# Otimização irrestrita

## Contents



## 1 Objetivos

Definições de minimização, pontos de ótimo local e global. Revisão de diferenciabilidade, gradiente e Hessiana. Minimização irrestrita: condições necessárias de primeira e segunda ordem, condições suficientes de segunda ordem, ponto de sela. Revisão de matrizes positiva (semi)definidas. Existência de pontos ótimos: Teorema de Weierstrass e coercividade. Condições suficientes para pontos ótimos globais. Funções quadráticas: pontos estacionários, ótimos e coercividade.

# 2 Ótimo global e local

Neste curso, iremos estudar problemas de minimização da forma

$$\min_{x \in C} f(x)$$

onde $C \subset \mathbb{R}^n$ é um conjunto e $f : C \to \mathbb{R}$ é uma função. Frequentemente, dizemos que $C$ é o *conjunto viável* e $f$ é a *função objetivo*.

**Definition 1** (Pontos de mínimo)**.** *Seja $f : C \to \mathbb{R}$ uma função definida num conjunto $C \subset \mathbb{R}^n$.*

- *Dizemos que $x^* \in C$ é um ponto de mínimo (global) de $f$ sobre $C$ sse*

$$\forall x \in C, \quad f(x^*) \leq f(x).$$

  *Neste caso, $f^* := f(x^*)$ é chamado valor mínimo (ou valor ótimo).*

- *Dizemos que $x^* \in C$ é um ponto de mínimo (global) estrito de $f$ sobre $C$ sse*

$$\forall x \in C, \quad f(x^*) < f(x).$$

- *Dizemos que $x^* \in C$ é um ponto de mínimo local de $f$ sobre $C$ sse existe $r > 0$ tal que*[1]

$$\forall x \in C \cap B(x^*, r), \quad f(x^*) \leq f(x).$$

- *Dizemos que $x^* \in C$ é um ponto de mínimo local estrito de $f$ sobre $C$ sse existe $r > 0$ tal que*

$$\forall x \in C \cap B(x^*, r) \setminus \{x^*\}, \quad f(x^*) < f(x).$$

**Remark 1** (Pontos de máximo)**.** As definições àcima possuem equivalentes ao problema de maximização da forma

$$\max_{x \in C} f(x).$$

Pontos de máximo (globais, locais e estritos) e valor máximo são definidos analogamente observando que maximizar $f$ é equivalente à minimizar $-f$. □

[1] Recorde que

$$B(x^*, r) = \{x \in \mathbb{R}^n : \|x - x^*\| \leq r\}$$

denota a bola de centro $x^*$ e raio $r$. Aqui, $\|\cdot\|$ denota a norma Euclidiana.

**Example 1.** Considere o caso em que

$$f(x,y) = -(x+y),$$

e

$$C = \{(x,y) : x^2 + y^2 \leq 1\}$$

é a bola unitária. Por Cauchy-Schwarz, temos que

$$-(x+y) = -\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \geq -\sqrt{x^2+y^2}\sqrt{1^2+1^2} \geq -\sqrt{2}.$$

É possível checar que $f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = -\sqrt{2}$. Portanto, $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ é um ponto de mínimo global de $f$ em $C$. Um raciocínio análogo mostra que $\left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$ é um ponto de máximo global com valor máximo $\sqrt{2}$ $\square$

**Example 2.** Considere a função

$$f(x,y) = \frac{x+y}{x^2+y^2+1}$$

definida em todo o espaço bi-dimensional $C = \mathbb{R}^2$. As curvas de contorno e gráfico da função são mostrados abaixo.

