Notas de estudo

Métodos do Gradiente Projetado

Emanuel Mendes Queiroz
Samara Viriato Vilar Dias

Vitória da Conquista, 2023

Conteúdo

1 Resultados preliminares

Nesta seção, apresentamos uma seleção de teoremas e definições fundamentais que servirão como base para o desenvolvimento das ideias e conceitos abordados ao longo destas notas de estudo.

1.1 Resultados básicos

Teorema 1.1. *(**Teorema de Weierstrass**) Sejam $D \subset \mathbb{R}^n$ um conjunto compacto não-vazio e $f : D \to \mathbb{R}$ uma função contínua. Então os problemas de minimizar e maximizar f em D têm soluções globais.*

Demonstração: Sabemos que qualquer problema de maximização

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

pode ser transformado em um problema de minimização equivalente:

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

Assim, é suficiente provar a existência de um minimizador ou de um maximizador. Mostraremos a existência de um minimizador.
Como a imagem de um conjunto compacto por uma função contínua é compacta, $\{v \in \mathbb{R}; v = f(x) \text{ para algum } x \in D\}$ é compacto. Em particular, este conjunto é limitado inferiormente, ou seja,

$$-\infty < \overline{v} = \inf_{x \in D} f(x).$$

Pela definição de ínfimo, para todo $k \in \mathbb{N}$ existe um $x_k \in D$ tal que

$$\overline{v} \leq f(x_k) \leq \overline{v} + 1/k.$$

Passando ao limite quando $k \to \infty$, concluímos que

$$\lim_{k \to \infty} f(x_k) = \overline{v}. \tag{1}$$

Como $\{x_k\} \subset D$ e D é compacto, segue-se que $\{x_k\}$ é um sequência limitada. Logo, ela possui uma subsequência $\{x_{k_j}\}$ que converge a um ponto de D:

$$\lim_{j \to \infty} x_{k_j} = \overline{x} \in D.$$

Pela continuidade de f,

$$\lim_{j \to \infty} f(x_{k_j}) = f(\overline{x}).$$

Usando (1), temos que $f(\overline{x}) = \overline{v}$, i. e., f assume o valor mínimo em D no ponto $\overline{x} \in D$. Em outras palavras, $\overline{x}$ é um minimizador global do problema de $\min f(x)$ sujeito a $x \in D$.

Definição 1.2. *O conjunto de nível da função $f : D \to \mathbb{R}$ associado a $c \in \mathbb{R}$, é o conjunto dado por $L_{f,D}(c) = \{x \in D; f(x) \leq c\}$.*

Corolário 1.3. *Sejam $D \subset \mathbb{R}^n$ e $f : D \to \mathbb{R}$ contínua no conjunto D. Suponhamos que existe $c \in \mathbb{R}$ tal que o conjunto de nível $L_{f,D}(c)$ seja não-vazio e compacto. Então o problema de minimizar f em D possui uma solução global.*

Demonstração: Pelo Teorema 1.1, o problema *min* $f(x)$ sujeito a $x \in L_{f,D}(c)$ tem uma solução global, digamos $\overline{x}$. Para todo ponto $x \in D \setminus L_{f,D}(c)$, temos que $f(x) > c \geq f(\overline{x})$, o que mostra que $\overline{x}$ é um minimizador global de f não só em $L_{f,D}(c)$, mas também em D. □

Definição 1.4. *Dados conjunto $D \subset \mathbb{R}^n$ e ponto $x \in \mathbb{R}^n$, uma projeção (ortogonal) de x sobre D é uma solução global do problema*

$$min\ \|y - x\|\ \textit{sujeito a}\ y \in D. \tag{2}$$

Em outras palavras, a projeção de x sobre D é um dos pontos de D que é mais próximo de x. Dependendo das hipóteses sobre o conjunto D, pode haver mais de um ponto com esta propriedade, um único ponto ou nenhum.

Exemplo 1.5. *Casos de existência da projeção:*

- *Todo ponto da esfera $D = \{y \in \mathbb{R}^n; \|y\| = 1\}$ é uma projeção da origem $x = 0$ sobre D.*
- *A projeção de $x = 1$ sobre $D = [-1, 0)$ não existe.*
- *A projeção de $x = 1$ sobre $D = [-1, 0]$ é o único ponto $\overline{x} = 0$.*

Teorema 1.6. ***(Teorema da projeção)***

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

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

$$\overline{x} \in D, \langle x - \overline{x}, y - \overline{x} \rangle \leq 0,\ \forall y \in D. \tag{3}$$

Também tem-se que

$$\|P_D(x) - P_D(y)\| \leq \|x - y\|, \forall x \in \mathbb{R}^n, \forall y \in \mathbb{R}^n. \tag{4}$$

Demonstração:

(a) Fixemos $x \in \mathbb{R}^n$ arbitrário. Seja

$$f : \mathbb{R}^n \to \mathbb{R}_+,\ f(x) = \|y - x\|.$$

Evidentemente, f é contínua no $\mathbb{R}^n$ e

$$L_{f,D}(c) = D \cap B(x, c),$$

onde

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

ou seja, o conjunto de nível $L_{f,D}(c)$ é a interseção do conjunto fechado D com o conjunto compacto $B(x,c)$. Portanto, $L_{f,D}(c)$ é compacto. Além disso, para $c > 0$ suficientemente grande, é óbvio que a bola $B(x,c)$ contém pontos de D. Logo, $L_{f,D}(c) \neq \varnothing$. Pelo Corolário 1.3 temos que o problema de minimizar f em D possui uma solução global. Assim, a projeção de x sobre D existe (pela definição de projeção de x sobre D).

(b) Sejam $\overline{x}$ uma solução do problema (2) e $y \in D$ qualquer. Como $\overline{x} \in D$ e D é convexo, temos pela definição de combinação convexa que $(1-\alpha)\overline{x} + \alpha y = x(\alpha) \in D$ para todo $\alpha \in (0,1]$. Temos, então, que $\|x - \overline{x}\| \leq \|x - x(\alpha)\| = \|x - ((1-\alpha)\overline{x} + \alpha y)\| = \|x - \overline{x} + \alpha\overline{x} + \alpha y\|$ e $\|x - x(\alpha)\| \geq \|x - \overline{x}\| \Longrightarrow \|x - x(\alpha)\|^2 \geq \|x - \overline{x}\|^2$. Assim,

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

Dividindo ambos os lados da desigualdade acima por $2\alpha > 0$, temos

$$0 \geq \langle x - \overline{x}, y - \overline{x}\rangle - \frac{\alpha}{2}\|y - \overline{x}\|^2.$$

Passando ao limite quando $\alpha \to 0+$, obtemos

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

sendo que $y \in D$ era arbitrário.

