

Capítulo 1

Introdução à Otimização

Neste capítulo estudaremos alguns fatos básicos sobre os problemas de otimização, que serão necessários ao longo deste livro. Serão apresentados os resultados de existência de soluções e as condições necessárias e suficientes de otimalidade para problemas irrestritos. As condições de otimalidade para problemas com restrições serão desenvolvidas primeiro em forma primal, usando as noções fundamentais de direções tangentes e de cone tangente. Acreditamos que esta abordagem é a mais adequada, porque ela resulta de uma análise direta e natural, sem introdução de técnicas e objetos artificiais ou hipóteses mais restritivas do que necessário. Além disso, a utilização do cone tangente permite um desenvolvimento unificado para várias formas de restrições e várias hipóteses de regularidade de restrições. As condições de otimalidade primais-duais serão obtidas via a descrição analítica do cone tangente e do seu dual para problemas com estrutura específica em Capítulos 2 e 4.

1.1 Definições e alguns fatos básicos

Sejam dados um conjunto $D \subset \mathbf{R}^n$ e uma função $f : D \rightarrow \mathbf{R}$. O problema principal a ser considerado neste livro é o de achar um minimizador de f no conjunto D . Este problema será escrito

como
$$\min f(x) \text{ sujeito a } x \in D. \quad (1.1)$$

O conjunto D será chamado *conjunto viável* do problema, os pontos de D serão chamados *pontos viáveis*, e f será chamada *função objetivo*.

Definição 1.1.1. Dizemos que um ponto $\bar{x} \in D$ é

(a) *minimizador global* de (1.1), se

$$f(\bar{x}) \leq f(x) \quad \forall x \in D; \quad (1.2)$$

(b) *minimizador local* de (1.1), se existe uma vizinhança U de $\bar{x}$ tal que

$$f(\bar{x}) \leq f(x) \quad \forall x \in D \cap U. \quad (1.3)$$

De forma equivalente, $\bar{x} \in D$ é minimizador local se existe $\varepsilon > 0$ tal que $f(\bar{x}) \leq f(x)$ para todo $x \in \{x \in D \mid \|x - \bar{x}\| \leq \varepsilon\}$.

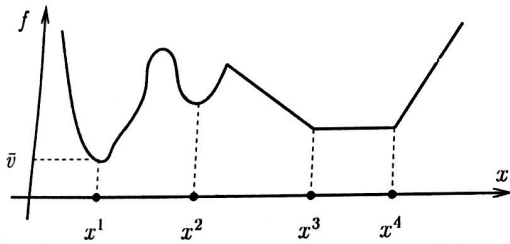


Figura 1.1.1: x^1 é o minimizador global ($\bar{v}$ é o valor ótimo), x^2 é um minimizador local estrito, $[x^3, x^4]$ é um conjunto de minimizadores locais não estritos.

Pela definição, é claro que todo minimizador global também é minimizador local, mas não reciprocamente. Se para todo $x \neq \bar{x}$ a desigualdade (1.2) ou (1.3) é estrita, $\bar{x}$ será chamado *minimizador estrito* (global ou local, respectivamente).

Definição 1.1.2. Dizemos que $\bar{v} \in [-\infty, +\infty)$ definido por

$$\bar{v} = \inf_{x \in D} f(x)$$

é o *valor ótimo* do problema (1.1).

Uma função pode admitir vários minimizadores globais, mas o valor ótimo (global) do problema, naturalmente, sempre é o mesmo. As Definições 1.1.1 e 1.1.2 são ilustradas na Figura 1.1.1.

As seguintes propriedades são fáceis de verificar.

Exercício 1.1.1. Sejam $f: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função qualquer, D_1 e D_2 conjuntos em $\mathbb{R}^n$ tais que $D_2 \subset D_1$.

Mostrar que

$$\inf_{x \in D_1} f(x) \leq \inf_{x \in D_2} f(x).$$

Provar que se $\bar{x}$ é um minimizador de f em D_1 tal que $\bar{x} \in D_2$, então $\bar{x}$ também é minimizador de f em D_2 .

É fácil ver que qualquer problema de maximização

$$\max f(x) \text{ sujeito a } x \in D, \quad (1.4)$$

pode ser transformado em um problema de minimização equivalente:

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

Em particular, as soluções locais e globais de ambos os problemas são as mesmas, com sinais opostos para os valores ótimos, veja a Figura 1.1.2. Por isso, do ponto de vista matemático, não existe nenhuma diferença relevante entre os problemas de minimização e de maximização: todos os resultados obtidos para uma classe de problemas podem ser estendidos para a outra classe sem dificuldade. Por razões de tradição, neste livro consideramos os problemas de minimização.

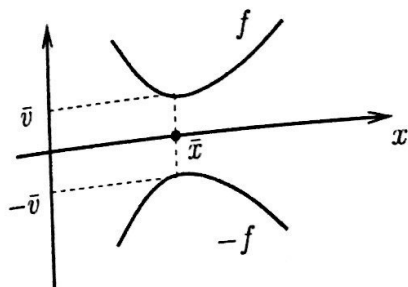


Figura 1.1.2: $\max f(x) = -\min(-f(x))$.

Tipicamente, o conjunto viável de um problema é definido por um sistema de igualdades e/ou desigualdades e/ou uma inclusão, como, por exemplo,

$$D = \left\{ x \in \Omega \mid \begin{array}{l} h_i(x) = 0, i = 1, \dots, l, \\ g_j(x) \leq 0, j = 1, \dots, m \end{array} \right\}, \quad (1.5)$$

ou, em notação mais compacta,

$$D = \{x \in \Omega \mid h(x) = 0, g(x) \leq 0\},$$

onde $\Omega \subset \mathbf{R}^n$, $h: \Omega \rightarrow \mathbf{R}^l$ e $g: \Omega \rightarrow \mathbf{R}^m$ (supomos que a função objetivo f também é definida no conjunto Ω). O conjunto Ω representa o que se chama *restrições diretas* e as restrições de igualdade e desigualdade se chamam *restrições funcionais*. Cabe frisar que a separação das restrições em diretas e funcionais na maioria dos casos é apenas uma questão de simples conveniência. Por exemplo, a restrição direta $x \in \Omega = \mathbf{R}_+^n$ pode ser escrita no formato funcional $x \in \{x \in \mathbf{R}^n \mid x_i \geq 0, i = 1, \dots, n\}$. Tipicamente, as restrições diretas têm uma estrutura “simples”, e.g., $\Omega = \mathbf{R}^n$ ou $\Omega = \mathbf{R}_+^n$ (ou algum outro conjunto de descrição fácil). Neste livro, quase sempre vamos supor que todas as restrições de um problema dado sejam de um tipo só: ou diretas ou funcionais (na maioria dos casos, funcionais). Segundo a discussão acima, isso não introduz perda de generalidade.

Em princípio, as restrições de igualdade sempre podem ser escritas em forma de desigualdades duplicando o número de restrições, já que

$$h_i(x) = 0 \Leftrightarrow h_i(x) \leq 0, -h_i(x) \leq 0.$$

As restrições de desigualdade também podem ser transformadas em igualdades mediante a introdução de variáveis artificiais, chamadas “de folga”:

$$g_i(x) \leq 0 \Leftrightarrow g_i(x) + z_i^2 = 0, z_i \in \mathbf{R}.$$

No entanto, as transformações deste tipo são indesejáveis, tanto do ponto de vista teórico quanto do ponto de vista computacional. Em primeiro lugar, em muitos casos essas transformações levam a condições de otimalidade mais fracas do que as obtidas numa análise direta das restrições originais. Em segundo lugar, o número das restrições e/ou variáveis aumenta, o que torna o problema mais difícil de resolver na prática. Por estas razões, estudaremos o formato geral de restrições mistas de igualdade e desigualdade.

Quando $D = \mathbf{R}^n$, dizemos que o problema (1.1) é *irrestrito*, e quando $D \neq \mathbf{R}^n$ falamos de *otimização com restrições*. O problema irrestrito pode ser considerado um caso particular do (1.1), (1.5), se tomamos $\Omega = \mathbf{R}^n$ e eliminamos as restrições funcionais ($l = m = 0$). Problemas deste tipo serão considerados em § 1.3. O caso de restrição direta $D = \Omega$, $l = 0$, $m = 0$, será analisado em § 1.4.

Quando $\Omega = \mathbf{R}^n$ e $l \neq 0$, $m = 0$, falamos sobre um *problema com restrições de igualdade*. As condições de otimalidade para este caso podem ser obtidas usando as técnicas clássicas da Análise. Estes assuntos serão considerados em Capítulo 2. Finalmente, a teoria para o caso de *restrições mistas* (de igualdade e desigualdade) será desenvolvida em Capítulo 4.

Dizemos que um conjunto é *poliedral* quando ele pode ser representado como o conjunto das soluções de um sistema finito de equações e inequações lineares. Por exemplo:

$$D = \{x \in \mathbf{R}^n \mid Ax = a, Bx \leq b\},$$

onde $A \in \mathbf{R}(l, n)$, $B \in \mathbf{R}(m, n)$, $a \in \mathbf{R}^l$, $b \in \mathbf{R}^m$. Neste contexto, dizemos que $h : \mathbf{R}^n \rightarrow \mathbf{R}^l$, $h(x) = Ax - a$, é uma *função afim*.

Uma função $f : \mathbf{R}^n \rightarrow \mathbf{R}$ definida por

$$f(x) = \langle Qx, x \rangle + \langle q, x \rangle, \quad (1.6)$$

onde $Q \in \mathbf{R}(n, n)$, $q \in \mathbf{R}^n$, é chamada função *quadrática*.

Uma classe especial de problemas de otimização se refere ao caso em que D é um conjunto poliedral. Se além disso f for quadrática, (1.1) se chama problema de *programação quadrática*, e se f for linear ($Q = 0$ em (1.6)), o problema é de *programação linear*.

Quando D é um conjunto convexo e f é uma função convexa (veja Capítulo 3), dizemos que (1.1) é um problema de *programação convexo*. Quando a matriz Q é semidefinida positiva, os problemas quadráticos são convexos (portanto, os problemas lineares sempre são convexos).

Exercício 1.1.2. Fazendo um desenho geométrico, formular uma hipótese sobre a solução global do problema (1.1), onde

(a) $f(x) = -x_1 + x_2$,
 $D = \{x \in \mathbf{R}^2 \mid x_1^2 + x_2^2 \leq 1, x_1 \geq x_2^2, x_1 + x_2 \geq 0\};$

(b) $f(x) = \max\{|x_1 - 2|, |x_2|\}$,
 $D = \{x \in \mathbf{R}^2 \mid 2|x_1| - |x_2| \leq 2\};$

(c) $f(x) = (x_1 - 1)^2 + x_2^2$, $D = \{x \in \mathbf{R}^2 \mid x_1^2 + x_2^2 = 2\sigma x_1 x_2\}$ e $\sigma \in \mathbf{R}$ é um parâmetro.

Exercício 1.1.3. Fazendo um desenho geométrico, responder à seguinte pergunta: para quais valores do parâmetro $\sigma \in \mathbf{R}$ o problema (1.1) com

$$f(x) = x_1 + \sigma x_2, \quad D = \{x \in \mathbf{R}^2 \mid x_1^3 + x_2^3 = 3x_1 x_2\},$$

tem uma solução global e para quais apenas uma solução local?

Exercício 1.1.4. Seja $f : \mathbf{R}^n \rightarrow \mathbf{R}$ uma função contínua. Provar que $\bar{x} \in \mathbf{R}^n$ é um minimizador global de f no $\mathbf{R}^n$ se, e somente se, todo x tal que $f(x) = f(\bar{x})$ é um minimizador local de f .

1.2 Existência de soluções globais

Quando na Definição 1.1.2 temos $\bar{v} = -\infty$, o problema (1.1) não possui solução global (neste caso f é ilimitada inferiormente no conjunto D). Mas mesmo quando $\bar{v}$ é finito, o minimizador global pode não existir. Esse é o caso em que $\bar{v}$ não é atingido em nenhum ponto viável, i.e., quando não existe $x \in D$ tal que $f(x) = \bar{v}$. Por exemplo, sejam $f : \mathbf{R} \rightarrow \mathbf{R}$, $f(x) = e^x$, $D = \mathbf{R}$. Evidentemente, $\bar{v} = \inf_{x \in \mathbf{R}} e^x = 0$. No entanto, não existe $x \in \mathbf{R}$ tal que $e^x = 0$.

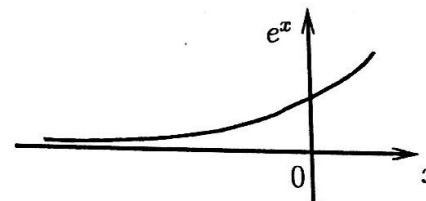


Figura 1.2.1: A função $f(x) = e^x$ não tem minimizador global no $\mathbf{R}$, mesmo que o valor ótimo do problema é finito: para todo $x \in \mathbf{R}$, tem-se que $e^x > 0 = \bar{v} = \inf_{x \in \mathbf{R}} e^x$.