**Figure 1**

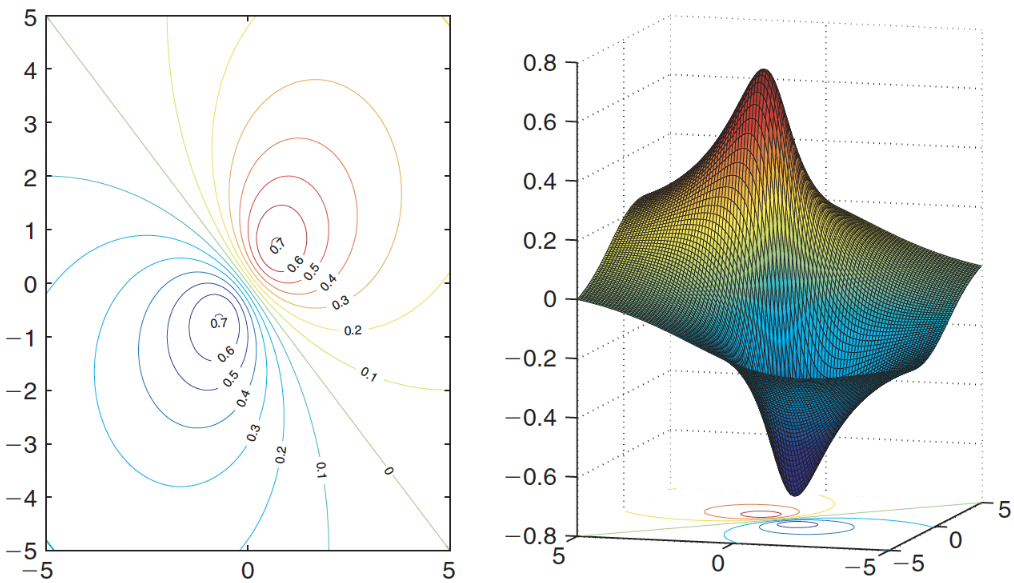


**Figure 2.1.** *Contour and surface plots of* $f(x,y) = \frac{x+y}{x^2+y^2+1}$.

Mostraremos mais tarde que essa função possui dois pontos ótimos: um maximizador global em $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$ e um minimizador global em $\left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right)$. O valor máximo é

$$\frac{\frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}}}{\left(\frac{1}{\sqrt{2}}\right)^2 + \left(\frac{1}{\sqrt{2}}\right)^2 + 1} = \frac{1}{\sqrt{2}},$$

e valor mínimo $\frac{-1}{\sqrt{2}}$. $\square$

**Example 3.** Considere a seguinte função de uma variável definida no intervalo $[-1, 8]$:

$$f(x) = \begin{cases} (x-1)^2 + 2, & -1 \le x \le 1, \\ 2, & 1 \le x \le 2, \\ -(x-2)^2 + 2, & 2 \le x \le 2.5, \\ (x-3)^2 + 1.5, & 2.5 \le x \le 4, \\ -(x-5)^2 + 3.5, & 4 \le x \le 6, \\ -2x + 14.5, & 6 \le x \le 6.5, \\ 2x - 11.5, & 6.5 \le x \le 8. \end{cases}$$

Seu gráfico é plotado a seguir:

**Figure 2**

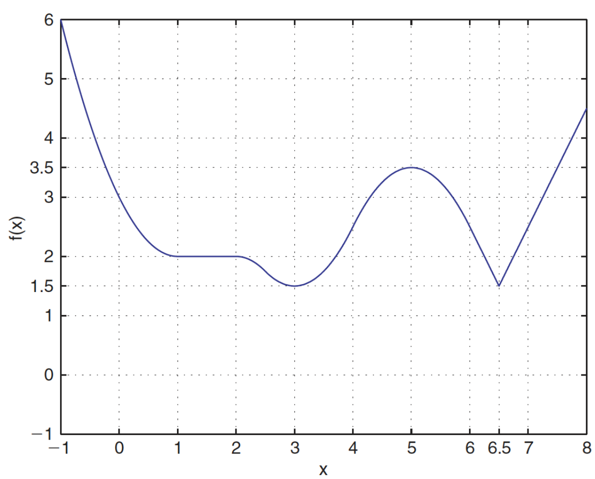


**Figure 2.2.** *Local and global optimum points of a one-dimensional function.*

O ponto $x = -1$ é um ponto de máximo global estrito. O ponto $x = 1$ é um ponto de mínimo local não-estrito. Os pontos no intervalo $(1, 2)$ são pontos de mínimo e máximo locais não-estritos. O ponto $x = 2$ é um ponto de máximo local não-estrito. O ponto $x = 3$ é um ponto de mínimo local estrito e um ponto de mínimo global não-estrito. O ponto $x = 5$ é um ponto de máximo local não-estrito. O ponto $x = 6.5$ é um ponto de mínimo local estrito e um ponto de mínimo global não-estrito. O ponto $x = 8$ é um ponto de máximo local estrito. $\square$

# 3 Recordando conceitos de Cálculo Multivariado

Recordemos o conceito de gradiente de uma função de várias variáveis. Relembramos que iremos usar a convenção $\mathbb{R}^n = \mathbb{R}^{n\times 1}$.

**Definition 2** (Derivada direcional)**.** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função e* $0 \neq d \in \mathbb{R}^n$*. Se o limite*

$$\lim_{t\to 0^+} \frac{f(x+td)-f(x)}{t}$$

*existe, então ele é chamado de* derivada direcional de $f$ ao longo da direção $d$*, e denotada por* $f'(x;d)$*.*

*Quando existente, para cada* $i \in [n]$*, a* i*-ézima derivada parcial de* $f$ *no ponto* $x \in \mathbb{R}^n$ *é o limite*

$$\frac{\partial f}{\partial x_i}(x) := f'(x;e_i) = \lim_{t\to 0^+} \frac{f(x+te_i)-f(x)}{t}.$$

**Definition 3** (Gradiente)**.** *Seja* $f : \mathbb{R}^n \to \mathbb{R}$ *uma função e suponha que existam todas as derivadas parciais no ponto* $x \in \mathbb{R}^n$*. O gradiente de* $f$ *em* $x$ *é o vetor coluna* $\nabla f(x) \in \mathbb{R}^{n\times 1}$ *com* i*-ézima coordenada* $\frac{\partial f}{\partial x_i}(x)$*. Isto é:*

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}.$$

**Remark 2** (Funções continuamente diferenciáveis)**.** Uma função $f : \mathbb{R}^n \to \mathbb{R}$ é dita *continuamente diferenciável* se existem todas derivadas parciais $\frac{\partial f}{\partial x_i}(x)$ em todo ponto $x \in \mathbb{R}^n$ e, além disso, as funções $\frac{\partial f}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$ são contínuas.

Podemos também expressar uma definição similar para uma função $f : C \to \mathbb{R}$ definida num conjunto $C \subset \mathbb{R}^n$. Dizemos que $f : C \to \mathbb{R}$ é *continuamente diferenciável em* $C$ se existem $U \supset C$ *conjunto aberto*[2] tal que todas derivadas parciais $\frac{\partial f}{\partial x_i}(x)$ em todo ponto $x \in U$ e, além disso, as funções $\frac{\partial f}{\partial x_i} : U \to \mathbb{R}$ são contínuas. □

**Remark 3** (Aproximação de primeira ordem)**.** Suponha que $f : U \to \mathbb{R}$ é uma função continuamente diferenciável no conjunto $U \subset \mathbb{R}^n$ aberto e seja $x \in U$. Uma propriedade importante vista em Cálculo é de que

$$\forall d \in \mathbb{R}^n, \quad f'(x;d) = \nabla f(x)^\top d.$$

Esta fórmula é de grande utilidade computacional: conseguimos exprimir um limite $f'(x;d)$ em termos de uma fórmula algébrica que pode ser implementada num programa de computador.

[2]Informalmente, um conjunto aberto é um conjunto "sem fronteiras". Este detalhe técnico é preciso porque para derivadas serem definidads eu preciso de uma vizinhança em que o limite quociente faça sentido. Não exploraremos este tipo de detalhe neste curso.