Suponhamos agora que um certo $\overline{x}$ satisfaça (3). Então, para todo $y \in D$,

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

Agora, vamos mostrar que a projeção é única. Seja $\hat{x}$ alguma outra solução de (2). Usando (3) para $\overline{x}$ com $y = \hat{x} \in D$ e para $\hat{x}$ com $y = \overline{x} \in D$, temos

$$\langle x - \overline{x}, \hat{x} - \overline{x} \rangle \leq 0 \text{ e } \langle x - \hat{x}, \overline{x} - \hat{x} \rangle \leq 0 \Longrightarrow -\langle x - \hat{x}, \hat{x} - \overline{x} \rangle \leq 0.$$

Somando, obtemos que

$$\begin{aligned} 0 &\geq \langle x - \overline{x}, \hat{x} - \overline{x} \rangle - \langle x - \hat{x}, \hat{x} - \overline{x} \rangle \\ &= \langle x - \overline{x} - (x - \hat{x}), \hat{x} - \overline{x} \rangle \\ &= \langle \hat{x} - \overline{x}, \hat{x} - \overline{x} \rangle \\ &= \|\hat{x} - \overline{x}\|^2. \end{aligned}$$

Logo, $\hat{x} = \overline{x}$.

Por fim, vamos provar (4). Como $P_D(x) \in D$ e $P_D(y) \in D$, usando (3) duas vezes (para x e y respectivamente), obtemos

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

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

Somando as duas desigualdades, temos que

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

Então temos $0 \geq \|P_D(x) - P_D(y)\|^2 + \langle y - x, P_D(x) - P_D(y) \rangle \Longrightarrow -\langle y - x, P_D(x) - P_D(y) \rangle \geq \|P_D(x) - P_D(y)\|^2 \Longrightarrow \langle x - y, P_D(x) - P_D(y) \rangle \geq \|P_D(x) - P_D(y)\|^2 \geq 0$.

Usando agora a desigualdade de Cauchy-Schwarz, obtemos

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

Se $P_D(x) = P_D(y)$, (4) vale trivialmente. Caso contrário, obtemos (4) dividindo os dois lados da desigualdade acima por $\|P_D(x) - P_D(y)\| > 0$.

$\square$

Teorema 1.7. ***(Teorema do Valor Médio)*** *Vale o seguinte*

(a) *Se para $x, y \in \mathbb{R}^n$ uma função $F : \mathbb{R}^n \to \mathbb{R}^l$ é continuamente diferenciável no intervalo $\{x + ty \mid t \in [0, 1]\}$, então*

$$F(x + y) = F(x) + \int_0^1 F'(x + ty) y \, dt.$$

(b) *Se $f : \mathbb{R}^n \to \mathbb{R}$ é continuamente diferenciável num conjunto convexo e aberto $\Omega \subset \mathbb{R}^n$, então para todo $x, y \in \Omega$ existe $t \in [0,1]$ tal que*

$$f(y) - f(x) = \langle \nabla f(tx + (1-t)y), y - x \rangle .$$

Lema 1.8. *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no $\mathbb{R}^n$, com gradiente Lipschitz-contínuo no $\mathbb{R}^n$ com módulo $L > 0$. Então*

$$|f(x+y) - f(x) - \nabla f(x)^T y| \leq \frac{L\|y\|^2}{2}, \ \forall x, y \in \mathbb{R}^n.$$

Demonstração: Pelo Teorema 1.7 (a), temos que $f(x+y) = f(x) + \int_0^1 \nabla f(x+ty)^T y \, dt$. Assim,

$$\begin{aligned} |f(x+y) - f(x) - \nabla f(x)^T y| &= |f(x) + \int_0^1 \nabla f(x+ty)^T y \, dt - f(x) - \nabla f(x)^T y| \\ &= |\int_0^1 \nabla f(x+ty)^T y \, dt - \nabla f(x)^T y| \\ &= |\int_0^1 (\nabla f(x+ty) - \nabla f(x))^T y \, dt| \\ &\leq \int_0^1 |(\nabla f(x+ty) - \nabla f(x))^T y| \, dt. \end{aligned}$$

Pela Desigualdade de Cauchy-Schawrz, temos que

$$\int_0^1 |(\nabla f(x+ty) - \nabla f(x))^T y| \, dt \leq \int_0^1 \|\nabla f(x+ty) - \nabla f(x)\| \|y\| \, dt.$$

Além disso, temos (por hipótese) que ∇f é Lipschitz-contínuo no $\mathbb{R}^n$, ou seja, $\|\nabla f(x+ty) - \nabla f(x)\| \leq L\|(x+ty) - x\| = L\|ty\| = Lt\|y\|$, logo

$$\begin{aligned} \int_0^1 \|\nabla f(x+ty) - \nabla f(x)\| \|y\| \, dt &\leq \int_0^1 Lt\|y\|^2 \, dt \\ &= \left[L\frac{t^2}{2}\|y\|^2 \right]_0^1 \\ &= \frac{L\|y\|^2}{2}. \end{aligned}$$

□

Teorema 1.9. ***(Caracterizações de funções convexas diferenciáveis)*** *Sejam $\Omega \subset \mathbb{R}^n$ um conjunto convexo e aberto e $f : \Omega \to \mathbb{R}$ uma função diferenciável em Ω. Então, as propriedades seguintes são equivalentes:*

(a) *A função f é convexa em Ω;*

(b) *para quaisquer $x, y \in \Omega$,*

$$f(y) \geq f(x) + \langle \nabla f(x_k), y - x \rangle .$$

Demonstração: Veja a demonstração em [2, Teorema 3.4.30].

1.2 Condições de otimalidade em forma primal para problemas com restrições. Cone tangente.

Vamos considerar agora um problema no formato geral

$$min\ f(x) \ \text{ sujeito à } x \in D \tag{5}$$

onde $D \subset \mathbb{R}^n$ é um dado conjunto (não vazio) cuja estrutura por enquanto não é especificada, e $f : \mathbb{R}^n \to \mathbb{R}$ é a função objetivo.
A fim de desenvolver as condições de otimalidade no caso de problemas com restrições, precisamos estudar não só o comportamento da função objetivo numa vizinhança de uma solução (como no caso irrestrito), mas também a estrutura do conjunto viável nessa vizinhança. As noções de direções viáveis e direções tangentes são fundamentais neste contexto e são muito importantes para um entendimento correto das condições de otimilidade. Para introduzir conjuntos de direções, a seguinte noção é útil.