A seguir estudaremos alguns critérios que garantem a existência de solução global.

Teorema 1.2.1. (Teorema de Weierstrass)

Sejam $D \subset \mathbf{R}^n$ um conjunto compacto não vazio e $f : D \rightarrow \mathbf{R}$ uma função contínua.

Então, os problemas (1.1) e (1.4) têm soluções globais.

Demonstração: Pelo observado em § 1.1, 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, $\{v \in \mathbf{R} \mid v = f(x) \text{ para algum } x \in D\}$ é compacto. Em particular, este conjunto é limitado inferiormente, ou seja,

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

Pela definição de ínfimo, para todo $k \in \mathbf{N}$ existe um $x^k \in D$ tal que

$$\bar{v} \leq f(x^k) \leq \bar{v} + 1/k.$$

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

$$\lim_{k \rightarrow \infty} f(x^k) = \bar{v}. \quad (1.7)$$

Como $\{x^k\} \subset D$ e D é compacto, segue-se que $\{x^k\}$ é uma seqüência limitada. Logo, ela possui uma subseqüência $\{x^{k_j}\}$ que converge a um ponto de D :

$$\lim_{j \rightarrow \infty} x^{k_j} = \bar{x} \in D.$$

Pela continuidade de f ,

$$\lim_{j \rightarrow \infty} f(x^{k_j}) = f(\bar{x}).$$

Usando (1.7), temos que $f(\bar{x}) = \bar{v}$, i.e., f assume o valor mínimo em D no ponto $\bar{x} \in D$. Em outras palavras, $\bar{x}$ é um minimizador global do problema (1.1). $\square$

A hipótese de que o conjunto viável seja compacto só pode ser eliminada nos resultados de existência de solução ao custo de fortalecimento das hipóteses sobre a função objetivo. Neste sentido, a noção de conjunto de nível é fundamental.

Definição 1.2.1. O conjunto de nível da função $f : D \rightarrow \mathbf{R}$ associado a $c \in \mathbf{R}$, é o conjunto dado por

$$L_{f,D}(c) = \{x \in D \mid f(x) \leq c\}.$$

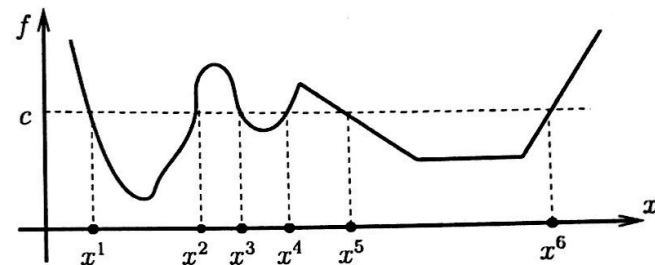


Figura 1.2.2: Conjunto de nível $L_{f,\mathbf{R}}(c) = [x^1, x^2] \cup [x^3, x^4] \cup [x^5, x^6]$.

Corolário 1.2.1. Sejam $D \subset \mathbf{R}^n$ e $f : D \rightarrow \mathbf{R}$ contínua no conjunto D . Suponhamos que existe $c \in \mathbf{R}$ tal que o conjunto de nível $L_{f,D}(c)$ seja não vazio e compacto.

Então o problema (1.1) possui uma solução global.

Demonstração: Pelo Teorema de Weierstrass (Teorema 1.2.1), o problema

$$\min f(x) \text{ sujeito a } x \in L_{f,D}(c)$$

tem uma solução global, digamos $\bar{x}$. Para todo $x \in D \setminus L_{f,D}(c)$, temos que $f(x) > c \geq f(\bar{x})$, o que mostra que $\bar{x}$ é um minimizador global de f não só em $L_{f,D}(c)$, mas também em D . $\square$

A seguir mostraremos um importante exemplo de aplicação do Corolário 1.2.1.

Definição 1.2.2. Dados conjunto $D \subset \mathbf{R}^n$ e ponto $y \in \mathbf{R}^n$, uma projeção (ortogonal) de y sobre D é uma solução global do problema

$$\min \|x - y\| \text{ sujeito a } x \in D.$$

Em outras palavras, a projeção de y sobre D é um dos pontos de D que é mais próximo de y . Dependendo de hipóteses sobre o conjunto D , pode haver mais de um ponto com esta propriedade, um único ponto, ou nenhum. A seguir mostraremos que a projeção sempre existe quando o conjunto é não vazio e fechado.

Corolário 1.2.2. *Seja $D \subset \mathbf{R}^n$ um conjunto fechado não vazio. Então a projeção de y sobre D existe para todo ponto $y \in \mathbf{R}^n$.*

Demonstração: Fixemos $y \in \mathbf{R}^n$ arbitrário. Seja

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

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

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

onde

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

O conjunto de nível $L_{f,D}(c)$ é a interseção do conjunto fechado D com o conjunto compacto $B(y, c)$. Portanto, $L_{f,D}(c)$ é compacto. Além disso, para $c > 0$ suficientemente grande, é óbvio que a bola $B(y, c)$ contém pontos de D . Logo, $L_{f,D}(c) \neq \emptyset$.

A afirmação agora segue pelo Corolário 1.2.1. $\square$

No problema de minimização do Corolário 1.2.2, o conjunto viável do problema poderia ser ilimitado, mas é fechado. A seguir consideremos um exemplo onde o conjunto viável não é nem limitado nem fechado.

Exemplo 1.2.1. Mostraremos que o problema (1.1), onde

$$D = \{x \in \mathbf{R}^2 \mid x_1 + x_2 > 0\},$$

$$f : D \rightarrow \mathbf{R}, \quad f(x) = x_1^2 + x_2 + 1/(x_1 + x_2),$$

possui uma solução global.

Observemos que D é ilimitado e aberto. Portanto, o Teorema 1.2.1 não se aplica neste caso.

Fixemos $c \in \mathbf{R}$ tal que $L(c) = L_{f,D}(c) \neq \emptyset$ (por exemplo, tomando $c = f(x)$ para um $x \in D$ qualquer). Verificaremos que $L(c)$ é compacto (i.e., limitado e fechado), e aplicaremos o Corolário 1.2.1.

Suponhamos primeiro que $L(c)$ seja ilimitado, i.e., que exista $\{x^k\} \subset L(c)$ tal que $\|x^k\| \rightarrow \infty$ ($k \rightarrow \infty$). Temos que $x_1^k + x_2^k > 0$ (porque $x^k \in D$) e $f(x^k) \leq c \forall k$. Se $|x_1^k| \rightarrow \infty$ ($k \rightarrow \infty$), então $c \geq f(x^k) > (x_1^k)^2 + x_2^k > (x_1^k)^2 - x_1^k \rightarrow +\infty$ ($k \rightarrow \infty$), o que é impossível. Se $\{x_1^k\}$ é limitada, então $x_2^k \rightarrow +\infty$ ($k \rightarrow \infty$), e para todo k suficientemente grande, temos que $c \geq f(x^k) > x_2^k \rightarrow +\infty$ ($k \rightarrow \infty$), o que também é impossível. Acabamos de mostrar que $L(c)$ é limitado.

Suponhamos agora que $L(c)$ não seja fechado, i.e., que exista $\{x^k\} \subset L(c)$ tal que $\{x^k\} \rightarrow x$ ($k \rightarrow \infty$), $x \in \mathbf{R}^2 \setminus L(c)$. A função f é contínua em D . Portanto $f(x^k) \leq c$ implica que $f(x) \leq c$. Então como $x \notin L(c)$, deve ser $x \in \text{cl } D \setminus D$, i.e., $x_1 + x_2 = 0$, mas neste caso $c \geq f(x^k) \rightarrow +\infty$ ($k \rightarrow \infty$), o que é impossível. Acabamos de mostrar então que $L(c)$ também é fechado.

Logo $L(c)$ é compacto e a existência de minimizador global segue do Corolário 1.2.1.

A seguir formalizaremos os fenômenos observados na análise do Exemplo 1.2.1.

Definição 1.2.3. Dizemos que uma seqüência $\{x^k\} \subset \mathbf{R}^n$ é *crítica* em relação ao conjunto D , se $\{x^k\} \subset D$ e $\|x^k\| \rightarrow \infty$ ou $\{x^k\} \rightarrow x \in \text{cl } D \setminus D$ ($k \rightarrow \infty$).

Dizemos que a função $f : D \rightarrow \mathbf{R}$ é *coerciva no conjunto D* , quando para toda seqüência $\{x^k\}$ crítica em relação a D , tem-se $\limsup_{k \rightarrow \infty} f(x^k) = +\infty$.

No Exemplo 1.2.1, f é coerciva no conjunto D . Outros exemplos vem dados na Figura 1.2.3. Observamos que quando D é fechado, a Definição 1.2.3 pode ser simplificada, afirmando que $x^k \in D$ e $\|x^k\| \rightarrow \infty$ implicam $\limsup_{k \rightarrow \infty} f(x^k) = +\infty$. Quando D é limitado, a definição pode ser simplificada, afirmando que quando $x^k \in D$ e $\{x^k\} \rightarrow x \in \text{cl } D \setminus D$, tem-se que $\limsup_{k \rightarrow \infty} f(x^k) = +\infty$. Finalmente, quando D é compacto, não há seqüências críticas (e, portanto, qualquer função $f : D \rightarrow \mathbf{R}$ é coerciva em D trivialmente).

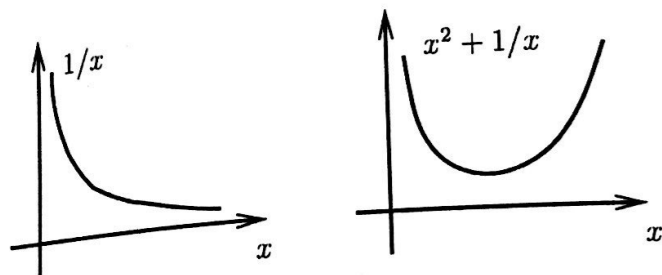


Figura 1.2.3: $f(x) = 1/x$ não é coerciva em $(0, +\infty]$, mas é coerciva em $(0, t]$ para $t > 0$ fixo qualquer. A função $f(x) = x^2 + 1/x$ é coerciva em $(0, +\infty]$.

Corolário 1.2.3. Sejam $D \subset \mathbb{R}^n$ e $f : D \rightarrow \mathbb{R}$ uma função contínua coerciva em D . Então, o problema (1.1) possui uma solução global.

Exercício 1.2.1. Provar o Corolário 1.2.3. (Ajuda: Usar o raciocínio do Exemplo 1.2.1.)

O Teorema de Weierstrass pode ser generalizado ligeiramente em relação às propriedades de continuidade da função objetivo. Como ao longo deste livro admitimos que a função objetivo é pelo menos contínua, as generalizações deste tipo não são muito importantes para nós. Portanto, vamos nos limitar a uma exposição curta e sem provas.

Definição 1.2.4. Dizemos que a função $f : D \rightarrow \mathbb{R}$ é *semicontínua inferiormente* no ponto $x \in D \subset \mathbb{R}^n$, quando para qualquer seqüência $\{x^k\} \subset D$ tal que $\{x^k\} \rightarrow x$ ($k \rightarrow \infty$), tem-se:

$$\liminf_{k \rightarrow \infty} f(x^k) \geq f(x). \quad (1.8)$$

Dizemos que f é *semicontínua superiormente* quando, nas mesmas condições,

$$\limsup_{k \rightarrow \infty} f(x^k) \leq f(x).$$

A função f é *semicontínua inferiormente (superiormente)* no conjunto D , quando ela é semicontínua inferiormente (superiormente) em todos os pontos de D .

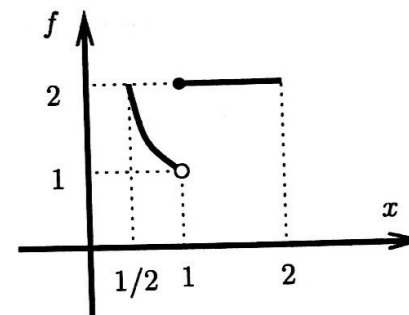


Figura 1.2.4: A função f definida no Exemplo 1.2.2 é semicontínua superiormente, mas não é semicontínua inferiormente em $D = [1/2, 2]$. O conjunto de maximizadores globais de f em D é o conjunto $\{1/2\} \cup [1, 2]$; f não tem minimizador global em D : tem-se que para todo $x \in D$, $f(x) > 1 = \bar{v} = \inf_{z \in D} f(z)$.