Uma segunda propriedade relacionada àcima, também vista em Cálculo, permite aproximar, para qualquer $y$ numa vizinhança próxima de $x$, o valor funcional $f(y)$ pela função *afim*

$$y \mapsto f(x) + \nabla f(x)^\top (y - x).$$

Note que a função $f$ pode ser complexa, mas a função afim aproximada é simples de computar. Rigorosamente, para qualquer $y \in \mathbb{R}^n$,

$$\lim_{y \to x} \frac{f(y) - \left[f(x) + \nabla f(x)^\top (y - x)\right]}{\|y - x\|} = 0.$$

Frequentemente, em Análise Numérica e Computação reescrevemos o limite àcima usando a seguinte notação: $o(\cdot) : \mathbb{R}_+ \to \mathbb{R}$ denota uma funcão uni-dimensional satisfazendo a propriedade: $\lim_{t \to 0^+} \frac{o(t)}{t} = 0$. Assim, podemos escrever

$$f(y) - \left[f(x) + \nabla f(x)^\top (y - x)\right] = o(\|y - x\|), \tag{1}$$

e dizemos que o 'resto' da aproximação é da ordem de $o(\|y - x\|)$.

Veremos que esta idéia 'simples' está por trás de vários métodos de otimização que usam 'informação de primeira ordem'. O algoritmo representativo é o *método do gradiente*. □

**Definition 4** (Hessiana)**.** *Seja $f : U \to \mathbb{R}$ uma função duas vezes derivável no conjunto $U \subset \mathbb{R}^n$ aberto e seja $x \in U$. A Hessiana de $f$ em $x$ é a matrix $\nabla^2 f(x) \in \mathbb{R}^{n \times n}$ com entradas $[\nabla^2 f(x)]_{i,j} = \frac{\partial^2 f}{\partial x_j \partial x_i}(x)$. Isto é:*

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}.$$

**Remark 4** (Funções duas vezes continuamente diferenciáveis)**.** Podemos também expressar uma definição similar para uma função $f : C \to \mathbb{R}$ definida num conjunto $C \subset \mathbb{R}^n$. Dizemos que $f : C \to \mathbb{R}$ é *duas continuamente diferenciável em $C$* se existem $U \supset C$ conjunto aberto tal que existem todas derivadas parciais de primeira e segunda ordem em todo ponto $x \in U$ e, além disso, as funções $\frac{\partial^2 f}{\partial x_i \partial x_j} : U \to \mathbb{R}$ são contínuas. □

Um resultado fundamental do Cálculo é o Teorema de Schwarz que diz que, se $f$ é continuamente duas vezes diferenciável em $U$, então, para todo $x \in U$, a Hessiana $\nabla^2 f(x)$ é simétrica:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x), \quad \forall (i, j).$$

**Remark 5** (Aproximação de segunda ordem)**.** Seja $f : U \to \mathbb{R}$ uma função duas vezes continuamente diferenciável num conjunto aberto $U \subset \mathbb{R}^n$. Seja $x \in U$. Neste caso, existem fórmulas mais precisas do que (1) para aproximar a função numa vizinhança de $x$. De fato, podemos aproximar, para qualquer $y$ numa vizinhança próxima de $x$, o valor funcional $f(y)$ pela função *quadrática*

$$y \mapsto f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x),$$

que, novamente, pode ser computada com certa facilidade num computador. Mais precisamente, pode-se mostrar que, para qualquer $y \in \mathbb{R}^n$,

$$\lim_{y \to x} \frac{f(y) - \left[f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x)\right]}{\|y-x\|^2} = 0.$$

Equivalentemente,

$$f(y) - \left[f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x)\right] = o(\|y-x\|^2).$$

Note que o 'resto' da aproximação é da ordem de $o(\|y-x\|^2)$ e portanto significadamente menor do que no caso uma vez diferenciável. Por outro lado, calcular uma quadrática é mais dispendioso do que uma função afim quando $n$ é muito grande.

