha > 0$ é mínimo de $\varphi(\alpha)$ então $\langle f'(x^k + \alpha d^k), d^k \rangle = 0$. Portanto a segunda desigualdade é um teste de qualidade para um $\alpha > 0$ dado como uma aproximação de minimizadores de f na semi-reta ao longo de d^k. (Isso ocorre pois como $\sigma_2 \langle f'(x^k), d^k \rangle$ e $\langle f'(x^k + \alpha d^k), d^k \rangle$ são negativos, se a desigualdade é satisfeita, a primeira parcela se aproxima de 0)

Então qualquer implementação da regra de Wolfe, para o qual tem-se que $\breve{\alpha} \to \infty$ no caso de um número infinito de extrapolações e $(\hat{\alpha} - \breve{\alpha}) \to 0$ no caso de um número infinito de interpolações, está bem definida, no sentido de que ela termina com um valor de comprimento de passo $\alpha_k > 0$

7 Método do gradiente

Método do gradiente é um dos métodos de descida que usa a direção do gradiente como direção de descida. Supnhamos que $f : \mathbb{R}^n \to \mathbb{R}$ seja uma função diferenciável, tomamos como direção de descida $d^k = -f'(x^k)$, note que $\langle f'(x^k), -f'(x^k)\rangle = -||f'(x^k)||^2 < 0$, para $f'(x^k) \neq 0$. A iteração no método do gradiente é dada da seguinte forma

$$x^{k+1} = x^k - \alpha_k d^k$$

Analisando no contexto da minimização Uni-dimensional, temos que vale a igualdade

$$\langle f'(x^{k+1}), f'(x^k)\rangle = 0$$

Que é válida para $0 = \varphi_k'(\alpha)$. Ou seja, $d^K = -f'(x^k)$ e $d^k = -f'(x^{k+1})$, são ortogonais, o que explica a trajetória em "zig - zag"do método do gradiente

No contexto da Regra de Armijo, temos que a desigualdade é da seguinte forma

$$f(x^k - \alpha f'(x^k)) \leq f(x^k) - \sigma\alpha||f'(x^k)||^2$$

Note que quando $f'(x^k) \neq 0$, temos que $\frac{\langle f'(x^k), d^k\rangle}{||d^k||^2} \leq \delta < 0$ é satisfeita tomando $\delta = -1$, logo a estimativa de passo mais lngo é dado por

$$\bar{\alpha_k} = \frac{2(1-\sigma)}{L}$$

Como essa estimativa não depende de k quando o gradiente é Lipschitz-contínuo pelo lema (6.2) e suas consequências temos

$$\alpha_k \geq \breve{\alpha} > 0$$

Onde $\breve{\alpha}$ não depende de k

Teorema 7.1. ***(Convergência Global do Método do Gradiente I)*** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável em $\mathbb{R}^n$ com derivada Lipschitz-contínua em $\mathbb{R}^n$ com módulo $L > 0$. Então, se uma sequência $\{x^k\}$ gerada pelo método do gradiente (equipado com a Regra de Armijo) possui um ponto de acumulação, ou se f é limitada inferiormente em $\mathbb{R}^n$, tem-se* *que*

$$\{f'(x^k)\} \to 0 \quad (k \to \infty)$$

Em particular, qualquer ponto de acumulação de $\{x^k\}$ é um ponto estacionário.

Demonstração. Se $f'(x^k) \neq 0, \forall k$, então $f(x^{k+1}) < f(x^k)$, o que implica que a sequência $\{f(x^k)\}$ é decrescente. Suponha que $\{x^{k_j}\} \to a$, onde $\{x^{k_j}\}$ é uma subsequência de $\{x^k\}$, e a é um ponto de acumulação da sequência. Pela continuidade de f, temos que $\{f(x^{k_j})\} \to f(a)$. No caso em que f é limitada inferiormente em $\mathbb{R}^n$, concluímos que $\{f(x^k)\}$ é monótona e decrescente, portanto converge. Sendo assim, $f(x^k) - f(x^{k+1}) \to 0, (k \to \infty)$.

Pela regra de Armijo, temos:

$$\begin{aligned}
&f(x^k - \alpha_k f'(x^k)) \leq f(x^k) - \sigma\alpha_k ||f'(x^k)||^2 \\
&\Rightarrow f(x^{k+1}) - f(x^k) \leq -\sigma\alpha_k ||f'(x^k)||^2 \\
&\Rightarrow f(x^k) - f(x^{k+1}) \geq \sigma\alpha_k ||f'(x^k)||^2
\end{aligned}$$

Como $f'(x^k)$ é Lipschitz-contínuo, temos que vale $\alpha_k \leq \breve{\alpha} \leq 0$. Sendo assim,

$$f(x^k) - f(x^{k+1}) \geq \sigma\alpha_k ||f'(x^k)||^2 \geq \sigma\breve{\alpha} ||f'(x^k)||^2 \geq 0$$

Tomando limite quando $k \to \infty$, temos pelo teorema do confronto que $f'(x^k) = 0$. Como f é contínua, segue que $\lim_{k\to\infty} f(x^k) = f(a)$. Como f é contínua, $\{f'(x^k)\} \to 0$.