Exemplo 1.2.2. Sejam $D = [1/2, 2] \subset \mathbb{R}$ e $f : D \rightarrow \mathbb{R}$ dada por

$$f(x) = \begin{cases} 1/x, & x \in [1/2, 1), \\ 2, & x \in [1, 2], \end{cases}$$

veja Figura 1.2.4. O conjunto D é compacto. A função f não é semicontínua inferiormente em D , pois ela não é semicontínua inferiormente no ponto $x = 1 \in D$. Tem-se que $\bar{v} = 1$, mas não existe $x \in D$ tal que $f(x) = 1$, i.e., não existe minimizador global.

No entanto, é fácil ver que f é semicontínua superiormente e que ela possui maximizadores globais em D .

Exercício 1.2.2. Seja D um conjunto compacto não vazio. Supondo que f seja semicontínua inferiormente em D , provar que

existe um minimizador global para o problema (1.1). Supondo que f seja semicontínua superiormente em D , provar que existe um maximizador global para o problema (1.4).

Exercício 1.2.3. Sejam $D \subset \mathbf{R}^n$ e $f : D \rightarrow \mathbf{R}$ uma função semicontínua inferiormente e coerciva em D . Provar que o problema (1.1) possui uma solução global.

Exercício 1.2.4. Seja f uma função quadrática:

$$f : \mathbf{R}^n \rightarrow \mathbf{R}, \quad f(x) = \langle Qx, x \rangle + \langle q, x \rangle + c,$$

onde $Q \in \mathbf{R}(n, n)$ é uma matriz simétrica, $q \in \mathbf{R}^n$, $c \in \mathbf{R}$.

Provar as seguintes afirmações:

- Se o problema de minimização irrestrita de f tem um minimizador local, então a matriz Q é semidefinida positiva. Mais ainda, todo minimizador local é global.
- Se Q é definida positiva, então f é coerciva em $\mathbf{R}^n$. Em particular, o problema de minimização de f sobre qualquer conjunto $D \subset \mathbf{R}^n$ não vazio e fechado tem um minimizador global. Mostrar que, em geral, isto não é verdadeiro quando D não é fechado.

Exercício 1.2.5. Suponha-se que o problema de minimização irrestrita de uma função contínua no $\mathbf{R}^n$ tenha uma solução global. Determinar se isso implica ou não que f tem um minimizador global em qualquer conjunto fechado em $\mathbf{R}^n$.

Exercício 1.2.6. Suponha-se que as duas funções $f_1 : \mathbf{R}^n \rightarrow \mathbf{R}$ e $f_2 : \mathbf{R}^n \rightarrow \mathbf{R}$ sejam contínuas e que tenham minimizadores globais. Determinar se isso implica ou não que $f_1(\cdot) + f_2(\cdot)$ também tem um minimizador global.

Exercício 1.2.7. Provar que $f : \mathbf{R}^n \rightarrow \mathbf{R}$ é semicontínua inferiormente no conjunto D se, e somente se, os seus conjuntos de nível $L_{f,D}(c)$ são fechados para todo $c \in \mathbf{R}$.

1.3 Condições de otimalidade para problemas sem restrições

Consideremos o problema de minimização irrestrita

$$\min f(x), \quad x \in \mathbf{R}^n, \quad (1.9)$$

onde $f : \mathbf{R}^n \rightarrow \mathbf{R}$. Estudaremos as condições que devem ser satisfeitas quando um $\bar{x} \in \mathbf{R}^n$ dado é minimizador (local) do problema (1.9). Condições deste tipo se chamam *condições necessárias de otimalidade*. Também estudaremos as condições que garantem que um ponto dado é minimizador local do problema. As condições deste último tipo se chamam *condições suficientes de otimalidade*.

Cabe notar que todos os resultados apresentados a seguir são também verdadeiros para um problema com restrições

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

desde que o ponto de interesse $\bar{x} \in D$ esteja no interior do conjunto viável, i.e., exista uma bola $B(\bar{x}, \varepsilon)$ em torno de $\bar{x}$ tal que $B(\bar{x}, \varepsilon) \subset D$. Em particular, este sempre é o caso quando D é um conjunto aberto.

Teorema 1.3.1. (Condição necessária de primeira ordem) Suponhamos que a função $f : \mathbf{R}^n \rightarrow \mathbf{R}$ seja diferenciável no ponto $\bar{x} \in \mathbf{R}^n$.

Se $\bar{x}$ é um minimizador local do problema (1.9), então

$$f'(\bar{x}) = 0. \quad (1.10)$$

Demonstração: Seja $d \in \mathbf{R}^n$ arbitrário, porém fixo. Pela definição de minimizador local, existe $\varepsilon > 0$ tal que

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

Pela diferenciabilidade de f em $\bar{x}$,

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

Logo,

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

Dividindo por $t > 0$, temos que

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

e tomando o limite quando $t \rightarrow 0+$, obtemos

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

Como $d \in \mathbb{R}^n$ é arbitrário, podemos escolher $d = -f'(\bar{x})$, o que resulta na condição $0 \leq \langle f'(\bar{x}), d \rangle = -\|f'(\bar{x})\|^2$. Donde segue que $f'(\bar{x}) = 0$. $\square$

Definição 1.3.1. Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é *estacionário* (ou *crítico*) para o problema (1.9), se vale a condição (1.10).

Portanto, se f é diferenciável, as soluções locais do problema (1.9) devem ser pontos estacionários. Claramente, o mesmo vale para os problemas de maximização.

A seguir, apresentamos as condições de segunda ordem.

Teorema 1.3.2. (Condição necessária de segunda ordem) Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja duas vezes diferenciável no ponto $\bar{x} \in \mathbb{R}^n$.

Se $\bar{x}$ é um minimizador local do problema (1.9), então vale (1.10) e a matriz Hessiana de f no ponto $\bar{x}$ é semidefinida positiva, i.e.,

$$\langle f''(\bar{x})d, d \rangle \geq 0 \quad \forall d \in \mathbb{R}^n. \quad (1.11)$$

Demonstração: A condição (1.10) já foi obtida acima.

Seja $d \in \mathbb{R}^n$ arbitrário, porém fixo. Se $\bar{x}$ é minimizador local do problema (1.9), então para todo $t > 0$ suficientemente pequeno

$$\begin{aligned} 0 &\leq f(\bar{x} + td) - f(\bar{x}) \\ &= \langle f'(\bar{x}), td \rangle + \langle f''(\bar{x})td, td \rangle / 2 + o(t^2) \\ &= t^2 \langle f''(\bar{x})d, d \rangle / 2 + o(t^2), \end{aligned}$$

onde usamos o Teorema 1.3.1 (em particular, a relação (1.10)). Dividindo os dois lados da desigualdade acima por $t^2 > 0$, temos que

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

Passando ao limite quando $t \rightarrow 0+$, obtemos (1.11). $\square$

Teorema 1.3.3. (Condição suficiente de segunda ordem) Suponhamos que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ seja duas vezes diferenciável no ponto $\bar{x} \in \mathbb{R}^n$.

Se $\bar{x}$ é um ponto estacionário (i.e., se vale (1.10)) e se a matriz Hessiana de f em $\bar{x}$ é definida positiva, i.e., se

$$\langle f''(\bar{x})d, d \rangle > 0 \quad \forall d \in \mathbb{R}^n \setminus \{0\}, \quad (1.12)$$

então $\bar{x}$ é minimizador local estrito do problema (1.9).

Demonstração: Suponhamos que $\bar{x}$ não seja minimizador local estrito, i.e., que exista $\{x^k\} \subset \mathbb{R}^n \setminus \{\bar{x}\}$ tal que $\{x^k\} \rightarrow \bar{x}$ ($k \rightarrow \infty$) e $f(x^k) \leq f(\bar{x})$ para todo k . Como a sequência $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\}$ é limitada, ela possui pontos de acumulação. Escolhendo (se for necessário) uma subsequência, podemos admitir que $\{(x^k - \bar{x})/\|x^k - \bar{x}\|\} \rightarrow d \in \mathbb{R}^n \setminus \{0\}$ (de fato, $\|d\| = 1$).

Para todo k , temos

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

onde usamos (1.10). Dividindo os dois lados desta desigualdade por $\|x^k - \bar{x}\|^2 > 0$ e tomando o limite quando $k \rightarrow \infty$, obtemos

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

em contradição com (1.12). Portanto, $\bar{x}$ tem que ser um minimizador local estrito. $\square$

Cabe ressaltar que as condições (1.10) e (1.11) não são suficientes para que um ponto $\bar{x}$ seja minimizador. Por exemplo, a função $f: \mathbf{R} \rightarrow \mathbf{R}$, $f(x) = x^3$, satisfaz essas condições no ponto $\bar{x} = 0$, mas este ponto não é minimizador local do problema (1.9). Por outro lado, as condições (1.10) e (1.12) não são necessárias para que um ponto $\bar{x}$ seja minimizador. A função $f: \mathbf{R} \rightarrow \mathbf{R}$, $f(x) = x^4$, não satisfaz (1.12) em $\bar{x} = 0$, mas este ponto é minimizador do problema (1.9).

Para problemas de maximização, resultados análogos dos Teoremas 1.3.2 e 1.3.3 são válidas se trocamos os sinais nas desigualdades (1.11) e (1.12).

Exemplo 1.3.1. Encontremos os minimizadores e maximizadores locais irrestritos da função

$$f: \mathbf{R}^2 \rightarrow \mathbf{R}, \quad f(x) = x_1^3 + x_2^3 - 3x_1x_2.$$

É fácil ver que f não é limitada nem inferiormente nem superiormente (tomando $x_1^k = k$ e $x_2^k = 0$, temos que $f(x^k) \rightarrow +\infty$ quando $k \rightarrow \infty$; tomando $x_1^k = -k$ e $x_2^k = 0$, temos que $f(x^k) \rightarrow -\infty$ quando $k \rightarrow \infty$). Então não há minimizadores ou maximizadores globais.

Para definir os pontos estacionários, resolvemos o seguinte sistema de equações:

$$f'_{x_1}(x) = 3x_1^2 - 3x_2 = 0, \quad f'_{x_2}(x) = 3x_2^2 - 3x_1 = 0.$$

As soluções são $x^1 = (0, 0)$ e $x^2 = (1, 1)$.

Para encontrar os minimizadores e maximizadores locais, usamos as condições de otimalidade de segunda ordem. Temos que

$$f''(x^1) = \begin{pmatrix} 0 & -3 \\ -3 & 0 \end{pmatrix}, \quad f''(x^2) = \begin{pmatrix} 6 & -3 \\ -3 & 6 \end{pmatrix}.$$

A matriz $f''(x^1)$ não é nem semidefinida positiva nem semidefinida negativa. Pelo Teorema 1.3.2, x^1 não é nem minimizador nem maximizador. Como a matriz $f''(x^2)$ é definida positiva, pelos Teoremas 1.3.1 e 1.3.3, x^2 é o único minimizador local (estrito). Não há maximizadores.

Exemplo 1.3.2. Encontremos as soluções globais do problema

$$\min f(x) = x_1 + \frac{x_2^2}{4x_1} + \frac{x_3^2}{x_2} + \frac{2}{x_3} \quad \text{sujeito a}$$

$$x \in D = \{x \in \mathbf{R}^3 \mid x_1 > 0, x_2 > 0, x_3 > 0\}.$$

O conjunto D é aberto. Logo, todo ponto viável pertence ao interior de D . Resolvendo (1.10), encontramos que o único ponto estacionário é $\bar{x} = (1/2, 1, 1)$. Verificamos que ele é minimizador global (por exemplo, mostrando que f é coerciva em D , veja Definição 1.2.3 e Exercício 1.2.1).

Exercício 1.3.1. Encontrar todos os valores do parâmetro σ , $\sigma \in \mathbf{R}$, para os quais o problema de minimização irrestrita da função

$$f: \mathbf{R}^2 \rightarrow \mathbf{R}, \quad f(x) = \sigma x_1^2 + 4x_1x_2 + 4x_2^2 + x_1 + x_2,$$

tem uma solução.

Exercício 1.3.2. Mostrar que a função

$$f: \mathbf{R}^2 \rightarrow \mathbf{R}, \quad f(x) = (x_2 - x_1^2)^2 + x_1^5$$

tem um único ponto estacionário que não é um maximizador nem um minimizador de f no $\mathbf{R}^n$.

Exercício 1.3.3. Sejam $n \geq 2$ e $f: \mathbf{R}^n \rightarrow \mathbf{R}$ definida por

$$f(x) = (1 + x_n)^3 \sum_{i=1}^{n-1} x_i^2 + x_n^2.$$

Mostrar que $\bar{x} = 0$ é o único ponto estacionário de f no $\mathbf{R}^n$ e que $\bar{x}$ é um minimizador local estrito de f , mas não é minimizador global de f no $\mathbf{R}^n$.

Exercício 1.3.4. Determinar se pode ou não existir uma função $f: \mathbf{R} \rightarrow \mathbf{R}$ que tem um único ponto estacionário e este ponto é minimizador local de f , mas não minimizador global de f .