Definição 1.10. *Um conjunto $K \subset \mathbb{R}^n$ chama-se cone quando ele contém todos os múltiplos não-negativos de seus elementos:*

$$d \in K \Longrightarrow td \in K \quad \forall t \in \mathbb{R}_+.$$

Pela definição, se K é um cone não vazio, necessariamente $0 \in K$ (basta tomar $t = 0$). Alguns exemplos de cone são: o espaço $\mathbb{R}^n$, qualquer subespaço de $\mathbb{R}^n$ e o ortante não-negativo $\mathbb{R}^n_+$. Informalmente, um cone é um conjunto de direções, veja Figura 7. Um cone não vazio sempre é ilimitado, com exceção do cone trivial $\{0\}$. Um cone pode ser não fechado, por exemplo $\{x \in \mathbb{R}^n | x_i > 0, i = 1, 2, \cdots, n\}$.

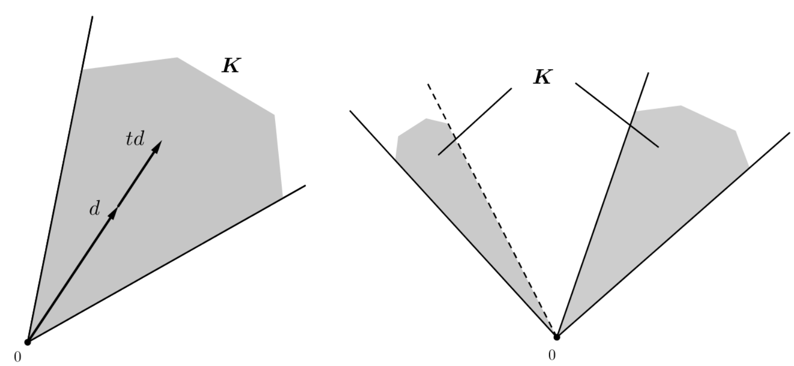


Figura 1: Exemplos de cones.

1.2.1 Direções viáveis

Definição 1.11. *Dizemos que $d \in \mathbb{R}^n$ é uma direção viável em relação ao conjunto D no ponto $\overline{x} \in D$, quando existe $\varepsilon > 0$ tal que*

$$\overline{x} + td \in D \quad \forall t \in [0, \varepsilon].$$

Denotamos por $\mathcal{V}_D(\overline{x})$ o conjunto de todas as direções viáveis em relação ao conjunto D no ponto $\overline{x} \in D$. A Figura 2 ilustra a definição de direções viáveis.

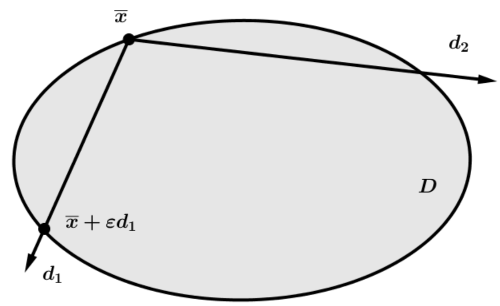


Figura 2: As direções d_1 e d_2 são viáveis em relação ao conjunto D no ponto $\overline{x} \in D$. Por exemplo, tem-se que $\overline{x} + td_1 \in D$ para todo $t \in [0, \varepsilon]$.

Para $D \subset \mathbb{R}^n$ e $\overline{x} \in D$ quaisquer, o conjunto de direções viáveis $\mathcal{V}_D(\overline{x})$ sempre é um cone, já que se $\overline{x}+td \in D$ para todo $t \in [0, \varepsilon]$ então $\overline{x}+t(\alpha d) \in D$ para todo $\alpha > 0$ e todo $t \in [0, \varepsilon/\alpha]$. Além disso, o cone $\mathcal{V}_D(\overline{x})$ sempre é não-vazio (pelo menos, $0 \in \mathcal{V}_D(\overline{x})$). Para $\overline{x} \in int\ D$, temos que $\mathcal{V}_D(\overline{x}) = \mathbb{R}^n$.

1.2.2 Direções de descida

Em relação ao comportamento da função objetivo, a seguinte noção é fundamental.

Definição 1.12. *Dizemos que $d \in \mathbb{R}^n$ é uma direção de descida de $f : \mathbb{R}^n \to \mathbb{R}$ no ponto $\overline{x} \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que*

$$f(\overline{x} + td) < f(\overline{x}) \quad \forall t \in (0, \varepsilon).$$

Denotamos por $\mathcal{D}_f(\overline{x})$ o conjunto de todas as direções de descida da função f no ponto $\overline{x} \in \mathbb{R}^n$.

É claro que o conjunto $\mathcal{D}_f(\overline{x})$ pode ser vazio (por exemplo, este o caso quando $\overline{x}$ é um minimizador irrestrito de f, mesmo que seja local). Quando $\mathcal{D}_f(\overline{x})$ é não-vazio, ele não é um cone (porque $0 \notin \mathcal{D}_f(\overline{x})$). No entanto, o conjunto $\mathcal{D}_f(\overline{x}) \cup 0$ é um cone não-vazio.

Exemplo 1.13. *Se $\mathcal{D}_f(\overline{x}) = \emptyset$ então nada podemos falar sobre cl $\mathcal{D}_f(\overline{x})$. Se $\mathcal{D}_f(\overline{x}) \neq \emptyset$ então cl $\mathcal{D}_f(\overline{x})$ (o fecho do conjunto $\mathcal{D}_f(\overline{x})$) é um cone que é não vazio. Para provar isto, basta mostrarmos que $0 \in$ cl $\mathcal{D}_f(\overline{x})$, i. e., que existe $\{d_k\} \in \mathcal{D}_f(\overline{x})$ tal que $\{d_k\} \to 0$. Se $\mathcal{D}_f(\overline{x}) \neq \emptyset$ então existe $d \in \mathcal{D}_f(\overline{x})$ tal que para todo $t \in (0, \varepsilon)$, com $\varepsilon > 0$, $f(\overline{x} + td) < f(\overline{x})$. Tome $t_k = \frac{1}{k}$, então $d_k = t_k d = \frac{1}{k} d \in \mathcal{D}_f(\overline{x})$. Fazendo $t \to 0+$, temos $\lim d_k = 0$. Logo, $0 \in$ cl $\mathcal{D}_f(\overline{x})$.*