Está idéia é o que motiva métodos de otimização que usam 'informação de segunda ordem' como o *método de Newton*. □

Terminamos com uma forma alternativa de aproximação de segunda ordem.

**Theorem 1.** *Seja $f : U \to \mathbb{R}$ uma função duas vezes continuamente diferenciável num conjunto aberto $U \subset \mathbb{R}^n$. Seja $x \in U$ e $r > 0$ tais que $B(x, r) \subset U$. Então para todo $y \in B(x, r)$, existe $\xi \in [x, y]$ tais que*

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(\xi)(y-x).$$

# 4 Soluções locais: condições de primeira ordem

Nesta aula, focaremos em problemas de minimização *irrestritos*. Isto é, da forma

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

Isto permitirá focar-nos apenas no comportamento da função objetivo. Veremos em outras aulas a influência do conjunto viável.

Um resultado bem conhecido do cálculo de uma variável é o seguinte. Seja $f$ uma função diferenciável de uma variável definida num intervalo aberto $(a, b)$. Se $x^* \in (a, b)$ é um ponto de máximo ou mínimo local então

$$f'(x^*) = 0.$$

Iremos generalizar este resultado para o caso multivariado usando o conceito de gradiente.

**Theorem 2** (Condições de primeira ordem)**.** *Seja $f : U \to \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Se $x^* \in U$ é um ponto ótimo local e todas as derivadas parciais de $f$ existem em $x^*$, então*

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

*Prova.* Seja $i \in [n]$ e defina a função $g(t) = f(x^* + te_i)$. Temos que $g$ é diferenciável em 0 e $g'(0) = \frac{\partial f}{\partial x_i}(x^*)$. Sendo $x^*$ um ponto ótimo local de $f$, segue que 0 é um ponto ótimo local de $g$; portanto $0 = g'(0) = \frac{\partial f}{\partial x_i}(x^*)$. O argumento vale para todo $i \in [n]$, implicando que $\nabla f(x^*) = 0$. □

O teorema anterior oferece uma condição *necessária* para um ponto ser ótimo local — permitindo restringir a busca de tais pontos. Entretanto, tal condição pode não ser *suficiente*. Um exemplo é a função $f(x) = x^3$ em $x = 0$. Temos que $f'(0) = 0$ mas 0 não é um ponto ótimo local.

**Definition 5.** *Seja $f : U \to \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Suponha que $x^* \in U$ e que todas as derivadas parciais de $f$ existem em $x^*$. Dizemos que $x^*$ é um ponto estacionário de $f$ se*

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

Concluindo, o Teorema 2 diz que pontos ótimos locais são estacionários, mas podem existir pontos estacionários que não são pontos ótimos locais.

# 5 Soluções locais: condições de segunda ordem

Um outro resultado bem conhecido é o seguinte. Seja $f$ uma função de uma variável continuamente duas vezes derivável no intervalo aberto $(a, b)$. Suponha que $x^* \in (a, b)$ é um ponto estacionário de $f$ tal que

$$f''(x^*) > 0.$$

Então $x^*$ é um ponto de mínimo local de $f$ em $(a, b)$. A seguir, iremos generalizar este resultado para o caso multivariado usando o conceito de Hessiana. Para tanto, precisamos definir a noção de "positividade" para matrizes simétricas.

**Definition 6.** *Seja $A \in \mathbb{R}^{n\times n}$ uma matriz simétrica.*

- *Dizemos que $A$ é positiva semidefinida, denotando-se por $A \succeq 0$, se*[3]

$$\forall x \in \mathbb{R}^n, \quad x^\top Ax \geq 0.$$

- *Dizemos que $A$ é positiva definida, denotando-se por $A \succ 0$, se*

$$\forall x \in \mathbb{R}^n \setminus \{0\}, \quad x^\top Ax > 0.$$

[3]Recorde que $A^\top$ denota a matrix transposta da matrix $A$. Além disso, $\langle x, y\rangle := x^\top y$ define o produto interno canônico em $\mathbb{R}^n$.