Exercício 1.3.5. Provar que, para qualquer $\sigma > 1$, o sistema de equações

$$\sigma \cos x_1 \sin x_2 + x_1 e^{x_1^2 + x_2^2} = 0, \quad \sigma \sin x_1 \cos x_2 + x_2 e^{x_1^2 + x_2^2} = 0$$

possui uma solução $x \in \mathbb{R}^2 \setminus \{0\}$.

1.4 Condições de otimalidade em forma primal para problemas com restrições. O cone tangente.

Consideramos agora um problema em formato geral

$$\min f(x) \text{ sujeito a } x \in D, \quad (1.13)$$

onde $D \subset \mathbb{R}^n$ é um dado conjunto cuja estrutura por enquanto não é especificada, e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é a função objetivo.

Para 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 foi feito para o caso irrestrito), mas também a estrutura do conjunto viável nessa vizinhança. As noções de direções tangentes e de cone tangente são fundamentais neste contexto e são muito importantes para um entendimento correto das condições de otimalidade.

Definição 1.4.1. Um conjunto $K \subset \mathbb{R}^n$ chama-se *cone* quando

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

Pela definição, se K é um cone não vazio, necessariamente $0 \in K$. Alguns exemplos de cone são: o espaço $\mathbb{R}^n$, qualquer subespaço de $\mathbb{R}^n$, o ortante não negativo $\mathbb{R}_+^n$. Informalmente, um cone é um conjunto de direções, veja Figura 1.4.1.

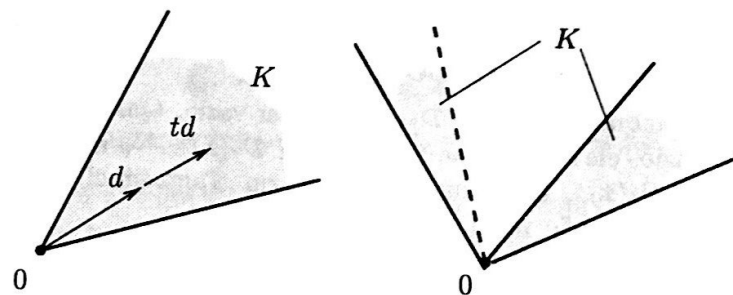


Figura 1.4.1: Exemplos de cones.

Definição 1.4.2. Dizemos que $d \in \mathbb{R}^n$ é uma *direção viável* em relação ao conjunto D no ponto $\bar{x} \in D$, quando existe $\varepsilon > 0$ tal que

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

Denotamos por $\mathcal{V}_D(\bar{x})$ o conjunto de todas as direções viáveis em relação ao conjunto D no ponto $\bar{x} \in D$.

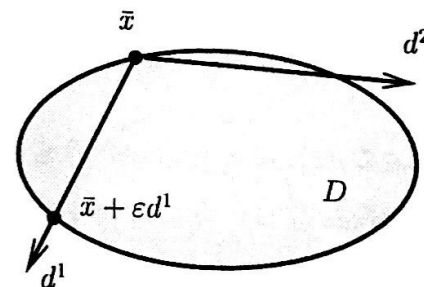


Figura 1.4.2: As direções d^1 e d^2 são viáveis em relação ao conjunto D no ponto $\bar{x} \in D$. Por exemplo, tem-se que $\bar{x} + td^1 \in D$ para todo $t \in [0, \varepsilon]$.

A Figura 1.4.2 ilustra a definição de direções viáveis. É fácil ver que $\mathcal{V}_D(\bar{x})$ é um cone não vazio (pelo menos, $0 \in \mathcal{V}_D(\bar{x})$).

Definição 1.4.3. Dizemos que $d \in \mathbb{R}^n$ é uma *direção de descida* de $f : \mathbb{R}^n \rightarrow \mathbb{R}$ no ponto $\bar{x} \in \mathbb{R}^n$, se existe $\varepsilon > 0$ tal que

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

Denotamos por $\mathcal{D}_f(\bar{x})$ o conjunto de todas as direções de descida da função f no ponto $\bar{x}$.

É claro que o conjunto $\mathcal{D}_f(\bar{x})$ pode ser vazio. Quando $\mathcal{D}_f(\bar{x})$ é não vazio, ele não é cone (porque $0 \notin \mathcal{D}_f(\bar{x})$). No entanto, o conjunto $\mathcal{D}_f(\bar{x}) \cup \{0\}$ é um cone não vazio. Também cl $\mathcal{D}_f(\bar{x})$ é um cone que é não vazio se $\mathcal{D}_f(\bar{x}) \neq \emptyset$.

A Figura 1.4.3 ilustra o Lema 1.4.1 a seguir.

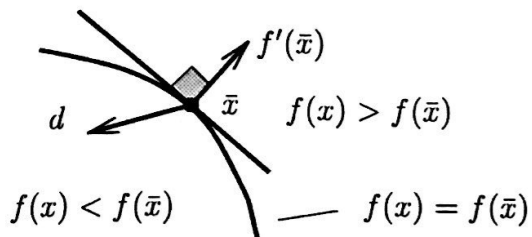


Figura 1.4.3: Como $\langle f'(\bar{x}), d \rangle < 0$, segue-se que d é uma direção de descida da função f no ponto $\bar{x}$, i.e., $f(\bar{x} + td) < f(\bar{x})$ para todo $t > 0$ suficientemente pequeno.

Lema 1.4.1. (Direções de descida)

Seja $f: \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in \mathbb{R}^n$.

Então:

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

Demonstração: Seja $d \in \mathcal{D}_f(\bar{x})$. Para todo $t > 0$ suficientemente pequeno,

$$0 > f(\bar{x} + td) - f(\bar{x}) = t(\langle f'(\bar{x}), d \rangle + o(t)/t).$$

Dividindo os dois lados da desigualdade acima por $t > 0$ e passando ao limite quando $t \rightarrow 0+$, obtemos $0 \geq \langle f'(\bar{x}), d \rangle$, o que prova o item (a).

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

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

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

$$\langle f'(\bar{x}), d \rangle + o(t)/t \leq \langle f'(\bar{x}), d \rangle/2 < 0,$$

o que implica que $f(\bar{x} + td) - f(\bar{x}) < 0$, i.e., $d \in \mathcal{D}_f(\bar{x})$. $\square$

Se $\bar{x} \in D$ é um minimizador local do problema (1.13), é óbvio que não pode existir uma direção de descida de f no ponto $\bar{x}$ que ao mesmo tempo seja viável em relação ao conjunto D , i.e., necessariamente temos

$$\mathcal{D}_f(\bar{x}) \cap \mathcal{V}_D(\bar{x}) = \emptyset.$$

Pelo Lema 1.4.1, temos então que

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

No entanto, esta condição nem sempre fornece informação útil, porque o conjunto de direções viáveis $\mathcal{V}_D(\bar{x})$ pode conter só o vetor nulo em muitas situações naturais (e neste caso a condição acima vale trivialmente, i.e., independente de f). Por exemplo, se $D = \{x \in \mathbb{R}^n \mid \|x\| = 1\}$ é a esfera unitária, temos que $\mathcal{V}_D(\bar{x}) = \{0\}$ para qualquer $\bar{x} \in D$ (veja Figura 1.4.4). Isto

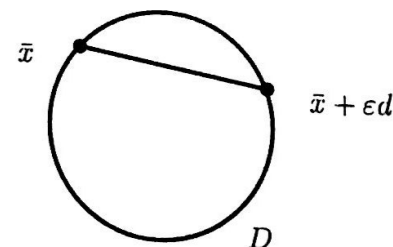


Figura 1.4.4: Para esfera D , quando $\bar{x} \in D$ e $\bar{x} + \epsilon d \in D$, onde $\epsilon > 0$ e $d \neq 0$, tem-se que $\bar{x} + td \notin D$ para todo $t \in (0, \epsilon)$. Logo, $d = 0$ é a única direção viável.

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 introduzimos a seguir.

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

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

Definição 1.4.4. Dizemos que $d \in \mathbb{R}^n$ é uma *direção tangente* em relação ao conjunto D no ponto $\bar{x} \in D$ quando

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

Denotamos por $\mathcal{T}_D(\bar{x})$ o conjunto de todas as direções tangentes ao conjunto D no ponto $\bar{x}$.

A Figura 1.4.5 ilustra a Definição 1.4.4.

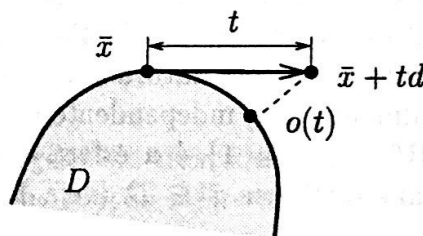


Figura 1.4.5: Ilustração da Definição 1.4.4: d é uma direção tangente em relação a D em $\bar{x}$. No desenho, $\|d\| = 1$ e, portanto, a distância entre os pontos $\bar{x}$ e $\bar{x} + td$ é igual a t . Quando $t \rightarrow 0+$, $\text{dist}(\bar{x} + td, D)$ é de ordem menor que t , o comprimento do passo na direção d a partir de $\bar{x}$.

Claramente,

$$\mathcal{V}_D(\bar{x}) \subset \mathcal{T}_D(\bar{x}), \quad (1.14)$$

i.e., todas as direções viáveis são tangentes (para $d \in \mathcal{V}_D(\bar{x})$ tem-se que $\text{dist}(\bar{x} + td, D) = 0$ para todo $t > 0$ suficientemente pequeno), mas não reciprocamente. Também é claro que o conjunto de todas as direções tangentes em relação ao conjunto D no ponto $\bar{x} \in D$ é um cone, chamado o *cone tangente* (em relação

ao conjunto D no ponto $\bar{x} \in D$). Como $0 \in \mathcal{T}_D(\bar{x})$, 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 na direção quase-viável, a violação de viabilidade é de ordem menor do que o comprimento do passo).

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

$$\mathcal{T}_D(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \forall \{t_k\} \subset \mathbb{R}_+, \{t_k\} \rightarrow 0+, \\ \exists \{d^k\} \subset \mathbb{R}^n, \{d^k\} \rightarrow d, \text{ tal que } \\ \bar{x} + t_k d^k \in D \text{ para todo } k \in \mathbb{N} \end{array} \right\}.$$

Esta última definição é ilustrada na Figura 1.4.6.

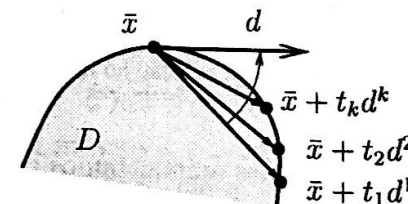


Figura 1.4.6: $\bar{x} + t_k d^k \in D$ para todo k , $\{t_k\} \rightarrow 0+$, e $\{d^k\} \rightarrow d \in \mathcal{T}_D(\bar{x})$.

Outra noção útil (um pouco mais geral do que a do cone tangente) é a do *cone (tangente) de Bouligand*¹:

$$\mathcal{B}_D(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \exists \{t_k\} \subset \mathbb{R}_+, \{t_k\} \rightarrow 0+, \text{ tal que } \\ \text{dist}(\bar{x} + t_k d, D) = o(t_k) \end{array} \right\},$$

ou, de forma equivalente,

$$\mathcal{B}_D(\bar{x}) = \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \exists \{t_k\} \subset \mathbb{R}_+, \{t_k\} \rightarrow 0+, \\ \exists \{d^k\} \subset \mathbb{R}^n, \{d^k\} \rightarrow d, \text{ tais que } \\ \bar{x} + t_k d^k \in D \text{ para todo } k \in \mathbb{N} \end{array} \right\}.$$

¹Em inglês, também se chama *contingent cone*

A Figura 1.4.7 contém desenho de um cone tangente típico quando o ponto $\bar{x}$ está na fronteira de conjunto, o que é a situação mais interessante. Observamos que o caso em que $\bar{x}$ está no interior do conjunto é o caso fácil de analisar: é óbvio que

$$\bar{x} \in \text{int } D \Rightarrow \mathcal{T}_D(\bar{x}) = \mathcal{B}_D(\bar{x}) = \mathcal{V}_D(\bar{x}) = \mathbb{R}^n.$$

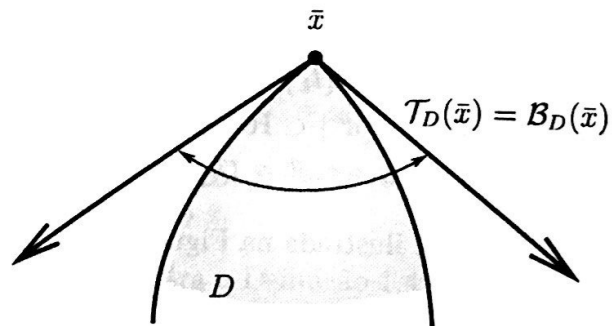


Figura 1.4.7: O cone tangente em relação ao conjunto D no ponto $\bar{x} \in D$. No caso do desenho, tem-se que $\mathcal{T}_D(\bar{x}) = \mathcal{B}_D(\bar{x})$.