Lema 1.14. ***(Direções de descida)*** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\overline{x} \in \mathbb{R}^n$. Então:*

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

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

Demonstração: Seja $d \in \mathcal{D}_f(\overline{x})$. Para todo t suficientemente pequeno,

$$\begin{aligned} &f(\overline{x}+td) = f(\overline{x}) + t\langle \nabla f(\overline{x}), d\rangle + o(t) \\ \Longrightarrow &f(\overline{x}+td) - f(\overline{x}) = t(\langle \nabla f(\overline{x}), d\rangle + o(t)/t) < 0 \end{aligned}$$

Dividindo ambos os lados da desigualdade por t e passando ao limite quando $t \to 0^+$, obtemos

$$\begin{aligned} 0 > t(\langle \nabla f(\overline{x}), d\rangle + o(t)/t) &\Longrightarrow 0 > \langle \nabla f(\overline{x}), d\rangle + o(t)/t \\ &\Longrightarrow 0 > \lim_{t\to 0^+} \langle \nabla f(\overline{x}), d\rangle + o(t)/t \\ &\Longrightarrow 0 \geq \langle \nabla f(\overline{x}), d\rangle \end{aligned}$$

Isto prova o item (a).

Suponhamos agora que $\langle \nabla f(\overline{x}), d\rangle < 0$. Temos que

$$f(\overline{x}+td) - f(\overline{x}) = t(\langle \nabla f(\overline{x}), d\rangle + o(t)/t).$$

Em particular, para todo $t > 0$ suficientemente pequeno, temos

$$\langle \nabla f(\overline{x}), d\rangle + o(t)/t \leq \frac{1}{2}\langle \nabla f(\overline{x}), d\rangle < 0,$$

o que implica que $f(\overline{x}+td) - f(\overline{x}) < 0$, ou seja, $d \in \mathcal{D}_f(\overline{x})$.

□

Se $\overline{x} \in D$ é um minimizador local do problema (5), então não pode existir uma direção de descida de f no ponto $\overline{x}$ que ao mesmo tempo seja viável em relação ao conjunto D. Temos isso porque

- se $\overline{x} \in D$ é um minimizador local, então $\mathcal{D}_f(\overline{x}) = \emptyset$ (não existem direções de descida). Veja um exemplo na Figura 3.

- se $\overline{x}$ pertence ao contorno de D, então $\mathcal{D}_f(\overline{x}) \neq \emptyset$, porém para qualquer $t \in \mathbb{R}_+$ e $d \in \mathcal{D}_f(\overline{x})$ temos $\overline{x} + td \notin D$, logo essas direções não serão viáveis. Veja um exemplo na Figura 4.

Isto implica, necessariamente, que $\mathcal{D}_f(\overline{x}) \cap \mathcal{V}_D(\overline{x}) = \emptyset$. Pelo Lema (1.14), temos então que

$$\langle \nabla f(\overline{x}), d\rangle \geq 0 \quad \forall d \in \mathcal{V}_D(\overline{x}).$$

No entanto, esta condição(mesmo sendo matematicamente correta) nem sempre fornece informação útil, pois o conjunto de direções viáveis $\mathcal{V}_D(\overline{x})$ pode conter apenas o vetor nulo em muitas situações naturais (e nestes casos a condição dada vale trivialmente, ou seja, não depende da função f e da otimalidade ou não do ponto $\overline{x}$). Por exemplo, para a esfera unitária $D = \{x \in \mathbb{R}^n; \|x\| = 1\}$, temos que $\mathcal{V}_D(\overline{x}) = 0$ para qualquer $\overline{x} \in D$.

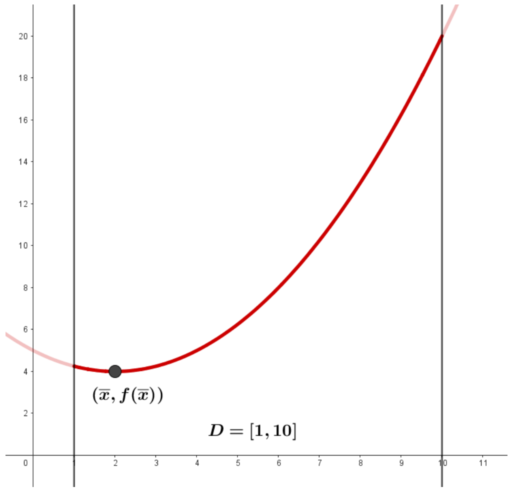


Figura 3: Minimização de $f(x) = \frac{x^2}{4} - x + 5$ em $D = [1, 10]$. Observa-se que o conjunto de direções de descida de f em relação ao ponto $\overline{x}$ é vazio.

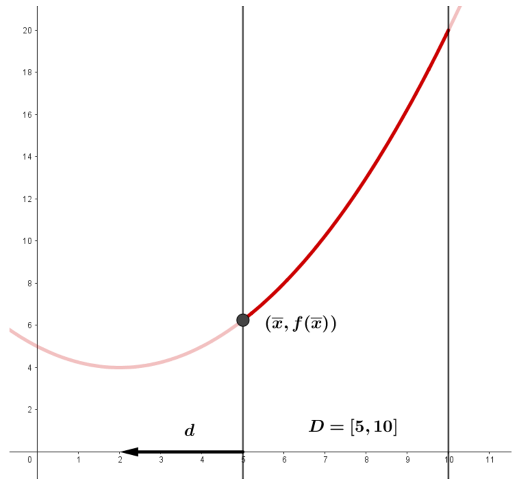


Figura 4: Minimização de $f(x) = \frac{x^2}{4} - x + 5$ em $D = [5, 10]$. Observa-se que d é uma direção de descida em relação ao ponto $\overline{x}$, mas não é uma direção viável neste ponto.

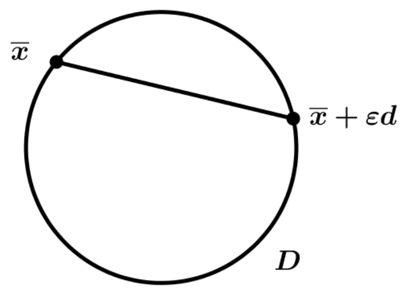


Figura 5: Para a esfera D, quando $\overline{x} \in D$ e $\overline{x} + \varepsilon d \in D$, onde $\varepsilon > 0$ e $d \neq 0$, tem-se que $\overline{x} + td \notin D$ para todo $t \in (0, \varepsilon)$. Logo, $d = 0$ é a única direção viável.