- *Dizemos que a matriz simétrica $A$ é* *negativa (semi)definida*, *denotando-se por $A \preceq 0$ ($A \prec 0$), se $-A$ é positiva (semi)definida.*

- *Dizemos que a matriz simétrica $A$ é* *indefinida* *se existem $x, y \in \mathbb{R}^n$ tais que $x^\top Ax > 0$ e $y^\top Ay < 0$.*

Em outras palavras, $A$ é indefinida se não é nem positiva ou negativa semidefinida. Ao nos referir a positividade ou negatividade de matrizes, assumiremos que a matriz é simétrica.

**Example 4.** Seja

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}.$$

Para todo $x = \begin{bmatrix} x_1 & x_2 \end{bmatrix}^\top$,

$$x^\top Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2x_1^2 - 2x_1x_2 + x_2^2 = x_1^2 + (x_1 - x_2)^2 \geq 0.$$

Portanto, $A \succeq 0$. De fato, como $x_1^2 + (x_1 - x_2)^2 = 0$ sse $x_1 = x_2 = 0$ tem-se que $A \succ 0$. $\square$

**Example 5.** Seja

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}.$$

Para todo $x = \begin{bmatrix} 1 & -1 \end{bmatrix}^\top$,

$$x^\top Ax = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2.$$

Portanto, $A$ não é positiva definida. Como $e_1^\top Ae_1 = A_{11} = 1$, também $A$ não é negativa definida. Portanto, $A$ é indefinida. $\square$

Os exemplos anteriores mostram que não podemos inferir a positivade/negatividade de uma matriz apenas verificando a positividade/negatividade de seus elementos. Entretanto, como mencionado no Ex. 3, se a matriz é positiva (semi)definida então os elementos de sua diagonal são positivos (não-negativos). Além disso, se a matriz tem um elemento positivo e um elemento negativo na diagonal, então tal matriz é indefinida. Veja Ex. 4. Usando o Teorema Espectral de Álgebra Linear, a caracterização de positividade de uma matrix simétrica é equivalente à positividade de seus auto-valores, conforme enunciado no teorema à seguir.

**Theorem 3.** *Seja matriz simétrica $A \in \mathbb{R}^{n\times n}$. Então*

(i) *$A \succeq 0$ sse todos autovalores de $A$ são não-negativos.*

(ii) *$A \succ 0$ sse todos autovalores de $A$ são positivos.*

(iii) *$A$ é indefinida sse $A$ possui um autovalor negativo e um auto-valor positivo.*

**Theorem 4** (Condições necessárias de segunda ordem)**.** *Seja $f : U \to \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Suponha que $f$ é continuamente duas vezes diferenciável sobre $U$ e seja $x^* \in U$. Então*

(i) *Se $x^*$ é um ponto de mínimo local, então $\nabla^2 f(x^*) \succeq 0$.*

(ii) *Se $x^*$ é um ponto de máximo local, então $\nabla^2 f(x^*) \preceq 0$.*

Note que as condições anteriores são necessárias mas não suficientes. Por exemplo, tome a função $f(x) = x^3$. Temos $f''(0) = 0$, mas $0$ não é nem ponto de máximo nem de mínimo.

**Theorem 5** (Condições suficientes de segunda ordem)**.** *Seja $f : U \to \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Suponha que $f$ é continuamente duas vezes diferenciável sobre $U$ e seja $x^* \in U$ um ponto estacionário de $f$. Então*

(i) *Se $\nabla^2 f(x^*) \succ 0$, então $x^*$ é um ponto de mínimo local estrito.*

(ii) *Se $\nabla^2 f(x^*) \prec 0$, então $x^*$ é um ponto de máximo local estrito.*

Notemos que a condição suficiente do Teorema 5 refere-se a pontos ótimos estritos. Entretanto, conforme o Teorema 4, positividade *estrita* da Hessiana num ponto mínimo local não ocorre necessariamente — apenas semi-positividade é garantida. Por exemplo, a função $f(x) = x^4$ tem como $0$ um ponto de mínimo estrito, mas $f''(0) = 0$ não é positiva.