Comparando as definições acima, temos que, em geral,

$$\mathcal{T}_D(\bar{x}) \subset \mathcal{B}_D(\bar{x}). \quad (1.15)$$

Como vamos ver mais adiante, quando o conjunto D é definido por restrições funcionais e é possível dar uma descrição construtiva (algébrica) das direções tangentes, quase sempre os cones tangente e de Bouligand coincidem (este também é o caso na Figura 1.4.7). Mas a igualdade em (1.15) não é válida em geral, veja o Exemplo 1.4.1 mais adiante. 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, como veremos a seguir. No entanto, o cone $\mathcal{T}_D(\bar{x})$ é mais fácil de se visualizar.

Teorema 1.4.1. (Condição necessária em forma primal)
Sejam $D \subset \mathbb{R}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in D$.

Se $\bar{x}$ é uma solução local do problema (1.13), então

$$\langle f'(\bar{x}), d \rangle \geq 0 \quad \forall d \in \mathcal{B}_D(\bar{x}). \quad (1.16)$$

Demonstração: Para $d = 0 \in \mathcal{B}_D(\bar{x})$, a condição (1.16) vale trivialmente. Fixemos $d \in \mathcal{B}_D(\bar{x}) \setminus \{0\}$ arbitrário e as seqüências associadas $\{t_k\} \subset \mathbb{R}_+ \setminus \{0\}$ e $\{d^k\} \subset \mathbb{R}^n$ tais que $\{t_k\} \rightarrow 0+$, $\{d^k\} \rightarrow d$ ($k \rightarrow \infty$) e $\bar{x} + t_k d^k \in D$ para todo k .

Como $\bar{x}$ é um minimizador local de (1.13) e $\{\bar{x} + t_k d^k\} \rightarrow \bar{x}$ ($k \rightarrow \infty$), para todo k suficientemente grande temos que

$$\begin{aligned} 0 &\leq f(\bar{x} + t_k d^k) - f(\bar{x}) \\ &= t_k \langle f'(\bar{x}), d^k \rangle + o(t_k \|d^k\|) \\ &= t_k \langle f'(\bar{x}), d \rangle + o(t_k). \end{aligned}$$

Dividindo os dois lados da desigualdade acima por $t_k > 0$ e passando ao limite quando $k \rightarrow \infty$, obtemos (1.16). $\square$

Definição 1.4.5. Dizemos que um ponto $\bar{x} \in \mathbb{R}^n$ é *ponto estacionário* do problema (1.13) se $\bar{x} \in D$ e vale (1.16).

O Teorema 1.4.1 mostra que para um ponto ser uma solução local do problema (1.13), ele tem que ser estacionário no sentido da condição (1.16).

Como veremos a seguir, a condição necessária de primeira ordem em problemas irrestritos (Teorema 1.3.1) é um caso particular do Teorema 1.4.1.

Corolário 1.4.1. Seja $\bar{x} \in \text{int } D$ uma solução local do problema (1.13) tal que $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é diferenciável em $\bar{x}$. Então $f'(\bar{x}) = 0$.

Demonstração: Como já mencionamos (e como pode ser verificado com facilidade), $\bar{x} \in \text{int } D$ implica $\mathbb{R}^n = \mathcal{T}_D(\bar{x}) \subset \mathcal{B}_D(\bar{x})$, i.e., $\mathcal{B}_D(\bar{x}) = \mathbb{R}^n$. Neste caso, (1.16) só pode ser verdadeiro quando $f'(\bar{x}) = 0$. $\square$

A condição de otimalidade no Teorema 1.4.1 se chama *condição primal*, porque ela é baseada nas variáveis "originais" (também chamadas primais) do problema (1.13), i.e., as variáveis $x \in \mathbb{R}^n$.

Definição 1.4.6. O cone dual² de um cone $K \subset \mathbb{R}^n$ é definido por

$$K^* = \{y \in \mathbb{R}^n \mid \langle y, d \rangle \leq 0 \quad \forall d \in K\}.$$

Usando a noção de cone dual, a condição de otimalidade (1.16) é equivalente a

$$-f'(\bar{x}) \in (\mathcal{B}_D(\bar{x}))^*. \quad (1.17)$$

Uma ilustração de cone dual (do cone tangente) e das condições de otimalidade (1.16) e (1.17) é dada na Figura 1.4.8.

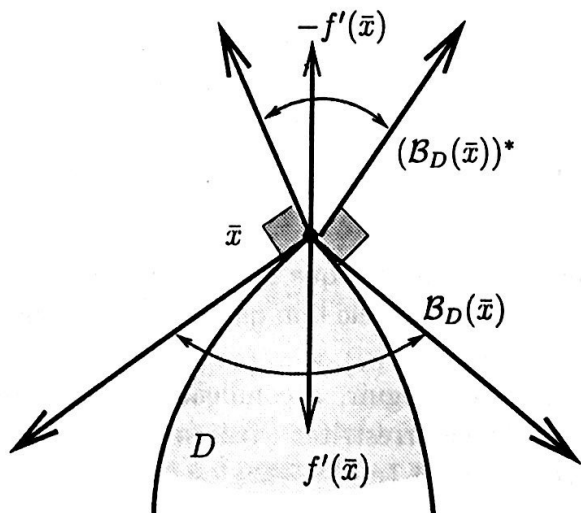


Figura 1.4.8: O cone dual de cone $\mathcal{B}_D(\bar{x})$ e as condições de otimalidade (1.16) e (1.17). Tem-se que $\langle f'(\bar{x}), d \rangle \geq 0$ para todo $d \in \mathcal{B}_D(\bar{x})$ ou, equivalentemente, $-f'(\bar{x}) \in (\mathcal{B}_D(\bar{x}))^*$.

Como veremos mais adiante, o cálculo do cone $(\mathcal{B}_D(\bar{x}))^*$ para conjuntos D com uma estrutura mais concreta faz uso de variáveis auxiliares chamadas *variáveis duais*. Por isso, a condição de otimalidade (1.17), bem como suas consequências, se chamam *condições primais-duais*.

²As vezes, este cone também é chamado cone polar

Teorema 1.4.2. (Condição suficiente em forma primal)

Sejam $D \subset \mathbb{R}^n$ e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in D$. Se

$$\langle f'(\bar{x}), d \rangle > 0 \quad \forall d \in \mathcal{B}_D(\bar{x}) \setminus \{0\}, \quad (1.18)$$

então $\bar{x}$ é um minimizador local estrito do problema (1.13).

Demonstração: Suponhamos que $\bar{x}$ não seja minimizador local estrito, i.e., que exista uma sequência $\{x^k\} \subset D \setminus \{\bar{x}\}$ tal que $\{x^k\} \rightarrow \bar{x}$ ($k \rightarrow \infty$) e $f(x^k) \leq f(\bar{x})$ para todo k . Definimos $t_k = \|x^k - \bar{x}\| > 0$ e $d^k = (x^k - \bar{x})/\|x^k - \bar{x}\|$. Temos que $\bar{x} + t_k d^k = x^k \in D$ para todo k . Como $\{d^k\}$ é limitada, podemos supor (tomando uma subsequência, se for necessário) que ela converge a $d \in \mathbb{R}^n$, $\|d\| = 1$. Pela definição do cone de Bouligand $\mathcal{B}_D(\bar{x})$, temos que $d \in \mathcal{B}_D(\bar{x}) \setminus \{0\}$.

Para todo k ,

$$\begin{aligned} 0 &\geq f(x^k) - f(\bar{x}) \\ &= \langle f'(\bar{x}), x^k - \bar{x} \rangle + o(\|x^k - \bar{x}\|) \\ &= t_k \langle f'(\bar{x}), d^k \rangle + o(t_k). \end{aligned}$$

Dividindo os dois lados por $t_k > 0$ e passando ao limite quando $k \rightarrow \infty$, obtemos que

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

em contradição com (1.18). Portanto, $\bar{x}$ deve ser um minimizador local estrito. $\square$

Em geral, o cone $\mathcal{T}_D(\bar{x})$ não pode ser usado no lugar do cone $\mathcal{B}_D(\bar{x})$ no Teorema 1.4.2, como mostra o Exemplo 1.4.2 um pouco mais adiante.

Observamos ainda que a condição (1.18) só pode ser satisfeita quando o cone $\mathcal{B}_D(\bar{x})$ não contém subespaços (fora do subespaço trivial $\{0\}$). Este fato indica as limitações de aplicação do Teorema 1.4.2, o que vai ficar bem claro mais adiante.

A questão do cálculo dos cones $\mathcal{B}_D(\bar{x})$ e $(\mathcal{B}_D(\bar{x}))^*$ para várias classes de problemas é fundamental para se obter condições de otimalidade úteis tanto do ponto de vista computacional quanto do ponto de vista teórico. Por exemplo, as descrições destes cones serão cruciais para o futuro desenvolvimento de condições de otimalidade para problemas com restrições funcionais em Capítulos 2 e 4.

A seguir, mostramos que os cones $\mathcal{T}_D(\bar{x})$ e $\mathcal{B}_D(\bar{x})$ sempre são fechados, para qualquer conjunto D .

Proposição 1.4.1. (Relações entre as direções viáveis, tangentes e de Bouligand)

Sejam $D \subset \mathbb{R}^n$ um conjunto qualquer e $\bar{x} \in D$.

Então os cones $\mathcal{T}_D(\bar{x})$ e $\mathcal{B}_D(\bar{x})$ são fechados e tem-se que

$$\{0\} \subset \text{cl } \mathcal{V}_D(\bar{x}) \subset \mathcal{T}_D(\bar{x}) \subset \mathcal{B}_D(\bar{x}).$$

Demonstração: Primeiro mostramos que $\mathcal{B}_D(\bar{x})$ é fechado. Seja $\{d^j\} \subset \mathcal{B}_D(\bar{x})$, $\{d^j\} \rightarrow d$ ($j \rightarrow \infty$). Temos que verificar que $d \in \mathcal{B}_D(\bar{x})$.

Pela definição de $d^j \in \mathcal{B}_D(\bar{x})$, para todo j existem seqüências $\{t_{j,k}\} \rightarrow 0+$ e $\{d^{j,k}\} \rightarrow d^j$ ($k \rightarrow \infty$), tais que $\bar{x} + t_{j,k}d^{j,k} \in D$ para todo k . Para todo j , escolhemos um índice k_j tal que $k_j \rightarrow \infty$ quando $j \rightarrow \infty$ e

$$0 = \lim_{j \rightarrow \infty} t_{j,k_j}, \quad 0 = \lim_{j \rightarrow \infty} \|d^{j,k_j} - d^j\|$$

(a possibilidade de fazer tal escolha é óbvia). Pela desigualdade triangular,

$$\lim_{j \rightarrow \infty} \|d^{j,k_j} - d\| \leq \lim_{j \rightarrow \infty} \|d^j - d\| + \lim_{j \rightarrow \infty} \|d^{j,k_j} - d^j\| = 0.$$

Como temos também que $\bar{x} + t_{j,k_j}d^{j,k_j} \in D$ para todo j , as seqüências $\{t_{j,k_j}\} \rightarrow 0+$ e $\{d^{j,k_j}\} \rightarrow d$ ($j \rightarrow \infty$) são aquelas que verificam que $d \in \mathcal{B}_D(\bar{x})$.

Mostramos agora que $\mathcal{T}_D(\bar{x})$ é fechado. Seja $\{d^k\} \subset \mathcal{T}_D(\bar{x})$, $\{d^k\} \rightarrow d$ ($k \rightarrow \infty$). Suponhamos que $d \notin \mathcal{T}_D(\bar{x})$, i.e., que

$$\limsup_{t \rightarrow 0+} \frac{\text{dist}(\bar{x} + td, D)}{t} > 0.$$

Neste caso, existem uma seqüência $\{t_j\} \rightarrow 0+$ ($j \rightarrow \infty$) e uma constante $\gamma > 0$ tais que

$$\text{dist}(\bar{x} + t_j d, D) \geq \gamma t_j \quad \forall j.$$

Portanto,

$$\begin{aligned} \forall x \in D, \|\bar{x} + t_j d - x\| &\geq \inf_{x \in D} \|\bar{x} + t_j d - x\| \\ &= \text{dist}(\bar{x} + t_j d, D) \geq \gamma t_j \quad \forall j. \end{aligned}$$

Fixamos k tal que $\|d^k - d\| \leq \gamma/2$. Obtemos que

$$\begin{aligned} \forall x \in D, \|\bar{x} + t_j d^k - x\| &\geq \|\bar{x} + t_j d - x\| - t_j \|d^k - d\| \\ &\geq \gamma t_j - \gamma t_j/2 = \gamma t_j/2. \end{aligned}$$

Logo,

$$\text{dist}(\bar{x} + t_j d^k, D) = \inf_{x \in D} \|\bar{x} + t_j d^k - x\| \geq \gamma t_j/2 \quad \forall j,$$

e portanto,

$$\limsup_{t \rightarrow 0+} \frac{\text{dist}(\bar{x} + td^k, D)}{t} \geq \gamma/2 > 0,$$

o que contradiz que $d^k \in \mathcal{T}_D(\bar{x})$. Temos então que $d \in \mathcal{T}_D(\bar{x})$, i.e., $\mathcal{T}_D(\bar{x})$ é fechado.