Isto mostra que considerando só as direções viáveis não conseguiremos captar a estrutura local de um conjunto "com curvatura"(principalmente quando o interior do conjunto é vazio). Para obter uma descrição mais informativa, precisamos considerar as direções tangentes, que introduziremos a seguir.

1.2.3 Direções tangentes

Definimos a distância entre um ponto $x \in \mathbb{R}^n$ e o conjunto $D \subset \mathbb{R}^n$ como

$$dist(x, D) = \inf_{y \in D} \|y - x\|.$$

Definição 1.15. *Dizemos que $d \in \mathbb{R}^n$ é uma direção tangente em relação ao conjunto D no ponto $\overline{x} \in D$ quando*

$$dist(\overline{x} + td, D) = o(t), \ t \in \mathbb{R}_+.$$

A propriedade $o(t)$ significa que quando $t \to 0+$, $dist(\overline{x} + td, D)$ é de ordem menor que t, o comprimento do passo na direção d a partir de $\overline{x}$.

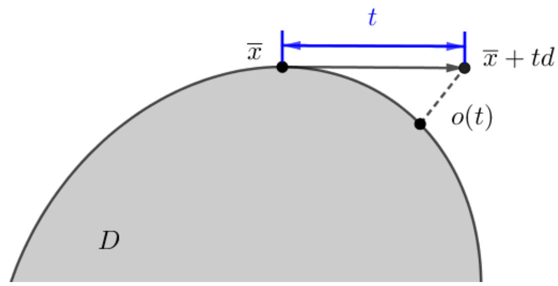


Figura 6: d é uma direção tangente em relação a D em $\overline{x}$. No desenho, $\|d\| = 1$ e, portanto, a distância entre os pontos $\overline{x}$ e $\overline{x}+td$ é igual a t. Quando $t \to 0+$, $dist(\overline{x}+td, D)$ é de ordem menor que t, o comprimento do passo na direção d a partir de $\overline{x}$.

Denotamos por $\mathcal{T}_D(\overline{x})$ o conjunto de todas as direções tangentes ao conjunto D no ponto $\overline{x} \in D$. Assim, temos que

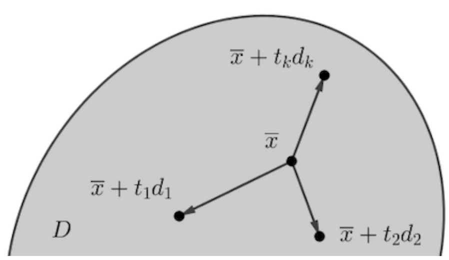
$$\mathcal{V}_D(\overline{x}) \subset \mathcal{T}_D(\overline{x}),$$

isto é, todas as direções viáveis são tangentes, pois para $d \in \mathcal{V}_D(\overline{x})$ temos $dist(\overline{x}+td, D) = 0$ para todo $t > 0$ suficientemente pequeno. Note que a recíproca não vale, pois existem direções que serão tangentes, mas que não serão direções viáveis.

Além disso, o conjunto de todas as direções tangentes em relação ao conjunto D no ponto $\overline{x} \in D$ é um cone, chamado *cone tangente* (em relação ao conjunto D no ponto $\overline{x} \in D$). Como $0 \in \mathcal{T}_D(\overline{x})$, segue-se que o cone tangente sempre é não vazio.

Informalmente falando, o cone tangente pode ser pensado como a união de todas as direções viáveis e direções "quase viáveis" (no sentido de que para um passo da direção quase viável a violação de viabiliade é de ordem menor do que o comprimento do passo).

Se $\overline{x} \in int\ D$, tem-se que $\mathcal{V}_D(\overline{x}) = \mathcal{T}_D(\overline{x}) = \mathbb{R}^n$. Por exemplo:



Observamos que, de forma equivalente, o cone tangente também pode ser definido como

$$\mathcal{T}_D(\overline{x}) = \{d \in \mathbb{R}^n \mid \forall\{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0+, \exists\{d_k\} \subset \mathbb{R}^n, \{d_k\} \to d; \overline{x}+t_kd_k \in D, \forall k \in \mathbb{N}\}$$

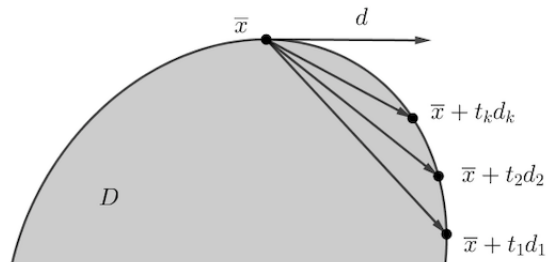


Figura 7: Ilustração da segunda definição de cone tangente: $\overline{x}+t_kd_k \in D$, para todo $k \in \mathbb{N}$, $\{t_k\} \to 0+$ e $\{d_k\} \to d \in \mathcal{T}_D(\overline{x})$.

Para ver isto, notamos que para todo $\{t_k\} \subset \mathbb{R}_+$, $\{t_k\} \to 0+$, temos

$$dist(\overline{x}+t_kd_k, D) \leq \|(\overline{x}+t_kd) - (\overline{x}+t_kd_k)\| = t_k\|d_k - d\| = o(t_k).$$

1.2.4 Cone de Bouligand

Outra noção útil (um pouco mais geral do que a do cone tangente) é a do *cone (tangente) de Bouligand*:

$$\mathcal{B}_D(\overline{x}) = \{d \in \mathbb{R}^n | \exists \{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0+; dist(\overline{x}+t_kd, D) = o(t_k)\},$$

ou, de forma equivalente

$$\mathcal{B}_D(\overline{x}) = \{d \in \mathbb{R}^n | \exists \{t_k\} \subset \mathbb{R}_+, \{t_k\} \to 0+, \exists \{d_k\} \subset \mathbb{R}^n, \{d_k\} \to d, \text{tais que}\, \overline{x}+t_kd_k \in D, \forall k \in \mathbb{N}\}.$$

Comparando as definições acima, temos que, em geral,

$$0 \in \mathcal{V}_D(\overline{x}) \subset \mathcal{T}_D(\overline{x}) \subset \mathcal{B}_D(\overline{x}). \tag{6}$$

Observação 1.16. *No caso em que $\overline{x} \in intD$ temos que $\mathcal{V}_D(\overline{x}) = \mathcal{T}_D(\overline{x}) = \mathcal{B}_D(\overline{x}) = \mathbb{R}^n$.*

Observação 1.17. *Quando $\overline{x}$ é um ponto isolado do conjunto D, tem-se que $\mathcal{V}_D(\overline{x}) = \mathcal{T}_D(\overline{x}) = \mathcal{B}_D(\overline{x}) = \{0\}$.*

A importância do cone de Bouligand deve-se ao fato de que é ele que aparece de maneira mais natural no desenvolvimento de condições necessárias e suficientes de otimalidade de primeira ordem. O resultado a seguir mostra o papel do cone $\mathcal{B}_D(\overline{x})$ nas condições de otimalidade.