Obs.: Esse mesmo argumento é válido para a subsequência $\{x^{k_j}\}$. □

No caso da minimização unidimensional e Armijo, podemos substirtuir a hipótese de $f'(x^k)$ ser Lipschitz-contínuo, por $f'(x^k)$ ser apenas contínua, que é uma condição mais frac. Note que $f'(x)$ não ser Lipschitz-contínuo, implica que nem sempre existe $\breve{\alpha} > 0$, tal que $\alpha_k \geq \breve{\alpha} > 0$, onde $\breve{\alpha}$ não de pende de k, ou seja, possa ser que $\alpha_k \to 0$

Teorema 7.2. ***(Convergência global do Método do gradiente II)*** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferênciável no $\mathbb{R}^n$ com derivada contínua, seja também a sequência $\{x^k\}$, gerada pelo método do gradiente (equipado com a Regra de Armijo). Então cada ponto de acumlção da sequência é um ponto estacionśrio do problema. Se $\{x^k\}$ é limitada vale*

$$\{f'(x^k)\} \to 0(k \to \infty)$$

Demonstração. Suponha que $\{x^k\}$ tenha um ponto de acumlção $\bar{x} \in \mathbb{R}^n$. Suponhamos que a subsequência $\{x^{k_j}\} \to \bar{x}$, caso exista $\breve{\alpha} > 0$ tal que $\alpha_{k_j} \geq \breve{\alpha} > 0, \forall j$, onde $\breve{\alpha}$, de maneira análoga ao teorema anterior concluimos que $f'(\bar{x}) = 0$, quando $(j \to \infty)$.

Caso não $\breve{\alpha}$ com tal propriedade, podemos supor que $\alpha_{k_j} \to 0$. Pela maneira que a Regra de armijo é construida (Tomando $\hat{\alpha}, \theta\hat{\alpha}, \theta^2\hat{\alpha}, \dots$), para algum j suficientemente grande, temos que α_{k_j} satisfaz Armijo, isso implica que $\theta^{-1}\hat{\alpha}$ não satifaz Armijo, ou seja

$$f(x^{k_j} - \theta^{-1}\alpha_{k_j} f'(x^{k_j})) > f(x^{k_j}) - \sigma\theta^{-1}\alpha_{k_j}||f'(x^{k_j})||^2$$

Denotando por $\tilde{\alpha}_{k_j} = \theta^{-1}\alpha_{k_j}||f'(x^{k_j})||$, temos

$$\frac{f(x^{k_j} - \tilde{\alpha}_{k_j}\frac{f'(x^{k_j})}{||f'(x^{k_j})||}) - f(x^{k_j})}{\tilde{\alpha}_{k_j}} > -\sigma||f'(x^{k_j})|| \tag{10}$$

Como $\varphi_{k_j}(\alpha) = f(x^{k_j} + \alpha d^{k_j})$, onde $d^{k_j} = -\frac{f'(x^{k_j})}{||f'(x^{k_j})||}$, de (10), temos

$$\frac{\varphi_{k_j}(\tilde{\alpha_{k_j}}) - \varphi_{k_j}(0)}{\tilde{\alpha_{k_j}}} \geq -\sigma||f'(x^{k_j})||$$

Note que pelo teorema do valor intermediário segue que

$$\frac{\varphi_{k_j}(\tilde{\alpha_{k_j}}) - \varphi_{k_j}(0)}{\tilde{\alpha_{k_j}}} = \varphi'_{k_j}(\overset{*}{\alpha}_{k_j}) = -\langle f'(x^{k_j} - \overset{*}{\alpha}_{k_j} d^{k_j}), d^{k_j}\rangle$$

Onde $0 \leq \overset{*}{\alpha}_{k_j} \leq \tilde{\alpha_{k_j}}$

Segue que

$$-\langle f'(x^{k_j} - \overset{*}{\alpha}_{k_j} d^{k_j}), d^{k_j}\rangle \geq -\sigma||f'(x^{k_j})||$$

Ou ainda

$$\langle f'(x^{k_j} - \overset{*}{\alpha}_{k_j} d^{k_j}), d^{k_j}\rangle \leq \sigma||f'(x^{k_j})||$$

Como $d^{k_j} = -\frac{f'(x^{k_j})}{||f'(x^{k_j})||}, \alpha_{k_j} \to 0$, quando $(j \to \infty)$ e $\{x^{k_j}\} \to \bar{x}$, temos que ao passar o limite quando $(j \to \infty)$

$$||f'(\bar{x})|| < \sigma||f'(\bar{x})|| \Rightarrow 0 \geq ||f'(\bar{x})||(1 - \sigma)$$

E como $\sigma \in (0,1)$, segue $f'(\bar{x}) = 0$

Como cada ponto de acumulação de qualquer sequência $\{x^k\}$ é um ponto estacionário do problema de minimizar $f(x)$, $x \in \mathbb{R}^n$, concluímos que

$$f'(x^k) \to 0 \quad (k \to \infty).$$

$\square$

Obs.: A existênciade pontos de acumulação ão faz parte da afirmação do teorema. uma hipótese que garante tal existência é supor que o conjuntp $L(f(x)) = \mathcal{L}_{f,\mathbb{R}^n} f(x^0)$ seja limitado. Como $d \in \mathcal{D}f(x^0)$, temos que $f(x^k) < f(x^0)$ que implica $\{x^k\} \subset L(f(x^0))$, portanto $\{x^k\}$ será limitado e terá algum ponto de acumulação

Referências

[1] A. Izmailov and M. Solodov. *Otimização, Volume 2: métodos computacionais*. IMPA, 2018.

[2] A. Izmailov and M. Solodov. *Otimização, Volume 1*. Rio de Janeiro: IMPA, 2020.