Pelo observado acima (veja (1.14)), $\mathcal{V}_D(\bar{x}) \subset \mathcal{T}_D(\bar{x})$. Como $\mathcal{T}_D(\bar{x})$ é fechado, segue-se que

$$\text{cl } \mathcal{V}_D(\bar{x}) \subset \text{cl } \mathcal{T}_D(\bar{x}) = \mathcal{T}_D(\bar{x}).$$

Todas as outras inclusões da afirmação são óbvias (em particular, veja (1.15)). $\square$

Em geral, os cones $\mathcal{T}_D(\bar{x})$ e $\mathcal{B}_D(\bar{x})$ podem ser diferentes, como mostra o Exemplo 1.4.1 a seguir. No entanto, tal situação é bastante atípica. O conjunto D no Exemplo 1.4.1 é “artificial” no sentido de que é muito improvável que um conjunto de tal estrutura possa vir a aparecer como o conjunto viável de um problema de otimização “de vida real”.

Exemplo 1.4.1. Sejam $q \in (0, 1)$ e $D = \{0\} \cup \{q^0, q^1, q^2, \dots\}$, $D \subset \mathbb{R}$. Consideremos o ponto $\bar{x} = 0 \in D$.

Para todo $d > 0$, escolhemos $t_k = q^k/d$ e $d^k = d$, $k \in \mathbb{N}$. Como $\bar{x} + t_k d^k = q^k \in D$ para todo k , temos que $d \in \mathcal{B}_D(\bar{x})$. Portanto, $\mathcal{B}_D(\bar{x}) = \mathbb{R}_+$.

Mostremos agora que $d = 1 \notin \mathcal{T}_D(\bar{x})$. Seja $t_k = (q^{k+1} + q^k)/2$, $k \in \mathbb{N}$. Temos que

$$\begin{aligned} \frac{\text{dist}(\bar{x} + t_k d, D)}{t_k} &= \frac{\text{dist}(t_k, D)}{t_k} \\ &= \frac{(q^k - q^{k+1})/2}{(q^{k+1} + q^k)/2} \rightarrow \frac{1 - q}{1 + q} \neq 0. \end{aligned}$$

Concluimos que $\mathcal{T}_D(\bar{x}) \neq \mathcal{B}_D(\bar{x})$.

De fato, um argumento similar mostra que todo $d > 0$ não pertence a $\mathcal{T}_D(\bar{x})$. Em particular, $\mathcal{T}_D(\bar{x}) = \{0\}$.

O exemplo seguinte mostra que o cone $\mathcal{B}_D(\bar{x})$ não pode ser substituído por $\mathcal{T}_D(\bar{x})$ na condição suficiente (1.18) do Teorema 1.4.2.

Exemplo 1.4.2. Sejam $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = -x$, e $D \subset \mathbb{R}$ como definido no Exemplo 1.4.1 acima.

Consideremos o ponto $\bar{x} = 0 \in D$. Em toda vizinhança de $\bar{x}$ existe $x = q^k \in D$ (tomando k suficientemente grande) com $f(x) = -q^k < 0 = f(\bar{x})$. Portanto, $\bar{x}$ não é minimizador local de f em D .

Como $f'(\bar{x}) = -1$ e $\mathcal{B}_D(\bar{x}) = \mathbb{R}_+$, a condição (1.18) não é satisfeita. No entanto, $\mathcal{T}_D(\bar{x}) = \{0\}$ e, portanto,

$$\langle f'(\bar{x}), d \rangle > 0 \quad \forall d \in \mathcal{T}_D(\bar{x}) \setminus \{0\}$$

vale trivialmente. Isto mostra que a última condição (ou seja, $\mathcal{T}_D(\bar{x})$ no lugar de $\mathcal{B}_D(\bar{x})$ em (1.18)) não é suficiente para otimalidade.

Exercício 1.4.1. Sejam $K_1 \subset K_2$ dois cones em $\mathbb{R}^n$. Mostrar que $K_2^* \subset K_1^*$.

Exercício 1.4.2. Sejam K_1 e K_2 cones não vazios em $\mathbb{R}^n$. Mostrar as seguintes propriedades:

- (a) $K_1 \times K_2$ é um cone, $(K_1 \times K_2)^* = K_1^* \times K_2^*$;
- (b) $K_1 \cup K_2$ é um cone, $(K_1 \cup K_2)^* = K_1^* \cap K_2^*$;
- (c) $K_1 + K_2$ é um cone, $(K_1 + K_2)^* = K_1^* \cap K_2^*$.

Exercício 1.4.3. Mostrar que o cone $\mathcal{B}_D(\bar{x})$ admite ainda a seguinte definição, equivalente às dadas acima:

$$\mathcal{B}_D(\bar{x}) = \{0\} \cup \left\{ d \in \mathbb{R}^n \mid \begin{array}{l} \exists \{x^k\} \subset D \text{ tal que } \{x^k\} \rightarrow \bar{x}, \\ \{(x^k - \bar{x})/\|x^k - \bar{x}\|\} \rightarrow d/\|d\| \end{array} \right\}.$$

Exercício 1.4.4. Mostrar que (em $\mathbb{R}^n$) a condição (1.16) é equivalente a seguinte:

$$\forall \varepsilon > 0, \quad \langle f'(\bar{x}), d \rangle > -\varepsilon \|d\| \quad \forall d \in \mathcal{B}_D(\bar{x}) \setminus \{0\}.$$

Exercício 1.4.5. Sejam $C \subset \mathbb{R}^n$, $D \subset \mathbb{R}^n$ conjuntos quaisquer. Provar que

$$\mathcal{B}_{C \cup D}(\bar{x}) = \mathcal{B}_C(\bar{x}) \cup \mathcal{B}_D(\bar{x}), \quad \mathcal{B}_{C \cap D}(\bar{x}) \subset \mathcal{B}_C(\bar{x}) \cap \mathcal{B}_D(\bar{x}).$$

Exercício 1.4.6. Sejam $\emptyset \neq D \subset \mathbb{R}^n$, $C = \mathbb{R}^n \setminus D$. Denotamos por $\text{fr } D$ a fronteira de D . Mostrar que se $\bar{x} \in \text{fr } D$, então

$$\mathcal{B}_C(\bar{x}) \cup \mathcal{B}_D(\bar{x}) = \mathbb{R}^n, \quad \mathcal{B}_{\text{fr } D}(\bar{x}) = \mathcal{B}_C(\bar{x}) \cap \mathcal{B}_D(\bar{x}).$$

Exercício 1.4.7. Sejam $D \subset \mathbb{R}^n$ um conjunto não vazio e fechado, $x \notin D$. Seja $P_D(x)$ o conjunto de projeções de x sobre D (veja a Definição 1.2.2 e o Corolário 1.2.2). Mostrar que

$$x - \bar{x} \in (\mathcal{B}_D(\bar{x}))^* \quad \forall \bar{x} \in P_D(x).$$

Capítulo 3

Elementos de Análise Convexa

Neste capítulo estudamos conjuntos convexos e funções convexas. Convexidade é uma noção muito importante na teoria de otimização. Com hipóteses de convexidade, as condições necessárias de otimalidade passam a ser suficientes. Em outras palavras, todo ponto estacionário torna-se uma solução do problema. Em particular, qualquer minimizador local é global. Além disso, no caso convexo podemos desenvolver a teoria de dualidade em sua forma mais completa, i.e., associar ao problema original (primal) um outro problema, chamado dual, que sob certas hipóteses é equivalente ao original e às vezes é mais fácil de resolver. Finalmente, as ferramentas da Análise Convexa serão necessárias para a caracterização do cone dual do cone tangente no caso de restrições mistas (de igualdade e desigualdade), o que resulta nas condições de otimalidade primais-duais de Karush-Kuhn-Tucker. Ressaltamos que este capítulo não constitui um curso da Análise Convexa completo; além de resultados básicos, apresentamos apenas o material que será necessário ao longo deste livro, ou que julgamos indispensável para um curso de Otimização em geral.

3.1 Definições de convexidade. O problema de minimização convexo

Um conjunto convexo se caracteriza por conter todos os segmentos cujos extremos pertencem ao conjunto (veja a Figura 3.1.1).

Definição 3.1.1. Um conjunto $D \subset \mathbb{R}^n$ é chamado *conjunto convexo* se para quaisquer $x \in D$, $y \in D$ e $\alpha \in [0, 1]$, tem-se $\alpha x + (1 - \alpha)y \in D$.

O ponto $\alpha x + (1 - \alpha)y$, onde $\alpha \in [0, 1]$, se chama a *combinação convexa* de x e y (com parâmetro α).

O conjunto vazio, o espaço $\mathbb{R}^n$, e um conjunto que contém um ponto só, são trivialmente convexos. Qualquer conjunto não conexo é trivialmente não convexo.

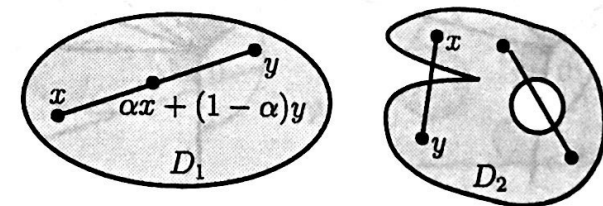


Figura 3.1.1: O conjunto D_1 é convexo; o conjunto D_2 não é convexo.

Exercício 3.1.1. Mostrar que os seguintes conjuntos em $\mathbb{R}^n$ são convexos:

- Qualquer subespaço de $\mathbb{R}^n$.
- Qualquer semi-espaço em $\mathbb{R}^n$, i.e., um conjunto $\{x \in \mathbb{R}^n \mid \langle a, x \rangle \leq c\}$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$.
- Qualquer hiperplano em $\mathbb{R}^n$, i.e., um conjunto $\{x \in \mathbb{R}^n \mid \langle a, x \rangle = c\}$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

(d) Uma bola em $\mathbf{R}^n$, i.e., um conjunto $\{x \in \mathbf{R}^n \mid \|x\| \leq c\}$, onde $c \in \mathbf{R}_+$.

Exercício 3.1.2. Mostrar que o conjunto $\{x \in \mathbf{R}^n \mid \|x\| = c\}$, onde $c \geq 0$, é convexo se, e somente se, $c = 0$.

Em particular, qualquer esfera não trivial (i.e., com $c > 0$) é não convexa.

Exercício 3.1.3. Seja $Q \in \mathbf{R}(n, n)$ uma matriz simétrica semi-definida positiva. Provar que o conjunto $\{x \in \mathbf{R}^n \mid \langle Qx, x \rangle \leq c\}$ é convexo para todo $c \in \mathbf{R}_+$.

A seguir mostramos que a dualização de um cone qualquer (mesmo não convexo) sempre produz um cone convexo, veja a Figura 3.1.2.

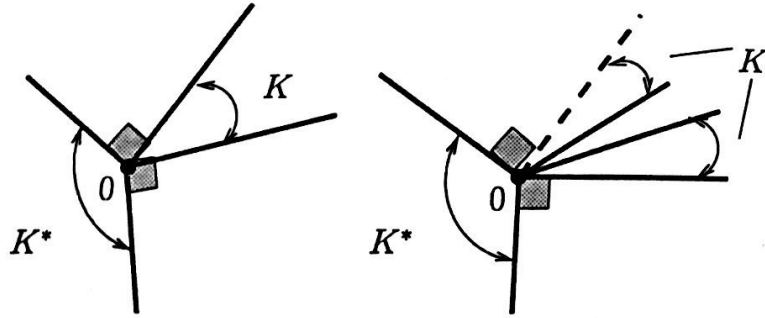


Figura 3.1.2: O cone dual K^* é convexo para K qualquer.

Proposição 3.1.1. Para todo cone $\emptyset \neq K \subset \mathbf{R}^n$, o cone dual K^* sempre é convexo e fechado.

Demonstração: Sejam $x \in K^*$ e $y \in K^*$, i.e., $\langle x, d \rangle \leq 0$ e $\langle y, d \rangle \leq 0$ para todo $d \in K$. Seja $\alpha \in [0, 1]$. Para qualquer $d \in K$, temos que

$$\langle \alpha x + (1 - \alpha)y, d \rangle = \alpha \langle x, d \rangle + (1 - \alpha) \langle y, d \rangle \leq 0,$$

i.e., $\alpha x + (1 - \alpha)y \in K^*$. Portanto, K^* é convexo.