1.2.5 Condições de otimalidade

Teorema 1.18. ***(Condição necessária em forma primal)*** *Sejam $D \subset \mathbb{R}^n$ e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\overline{x} \in D$. Se $\overline{x}$ é uma solução local do problema de minimizar f em D, então*

$$\langle \nabla f(\overline{x}), d\rangle \geq 0, \ \forall d \in \mathcal{B}_D(\overline{x}). \tag{7}$$

Demonstração: Para $d = 0 \in \mathcal{B}_D(\overline{x})$, teremos $\langle f'(\overline{x}), d\rangle = \langle f'(\overline{x}), 0\rangle = 0$ e vale a condição (7). Fixemos $d \in \mathcal{B}_D(\overline{x}) - \{0\}$ arbitrário e as sequências associadas $\{t_k\} \subset \mathbb{R}_+ - \{0\}$ e $\{d_k\} \subset \mathbb{R}^n$ tais que $\{t_k\} \to 0+$, $\{d_k\} \to d$ $(k \to \infty)$ e $\overline{x}+t_kd_k \in D$ para todo k.

Como $\overline{x}$ é um minimizador local de (5), temos que $\overline{x} + t_k d_k \in D$ e $\{\overline{x} + t_k d_k\} \to \overline{x}$ com $(k \to \infty)$, para todo k suficientemente grande temos que

$$0 \leq f(\overline{x} + t_k d_k) - f(\overline{x}) = t_k \langle \nabla f(\overline{x}), d_k \rangle + o(t_k \|d_k\|)$$

Vamos tomar $t_k||d_k|| = s_k$ e dividir ambos os lados da desigualdade acima por $t_k = \dfrac{s_k}{\|d_k\|}$:

$$\begin{aligned} 0 &\leq \frac{s_k}{\|d_k\|} \langle \nabla f(\overline{x}), d_k \rangle + o(s_k) \\ \Longrightarrow 0 &\leq \langle \nabla f(\overline{x}), d_k \rangle + \|d_k\| \frac{o(s_k)}{s_k} \end{aligned}$$

Passando o limite quando $k \to \infty$, obtemos

$$0 \leq \langle \nabla f(\overline{x}), d_k \rangle,$$

que é (7), que queríamos provar.

□

Observamos que se $\overline{x}$ é um minimizador local de f em D, então (7) implica a condição

$$\langle \nabla f(\overline{x}), d \rangle \geq 0 \quad \forall d \in \mathcal{T}_D(\overline{x}),$$

já que $\mathcal{T}_D(\overline{x}) \subset \mathcal{B}_D(\overline{x})$. No entanto, esta condição é mais fraca, considerando que nem sempre os dois cones são iguais.

Teorema 1.19. ***(Condições de otimalidade de primeira ordem no caso de conjunto viável convexo)*** *Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e fechado, e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no ponto $\overline{x} \in D$. Se $\overline{x}$ é um minimizador local de f no conjunto D, então*

$$\langle f'(\overline{x}), x - \overline{x} \rangle \geq 0, \ \forall x \in D, \tag{8}$$

ou, equivalentemente,

$$\overline{x} = P_D(\overline{x} - \alpha f'(\overline{x})), \ \forall \alpha \in \mathbb{R}_+. \tag{9}$$

Se f é uma função convexa, $\overline{x} \in D$ e vale (8) (ou (9)), então $\overline{x}$ é um minimizador de f em D. Além disso, neste caso (8) e (9) são equivalentes à seguinte condição:

$$\overline{x} = P_D(\overline{x} - \alpha f'(\overline{x})) \text{ para algum } \alpha > 0. \tag{10}$$

Demonstração: Note que $\{d \in \mathbb{R}^n | d = x - \overline{x}, x \in D\}$ é um conjunto de direções viáveis, pois sempre ocorre $x - \overline{x} \in D$, então $\{d \in \mathbb{R}^n | d = x - \overline{x}, x \in D\} \subset \mathcal{V}_D(\overline{x})$. Por (6) temos também que $\{d \in \mathbb{R}^n | d = x - \overline{x}, x \in D\} \subset \mathcal{V}_D(\overline{x}) \subset \mathcal{B}_D(\overline{x})$. Além disso, pelo Teorema 1.18, $\langle f'(\overline{x}), d \rangle \geq 0$ para todo $d \in \mathcal{B}_D(\overline{x})$. Logo, vale

$$\langle \nabla f(\overline{x}), x - \overline{x} \rangle \geq 0, \ \forall x \in D. \tag{11}$$

Para mostrarmos que (8) é equivalente a (9) utilizaremos também a hipótese de D ser um conjunto fechado. Seja $\overline{x}$ um minimizador local de f no conjunto D. Fixemos $\alpha \in \mathbb{R}_+$

qualquer. Pelo Teorema de Projeção, utilizando a relação (3) com $x = \overline{x} - \alpha\nabla f(\overline{x})$ e $y = \overline{x} \in D$, temos

$$\begin{aligned}
&\langle \overline{x} - \alpha\nabla f(\overline{x}) - P_D(\overline{x} - \alpha\nabla f(\overline{x})), \overline{x} - P_D(\overline{x} - \alpha\nabla(\overline{x}))\rangle \leq 0 \\
\Longrightarrow &\langle \overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x})), \overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\rangle \leq -\langle -\alpha\nabla f(\overline{x}), \overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\rangle
\end{aligned}$$

De onde obtemos

$$\|\overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x})\|^2 \leq \alpha\langle \nabla f(\overline{x}), \overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\rangle. \tag{12}$$

Como $P_D(\overline{x} - \alpha\nabla f(\overline{x})) \in D$, pela desigualdade (11) temos que

$$\langle \nabla f(\overline{x}), P_D(\overline{x} - \alpha\nabla f(\overline{x})) - \overline{x}\rangle \geq 0 \Longrightarrow \langle \nabla f(\overline{x}), \overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\rangle \leq 0.$$