Seja $\{y^k\} \subset K^*$, $\{y^k\} \rightarrow y$ ($k \rightarrow \infty$). Fixando $d \in K$ arbitrário, e passando ao limite quando $k \rightarrow \infty$ na relação $\langle y^k, d \rangle \leq 0$, obtemos que $\langle y, d \rangle \leq 0$. Portanto, como $d \in K$ era arbitrário, $y \in K^*$. Isto que mostra que K^* é fechado. $\square$

Outros critérios de convexidade serão apresentados em § 3.2.

Mostramos a seguir que, para um conjunto convexo, o cone tangente (usual ou do Bouligand) é dado pelo fecho de todas as direções viáveis. Mas primeiro introduzimos a seguinte noção.

Definição 3.1.2. Dado um conjunto $D \subset \mathbf{R}^n$ qualquer, o fecho cônico de D , denotado por $\text{cone } D$, é o menor cone convexo em $\mathbf{R}^n$ que contém D (ou, equivalentemente, a interseção de todos os cones convexos em $\mathbf{R}^n$ que contêm D).

A Figura 3.1.3 ilustra a definição de fecho cônico.

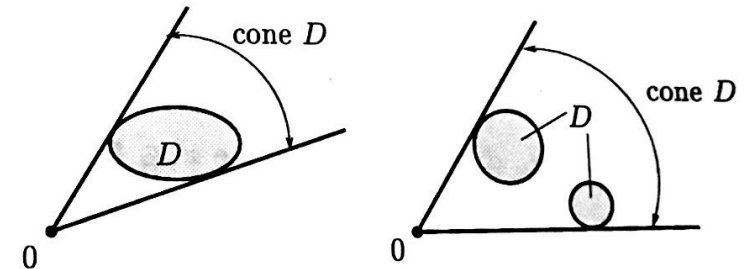


Figura 3.1.3: Exemplos de fecho cônico de um conjunto.

Para um conjunto convexo, o fecho cônico é composto por todos os múltiplos não negativos de elementos do conjunto.

Proposição 3.1.2. (Fecho cônico de conjunto convexo) Seja $D \subset \mathbf{R}^n$ um conjunto convexo. Tem-se que

$$\text{cone } D = \{d \in \mathbf{R}^n \mid d = \alpha x, x \in D, \alpha \in \mathbf{R}_+\}.$$

Demonstração: Denotamos

$$C = \{d \in \mathbf{R}^n \mid d = \alpha x, x \in D, \alpha \in \mathbf{R}_+\}.$$

Como o conjunto cone D é um cone, para todo $x \in D \subset \text{cone } D$, temos que $\alpha x \in \text{cone } D$ para todo $\alpha \in \mathbf{R}_+$. Logo, $C \subset \text{cone } D$.

Como C é um cone que contém D (basta tomar $\alpha = 1$ na definição de C), se provamos que C é convexo, a inclusão $\text{cone } D \subset C$ será óbvia (pela Definição 3.1.2). A Figura 3.1.4 ilustra a demonstração a seguir.

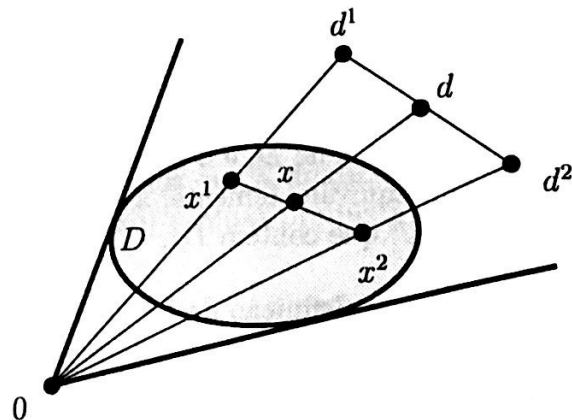


Figura 3.1.4: Ilustração para a demonstração da Proposição 3.1.2.

Sejam $d^i \in C$, i.e., $d^i = \alpha_i x^i$, $\alpha_i \in \mathbf{R}_+$ e $x^i \in D$, $i = 1, 2$.
Seja

$$d = t d^1 + (1-t) d^2 = t \alpha_1 x^1 + (1-t) \alpha_2 x^2, \quad t \in [0, 1].$$

Quando $t \in \{0, 1\}$ ou $\alpha_i = 0$ para $i \in \{1, 2\}$, a inclusão $d \in C$ é óbvia (nestes casos, temos que d é um múltiplo não negativo de x^1 ou de x^2).

Suponhamos então que $t \in (0, 1)$, $\alpha_i > 0$, $i = 1, 2$. Definimos

$$\beta = \left(1 + \frac{\alpha_2(1-t)}{\alpha_1 t}\right)^{-1} \in (0, 1).$$

Pela convexidade de D , temos que $x = \beta x^1 + (1-\beta) x^2 \in D$. Além disso,

$$\frac{\alpha_1 t}{\beta} x = \alpha_1 t (x^1 + (1/\beta - 1) x^2) = t \alpha_1 x^1 + (1-t) \alpha_2 x^2 = d,$$

mostrando que $d \in C$. Isto prova que C é convexo. $\square$

A seguir provaremos um resultado importante sobre o cone tangente de um conjunto convexo. Este resultado é ilustrado na Figura 3.1.5.

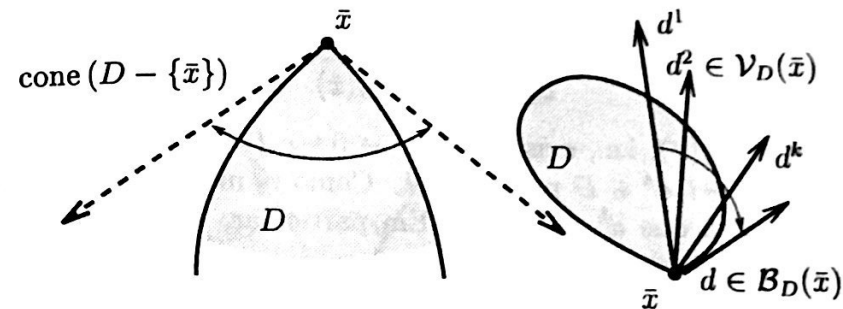


Figura 3.1.5: As direções tangentes e o cone tangente para um conjunto convexo. Relações com as direções viáveis e com o fecho cônico da translação $D - \{\bar{x}\}$.

Teorema 3.1.1. (Cone tangente de um conjunto convexo)

Sejam $D \subset \mathbf{R}^n$ um conjunto convexo, $\bar{x} \in D$. Então

$$\mathcal{T}_D(\bar{x}) = \mathcal{B}_D(\bar{x}) = \text{cl } \mathcal{V}_D(\bar{x}) = \text{cl } \text{cone}(D - \{\bar{x}\}).$$

Em particular, $\mathcal{T}_D(\bar{x})$ é convexo e fechado.

Demonstração: É fácil ver que o conjunto $D - \{\bar{x}\}$ é convexo (Exercício 3.2.1). Portanto, pela Proposição 3.1.2, temos que

$$\text{cone}(D - \{\bar{x}\}) = \{d \in \mathbf{R}^n \mid d = \alpha(x - \bar{x}), x \in D, \alpha \in \mathbf{R}_+\}. \quad (3.1)$$

Seja $d \in \text{cone}(D - \{\bar{x}\})$, $d \neq 0$ (logo, $\alpha > 0$). Pela definição acima, isto significa que $\bar{x} + d/\alpha = x \in D$. Pela convexidade de D , segue-se que $\bar{x} + td \in D$ para todo $t \in [0, 1/\alpha]$, i.e., $d \in \mathcal{V}_D(\bar{x})$ (d é uma direção viável, veja a Definição 1.4.2). Reciprocamente, para $d \in \mathcal{V}_D(\bar{x})$ tem-se que $\bar{x} + td \in D$ para todo $t \in [0, \varepsilon]$, onde

$\varepsilon > 0$. Portanto, existe $x \in D$ tal que $x = \bar{x} + td$, $t > 0$. Logo, $d = (x - \bar{x})/t$, i.e., $d \in \text{cone}(D - \{\bar{x}\})$.

Acabamos de mostrar então que $\mathcal{V}_D(\bar{x}) = \text{cone}(D - \{\bar{x}\})$. Pela Proposição 1.4.1, sempre tem-se que

$$\text{cl } \mathcal{V}_D(\bar{x}) \subset \mathcal{T}_D(\bar{x}) \subset \mathcal{B}_D(\bar{x}).$$

Provamos a afirmação mostrando que

$$\mathcal{B}_D(\bar{x}) \subset \text{cl } \mathcal{V}_D(\bar{x}).$$

Seja $d \in \mathcal{B}_D(\bar{x})$, i.e., existam $\{t_k\} \rightarrow 0+$ e $\{d^k\} \rightarrow d$ ($k \rightarrow \infty$) tais que $\bar{x} + t_k d^k \in D$ para todo k . Como já mostramos acima, isso significa que $d^k \in \mathcal{V}_D(\bar{x})$. Em particular, concluímos que $d = \lim_{k \rightarrow \infty} d^k \in \text{cl } \mathcal{V}_D(\bar{x})$. $\square$

A seguir mostraremos que um cone e o seu fecho têm o mesmo cone dual.

Proposição 3.1.3. *Seja $K \subset \mathbb{R}^n$ um cone qualquer. Então $\text{cl } K$ é um cone e tem-se que*

$$K^* = (\text{cl } K)^*.$$

Em particular, se $D \subset \mathbb{R}^n$ é um conjunto convexo e $\bar{x} \in D$, tem-se que

$$(\mathcal{T}_D(\bar{x}))^* = (\mathcal{V}_D(\bar{x}))^* = (\text{cone}(D - \{\bar{x}\}))^*.$$

Demonstração: O fato que $\text{cl } K$ é um cone é fácil de verificar e é deixado como um exercício para o leitor.

Pela definição de cone dual, o fato que $K \subset \text{cl } K$ implica $(\text{cl } K)^* \subset K^*$ (veja também o Exercício 1.4.1).

Sejam $y \in K^*$ e $d \in \text{cl } K$ quaisquer. Existe $\{d^k\} \subset K$ tal que $\{d^k\} \rightarrow d$ ($k \rightarrow \infty$). Temos que $\langle y, d^k \rangle \leq 0$ para todo k . Passando ao limite quando $k \rightarrow \infty$, obtemos que $\langle y, d \rangle \leq 0$. Portanto, $y \in (\text{cl } K)^*$, i.e., $K^* \subset (\text{cl } K)^*$.

Agora, a última afirmação da proposição segue do Teorema 3.1.1. $\square$

Outra noção útil em Análise Convexa é a do cone normal.

Definição 3.1.3. Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in D$. O *cone normal* (cone de direções normais) no ponto $\bar{x}$ em relação ao conjunto D é dado por

$$\mathcal{N}_D(\bar{x}) = \{d \in \mathbb{R}^n \mid \langle d, x - \bar{x} \rangle \leq 0 \ \forall x \in D\}.$$

A seguir, mostramos que, no caso convexo, o dual do cone tangente é exatamente o cone normal definido acima (veja Figura 3.1.6).

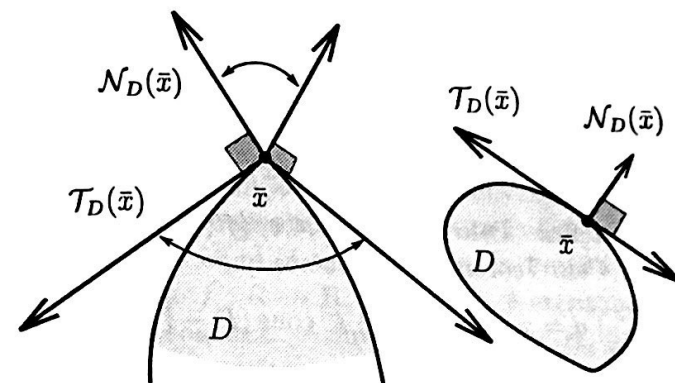


Figura 3.1.6: Exemplos do cone normal de um conjunto convexo. Tem-se que $\mathcal{N}_D(\bar{x}) = (\mathcal{T}_D(\bar{x}))^*$.

Teorema 3.1.2. (Cone normal é dual do cone tangente)
Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $\bar{x} \in D$. Então

$$(\mathcal{T}_D(\bar{x}))^* = (\mathcal{V}_D(\bar{x}))^* = (\text{cone}(D - \{\bar{x}\}))^* = \mathcal{N}_D(\bar{x}).$$

Demonstração: As primeiras duas igualdades já foram provadas (veja a Proposição 3.1.3).

Suponhamos que $y \in (\text{cone}(D - \{\bar{x}\}))^*$, i.e., $\langle y, d \rangle \leq 0$ para todo $d \in \text{cone}(D - \{\bar{x}\})$. Em particular, $\langle y, x - \bar{x} \rangle \leq 0$ para todo $x \in D$, i.e., $y \in \mathcal{N}_D(\bar{x})$.

Suponhamos que $y \in \mathcal{N}_D(\bar{x})$. Temos que $\langle y, x - \bar{x} \rangle \leq 0$ para todo $x \in D$, i.e., $\langle y, d \rangle \leq 0$ para todo $d \in (D - \{\bar{x}\})$.

Logo, $\langle y, d \rangle \leq 0$ para todo $d \in \text{cone}(D - \{\bar{x}\})$. Concluimos que $y \in (\text{cone}(D - \{\bar{x}\}))^*$. $\square$

Como consequência da caracterização do cone tangente e do seu dual, obtemos as seguintes condições de otimalidade para um problema com conjunto viável convexo.

Teorema 3.1.3. (Condição necessária de primeira ordem) Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função diferenciável no ponto $\bar{x} \in D$.

Se $\bar{x}$ é um minimizador local de f no conjunto D , então

$$\langle f'(\bar{x}), x - \bar{x} \rangle \geq 0 \quad \forall x \in D, \quad (3.2)$$

ou, equivalentemente,

$$-f'(\bar{x}) \in \mathcal{N}_D(\bar{x}). \quad (3.3)$$

Demonstração: Pelo Teorema 1.4.1, $\langle f'(\bar{x}), d \rangle \geq 0$ para todo $d \in \mathcal{B}_D(\bar{x})$. Pelo Teorema 3.1.1, temos que

$$\{d \in \mathbb{R}^n \mid d = x - \bar{x}, x \in D\} \subset \text{cone}(D - \{\bar{x}\}) \subset \mathcal{B}_D(\bar{x}),$$

o que implica (3.2). Pela Definição 3.1.3, (3.2) e (3.3) são equivalentes. $\square$

Como vamos provar em § 3.4.2, se a função f é convexa (veja a definição a seguir), a condição necessária de otimalidade dada no Teorema 3.1.3 também é suficiente.

Definição 3.1.4. Se $D \subset \mathbb{R}^n$ é um conjunto convexo, diz-se que a função $f : D \rightarrow \mathbb{R}$ é *convexa* em D quando para quaisquer $x \in D$, $y \in D$ e $\alpha \in [0, 1]$, tem-se

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

A função f diz-se *estritamente convexa* quando a desigualdade acima é estrita para todos $x \neq y$ e $\alpha \in (0, 1)$.

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

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

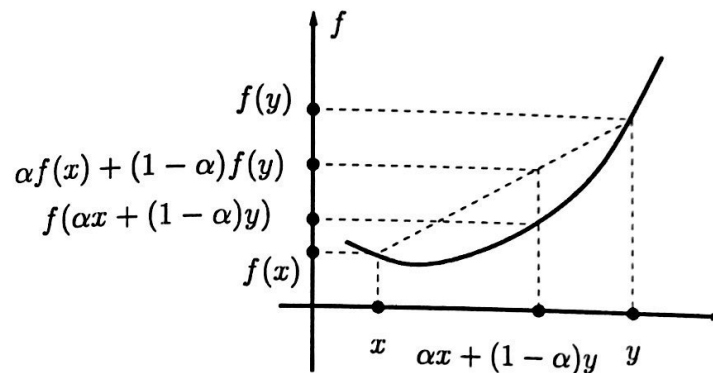


Figura 3.1.7: Ilustração da definição de função convexa: tem-se que $f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y)$.

É óbvio que uma função fortemente convexa é estritamente convexa, e uma função estritamente convexa é convexa. A função $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x^2$, é um exemplo de função fortemente convexa. A função $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = e^x$, é estritamente (mas não fortemente) convexa. A função $f : \mathbb{R} \rightarrow \mathbb{R}$, $f(x) = x$, é convexa (mas não é estritamente convexa).

Em § 3.4 apresentaremos alguns critérios que podem ser usados para reconhecer a que classe uma função dada pertence.

Definição 3.1.5. O *epígrafo* da função $f : D \rightarrow \mathbb{R}$ é o conjunto

$$E_f = \{(x, c) \in D \times \mathbb{R} \mid f(x) \leq c\}.$$

A relação entre convexidade de conjuntos e de funções é dada pelo seguinte teorema.

Teorema 3.1.4. Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Uma função $f : D \rightarrow \mathbb{R}$ é convexa em D se, e somente se, o epígrafo de f é um conjunto convexo em $\mathbb{R}^n \times \mathbb{R}$.

Demonstração: Suponhamos primeiro que E_f seja convexo. Sejam $x \in D$ e $y \in D$ quaisquer. Obviamente, $(x, f(x)) \in E_f$ e $(y, f(y)) \in E_f$. Pela convexidade de E_f , para todo $\alpha \in [0, 1]$

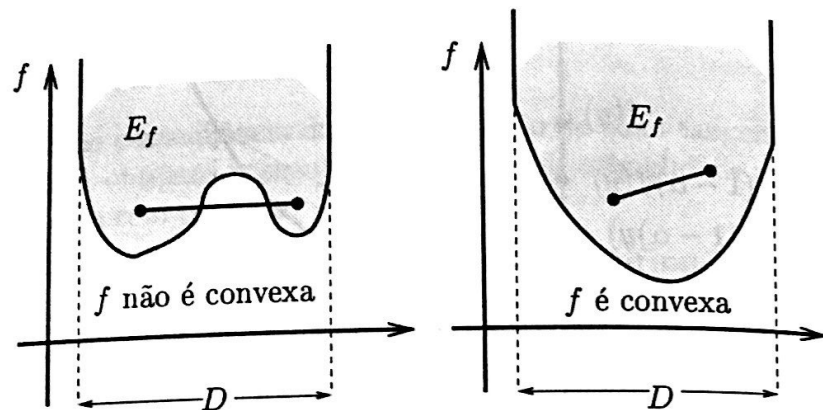


Figura 3.1.8: Convexidade da função $f \Leftrightarrow$ convexidade do epígrafo de f .

temos que

$$\begin{aligned} & (\alpha x + (1 - \alpha)y, \alpha f(x) + (1 - \alpha)f(y)) \\ &= \alpha(x, f(x)) + (1 - \alpha)(y, f(y)) \in E_f. \end{aligned}$$

Pela definição do epígrafo, isto é equivalente a dizer que

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

i.e., f é convexa.

Suponhamos agora que f seja convexa. Sejam $(x, c_1) \in E_f$ e $(y, c_2) \in E_f$. Como $f(x) \leq c_1$ e $f(y) \leq c_2$, pela convexidade de f , para todo $\alpha \in [0, 1]$ tem-se

$$\begin{aligned} f(\alpha x + (1 - \alpha)y) &\leq \alpha f(x) + (1 - \alpha)f(y) \\ &\leq \alpha c_1 + (1 - \alpha)c_2, \end{aligned}$$

o que significa que

$$\begin{aligned} & \alpha(x, c_1) + (1 - \alpha)(y, c_2) \\ &= (\alpha x + (1 - \alpha)y, \alpha c_1 + (1 - \alpha)c_2) \in E_f, \end{aligned}$$

i.e., E_f é convexo. $\square$

Pelo Teorema 3.1.4, de forma equivalente poderíamos definir a classe de funções convexas como as funções cujos epígrafos são convexos.

Dizemos que

$$\min f(x) \text{ sujeito a } x \in D \quad (3.4)$$

é um problema de minimização convexo quando $D \subset \mathbb{R}^n$ é um conjunto convexo e $f : D \rightarrow \mathbb{R}$ é uma função convexa no conjunto D . A importância da convexidade já pode ser vista no resultado seguinte.

Teorema 3.1.5. (Teorema de minimização convexa)

Sejam $D \subset \mathbb{R}^n$ um conjunto convexo e $f : D \rightarrow \mathbb{R}$ uma função convexa em D .

Então todo minimizador local em problema (3.4) é 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 $y \in D$ tal que $f(y) < f(\bar{x})$.

Definimos $x(\alpha) = \alpha y + (1 - \alpha)\bar{x}$. Pela convexidade de D , $x(\alpha) \in D$ para todo $\alpha \in [0, 1]$. Agora, pela convexidade de f , para todo $\alpha \in (0, 1]$, tem-se

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

Tomando $\alpha > 0$ suficientemente pequeno, podemos garantir que o ponto $x(\alpha)$ é arbitrariamente próximo ao ponto $\bar{x}$, e ainda tem-se que $f(x(\alpha)) < f(\bar{x})$ e $x(\alpha) \in D$. Isto contradiz o fato de que $\bar{x}$ é minimizador local do problema (3.4). Portanto, qualquer solução local deve ser global.

Sejam $S \subset D$ o conjunto dos minimizadores (globais) e $\bar{v} \in \mathbb{R}$ o valor ótimo do problema $f(x) = \bar{v}$ para qualquer $x \in S$.

Para quaisquer $x \in S$, $\bar{x} \in S$ e $\alpha \in [0, 1]$, pela convexidade de f obtemos

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

o que implica que de fato $f(\alpha x + (1 - \alpha)\bar{x}) = \bar{v}$ e, portanto, $\alpha x + (1 - \alpha)\bar{x} \in S$. Acabamos de mostrar então que S é convexo.

Suponhamos agora que f seja estritamente convexa e que existam $x \in S$ e $\bar{x} \in S$, $x \neq \bar{x}$. Seja $\alpha \in (0, 1)$. Como x e $\bar{x}$ são minimizadores globais e $\alpha x + (1 - \alpha)\bar{x} \in D$, pela convexidade de D , segue-se

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

No entanto, pela convexidade estrita,

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

o que resulta em contradição. Concluimos que neste caso o minimizador é único. $\square$

Definição 3.1.6. Se $D \subset \mathbb{R}^n$ é um conjunto convexo, dizemos que $f : D \rightarrow \mathbb{R}$ é uma *função côncava* em D , quando a função $(-f)$ é convexa em D .

É fácil ver que as afirmações do Teorema 3.1.5 são verdadeiras se substituirmos minimização de uma função convexa num conjunto convexo por maximização de uma função côncava num conjunto convexo.

Exercício 3.1.4. Seja $D \subset \mathbb{R}^n$ um conjunto convexo. Suponhamos que $x \in \text{int } D$ e $y \in \text{fr } D$. Provar que

$$((1 - \alpha)x + \alpha y) \begin{cases} \in \text{int } D, & \text{se } \alpha \in [0, 1), \\ \notin D, & \text{se } \alpha = 1. \end{cases}$$

Exercício 3.1.5. Seja $K \subset \mathbb{R}^n$ um cone. Provar que K é convexo se, e somente se, $K = K + K$.

Exercício 3.1.6. Sejam D um conjunto convexo em $\mathbb{R}^n$, $c_1 > 0$, $c_2 > 0$. Provar que $(c_1 + c_2)D = c_1D + c_2D$.

Mostrar que a afirmação pode ser falsa quando D é não convexo.

Exercício 3.1.7. Provar que quando $D = \mathbb{R}_+^n$, a condição de otimalidade no Teorema 3.1.3 é equivalente a seguinte *condição de complementaridade*:

$$\bar{x}_i \geq 0, \quad (f'(\bar{x}))_i \geq 0, \quad \bar{x}_i(f'(\bar{x}))_i = 0, \quad i = 1, \dots, n.$$

Exercício 3.1.8. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função (fortemente) convexa e sejam $x \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ quaisquer. Provar que a função $\psi : \mathbb{R} \rightarrow \mathbb{R}$, $\psi(\alpha) = f(x + \alpha d)$, é (fortemente) convexa.

Exercício 3.1.9. Seja $f : \mathbb{R}^m \rightarrow \mathbb{R}$ uma função convexa. Sejam $A \in \mathbb{R}(m, n)$ e $a \in \mathbb{R}^m$. Provar que a função $f(x) = f(Ax + a)$ é convexa em $\mathbb{R}^n$.

Exercício 3.1.10. Provar que a função $f(x) = \|x\|^2$ é fortemente convexa com módulo $\gamma = 1$.

Exercício 3.1.11. Seja $f : \mathbb{R}^n \rightarrow \mathbb{R}$ uma função convexa e, ao mesmo tempo, côncava. Mostrar que isto implica que f é uma função afim, i.e., $f(x) = \langle a, x \rangle + c$ para todo $x \in \mathbb{R}^n$, onde $a \in \mathbb{R}^n$ e $c \in \mathbb{R}$.

3.2 Conjuntos convexos. Teoremas de separação

3.2.1 Propriedades básicas de conjuntos convexos

Continuamos com nosso estudo de estrutura de conjuntos convexos. Mostramos primeiro que a interseção de conjuntos convexos é um conjunto convexo.