Como $\alpha \geq 0$, de (12) segue que

$$\begin{aligned}
&\|\overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\|^2 \leq 0 \\
\Longrightarrow &\|\overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\|^2 = 0 \\
\Longrightarrow &\|\overline{x} - P_D(\overline{x} - \alpha\nabla f(\overline{x}))\| = 0 \\
\Longrightarrow &P_D(\overline{x} - \alpha\nabla f(\overline{x})) = \overline{x}.
\end{aligned}$$

Se $\overline{x}$ é um minimizador (8) e (10) são satisfeitas por (12) e (9) respectivamente. Suponhamos que valha (8). Usando o item (b) do Teorema 1.9, obtemos que para qualquer $x \in D$,

$$f(x) \geq f(\overline{x}) + \langle \nabla f(\overline{x}), x - \overline{x}\rangle \geq f(\overline{x}),$$

i. e., $\overline{x}$ é um minimizador global.

Suponhamos que (10), i. e., $\overline{x}$ é uma solução do problema

$$min\ \psi(x) \text{ sujeito a } x \in D,$$

onde

$$\psi(x) = \frac{1}{2}\|x - (\overline{x} - \alpha\nabla f(\overline{x}))\|^2, \alpha > 0.$$

Por (11), $\forall x \in D$, tem-se que

$$0 \leq \langle \psi'(\overline{x}), x - \overline{x}\rangle = \langle \overline{x} - (\overline{x} - \alpha\nabla f(\overline{x})), x - \overline{x}\rangle = \alpha\langle \nabla f(\overline{x}), x - \overline{x}\rangle.$$

Como $\alpha > 0$, vale (8). Como já mostramos, isto implica que $\overline{x}$ é minimizador global.

□

2 Métodos do gradiente projetado

Nesta seção, trataremos de um método para problemas de otimização com restrições no caso em que o conjunto viável (o conjunto de todas as soluções possíveis que satisfazem as restrições do problema) tenha descrição abstrata, isto é, é definido por um conceito e não listado explicitamente:

$$min\ f(x) \text{ sujeito a } x \in D, \tag{13}$$

onde $f : \mathbb{R}^n \to \mathbb{R}$ é uma função diferenciável e $D \subset \mathbb{R}^n$. Vamos considerar que o conjunto D possui estrutura simples, isto é, D é pelo menos convexo e fechado.

2.1 Introdução

Com frequência, problemas com restrições simples aparecem como subproblemas em métodos para resoluções de problemas com restrições mais complexas. Notamos que quando o conjunto D é convexo e fechado, para todo ponto $x \in \mathbb{R}^n$ existe uma projeção $P_D(x)$ de x sobre D que é única (veja o Teorema 1.6).
Relembramos também que o ponto $\overline{x} \in D$ é um ponto estacionário do problema (13) quando

$$\langle \nabla f(\overline{x}), x - \overline{x} \rangle \geq 0, \forall x \in D, \tag{14}$$

ou, equivalentemente,

$$P_D(\overline{x} - \alpha \nabla f(\overline{x})) = \overline{x} \tag{15}$$

para algum $\alpha > 0$. Mais ainda, a validade de (15) para algum $\alpha > 0$ implica a validade desta relação para todo $\alpha > 0$.
Métodos do gradiente projetado podem ser pensados como uma combinação natural de Métodos do Gradiente para otimização irrestrita com projeção de iterandos assim obtidos no conjunto viável do problema. Isso resulta no esquema seguinte:

$$x_{k+1} = P_D(x_k - \alpha_k \nabla f(x_k)), k = 0, 1, ... \tag{16}$$

onde $x_0 \in D$, e os parâmetros de comprimentos de passos $\alpha_k > 0$ são calculados utilizando extensões de técnicas de busca linear vistas no capítulo anterior para o caso com restrições. Em particular, o valor de comprimento de passo é calculado para obter decréscimo da função

$$\varphi_k : \mathbb{R}_+ \to \mathbb{R}, \varphi(\alpha) = f(x_k(\alpha)),$$

onde

$$x_k(\alpha) = P_D(x_k - \alpha \nabla f(x_k)) \tag{17}$$

é o arco da projeção (veja Figura 8).

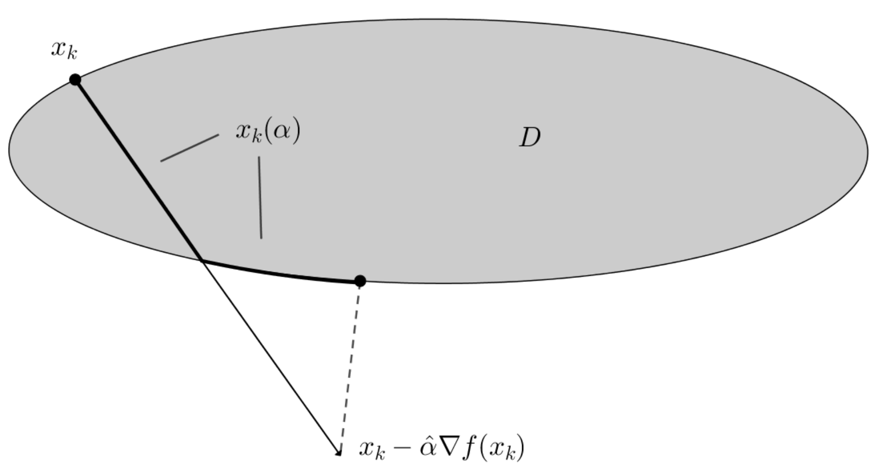


Figura 8: O arco de projeção $x_k(\alpha)$ para o caso da minimização unidimensional, com $\alpha \in [0, \hat{\alpha}]$.

2.2 Buscas lineares para o caso com restrições

Discutiremos brevemente sobre duas buscas lineares aplicadas em problemas com restrições.

2.2.1 Busca da minimização unidimensional

No caso da *minimização unidimensional*, α_k é uma solução do problema

$$\min\ \varphi_k(\alpha) \text{ sujeito a } \alpha \in \mathbb{R}_+, \tag{18}$$

ou do problema

$$\min\ \varphi_k(\alpha) \text{ sujeito a } \alpha \in [0, \hat{\alpha}]\,,$$

onde $\hat{\alpha} > 0$ é um parâmetro.

Observação 2.1. *No entanto, como no caso irrestrito, as regras de minimização unidimensional não são muito atrativas do ponto de vista prático, pois a resolução de um problema unidimensional em cada iteração de um método, mesmo que seja inexata, é relativamente cara. Mais ainda, cabe frisar que o nosso objetivo eventual tem a ver com minimização de f no $\mathbb{R}^n$, e não numa semireta (fixa na iteração k). Por isso, não vale "perder tempo" insistindo em resolver o problema unidimensional (18) com precisão. Na prática, outras regras de busca linear (desenvolvidas a seguir) são mais úteis e mais usadas. Essencialmente, elas buscam assegurar decréscimo suficiente da função f em relação ao valor $f(x_k)$, sem se preocupar com a resolução do problema (18) em si.*

2.2.2 Busca de Armijo

Mais interessantes são as extensões da *regra de Armijo*. Isto pode ser feito, por exemplo, pelo mesmo esquema que foi considerado anteriormente, substituindo a condição de Armijo por

$$f(x_k(\alpha)) \leq f(x_k) + \eta\langle \nabla f(x_k), x_k(\alpha) - x_k\rangle \tag{19}$$

$$f(P_D(x_k - \alpha\nabla f(x_k))) \leq f(x_k) + \eta\langle \nabla f(x_k), P_D(x_k - \alpha\nabla f(x_k)) - x_k\rangle.$$

Relembramos que, para começar o processo, devem ser escolhidos um parâmetro $\eta \in (0, 1)$, valor inicial do comprimento do passo $\hat{\alpha} > 0$ e o parâmetro de redução do passo $\theta \in (0, 1)$. Em particular, α_k é calculado como o maior entre todos os números de forma $\alpha = \hat{\alpha}\theta_s$, com $s \in \mathbb{N}$, satisfazendo a desigualdade (19). Como é fácil ver, no caso quando $D = \mathbb{R}^n$, a desigualdade (19) se reduz à condição de Armijo no caso irrestrito.

2.2.3 Outras buscas lineares

Existem também várias modificações da busca de Armijo, por exemplo, similares à *busca de Goldstein*.
Poderíamos considerar também a *busca de comprimento de passo fixo*, onde $\alpha_k = \overline{\alpha}$ para todo k, com $\overline{\alpha} > 0$ fixo.

2.3 Algoritmos

Apresentamos, a seguir, o algoritmo do método do gradiente projetado e o algoritmo do método do gradiente projetado espectral, baseado em [1].

Algoritmo 1: MÉTODO DO GRADIENTE PROJETADO

1 Tome um ponto inicial $x_0 \in D$ e $k := 0$.
2 Escolha umas das três buscas de cálculo de comprimento do passo a ser utilizada, e os parâmetros associados: no caso da minimização unidimensional, $\hat{\alpha} > 0$ (ou $\hat{\alpha} = +\infty$); no caso da busca de Armijo, $\hat{\alpha} > 0$, $\eta, \theta \in (0, 1)$; e no caso de comprimento do passo fixo, $\overline{\alpha} > 0$.
3 Calcular α_k de acordo com a busca escolhida.
4 Calcular x_{k+1} utilizando a fórmula (16).
5 Tomar $k := k + 1$ e retornar ao passo 1.

Observamos que se para algum k o ponto x_k é estacionário para o problema (13), então o método gera $x_k = x_{k+1} = ...$, para qualquer que seja comprimento do passo x_k (na prática, o método para, é claro). Notemos também que para $D = \mathbb{R}^n$, um método do gradiente projetado se reduz ao método do gradiente (irrestrito) correspondente.

Algoritmo 2: MÉTODO DO GRADIENTE PROJETADO ESPECTRAL

1 Tome um ponto inicial $x_0 \in D$, $\varepsilon > 0$ e faça $k := 0$.
2 Se $\|P_D(x_k - \nabla f(x_k)) - x_k\| \leq \varepsilon$ então pare.
3 Calcule α_k por (19).
4 $x_{k+1} := P_D(x_k - \alpha_k \nabla f(x_k))$.
5 Tomar $k := k + 1$ e retornar ao passo 2.

2.4 Convergência

As propriedades de convergência de métodos do gradiente projetado são muito parecidas com as propriedades de métodos do gradiente.

Na análise a seguir, a seguinte desigualdade tem um papel importante:

$$\langle \nabla f(x_k), x_k(\alpha) \rangle \leq \frac{1}{\alpha} \|x_k(\alpha) - x_k\|^2 \quad \forall \alpha > 0. \tag{20}$$

Para provar esta desigualdade, notamos que para todo $\alpha > 0$,

$$\begin{aligned} \langle \nabla f(x_k), x_k(\alpha) \rangle &= \frac{\alpha}{\alpha} \langle \nabla f(x_k), x_k(\alpha) \rangle \\ &= \frac{1}{\alpha} \langle \alpha \nabla f(x_k), x_k(\alpha) + x_k - x_k \rangle. \end{aligned}$$

Lema 2.2. *Sejam $D \subset \mathbb{R}^n$ um conjunto fechado e convexo, e $f : \mathbb{R}^n \to \mathbb{R}$ uma função diferenciável no $\mathbb{R}^n$, com derivada Lipschitz-contínua no $\mathbb{R}^n$ com módulo $L > 0$. Então para qualquer $x_k \in \mathbb{R}^n$ a desigualdade (20) é satisfeita para todo $\alpha \in (0, \overline{\alpha}]$, onde $\overline{\alpha} = 2(1 - \eta)/L$.*

2.5 Experimentos numéricos

2.5.1 Função Dixon-Price

A função Dixon-Price, de dimensão n, é definida por

$$f(x) = (x_1 - 1) + \sum_{i=2}^{n} i(2x_i^2 - x_{i-1})^2. \tag{21}$$

O mínimo global de (21) é dado por

$$f(\overline{x}) = 0, \text{ em } x_i = 2^{-\frac{2^i - 2}{2^i}}, \text{para } i = 1, 2, ..., n. \tag{22}$$

Nos testes numéricos, foi delimitado o conjunto viável $D = \{x \in \mathbb{R}^{\ltimes} | \langle a, x \rangle = 0\}$, onde $a \in \mathbb{R}^n \ \{0\}$ de forma que $a_1 = -\dfrac{1}{\sqrt{2}}$, $a_2 = 1$ e $a_i = 0$, para $i = 3, 4, ..., n$.

Referências

[1] E. G. Birgin, J. M. Martínez, and M. Raydan. Spectral projected gradient methods: review and perspectives. *Journal of Statistical Software*, 60:1–21, 2014.

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