**Definition 7** (Ponto de sela)**.** *Seja $f : U \to \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Suponha que $f$ é continuamente duas vezes diferenciável sobre $U$. Um ponto estacionário $x^* \in U$ de $f$ é chamado ponto de sela de $f$ em $U$ se não é nem um ponto de mínimo nem um ponto de máximo de $f$ em $U$.*

**Theorem 6** (Condições suficiente para ponto de sela)**.** *Seja $f : U \to \mathbb{R}$ uma função definida no conjunto aberto $U \subset \mathbb{R}^n$. Suponha que $f$ é continuamente duas vezes diferenciável sobre $U$ e seja $x^* \in U$ um ponto estacionário de $f$. Se $\nabla^2 f(x^*)$ é indefinida, então $x^*$ é um ponto de sela.*

O leitor curioso pode ver as demonstrações dos teoremas no Apêndice.

# 6 Existência de pontos ótimos

Até agora assumimos que existem pontos de mínimos/máximos. Nessa seção apresentamos alguns critérios sobre o problema de otimização que permitem garantir a existência de pontos ótimos.

Recorde que um conjunto $C \subset \mathbb{R}^n$ é *compacto* se é fechado e limitado.[4]

[4]Um conjunto $C \subset \mathbb{R}^n$ é fechado seu complementar $C^c$ é aberto. Tal conjunto $C$ é limitado se existe $r > 0$ tal que $C \subset B[0, r]$.

**Theorem 7** (Weierstrass)**.** *Seja uma função contínua definida sobre um conjunto compacto $C \subset \mathbb{R}^n$. Então $f$ possui um ponto de mínimo global e um ponto de máximo global em $C$.*

Quando o conjunto não é compacto o Teorema de Weierstrass não garante a existência de pontos ótimos. Nesse caso, a propriedade de *coercividade* da função pode ser usada. Informalmente, uma função é coerciva se seu gráfico tem a forma de um "tanque" em que podemos encher de água sem nunca transbordar.

**Definition 8** (Coercividade)**.** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função contínua definida sobre $\mathbb{R}^n$. A função $f$ é dita coerciva se*[5]

$$\lim_{\|x\| \to \infty} f(x) = \infty.$$

**Theorem 8** (Existência de soluções: coercividade)**.** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ uma função contínua e coerciva e $C \subset \mathbb{R}^n$ um conjunto fechado não-vazio. Então $f$ tem um mínimo global em $C$.*

**Example 6.** Considere a função

$$f(x_1, x_2) = x_1^2 + x_2^2,$$

e conjunto

$$C = \{(x_1, x_2) : x_1 + x_2 \leq -1\}.$$

O conjunto $C$ é fechado mas não-compacto, portanto, não podemos usar o Teorema de Weierstrass. Entretanto, a função $f$ é coerciva, seguindo que existe ponto de mínimo global.

Temos duas possibilidades. Em primeiro lugar, se existe um ponto de mínimo $x$ no interior $\mathrm{int}(C) = \{(x_1, x_2) : x_1 + x_2 < -1\}$, então do Teorema 2 devemos ter $\nabla f(x) = 0$, implicando que $x = 0$. Entretanto, $0 \notin C$, sendo um absurdo.

Em segundo lugar, devemos então ter um ponto de mínimo $x$ na fronteira $\partial C = \{(x_1, x_2) : x_1 + x_2 = -1\}$ de $C$. Substituindo $x_1 = -x_2 - 1$, transformamos o problema em minimizar a função unidimensional $g(x_2) = (-1 - x_2)^2 + x_2^2$ sobre $\mathbb{R}$. A solução de $0 = g'(x_2) \equiv 2(1 + x_2) + 2x_2$ é $x_2^* = -0.5$. Substituindo $x_1^* = -x_2^* - 1 = -0.5$, segue que $(x_1^*, x_2^*) = (-0.5, -0.5)$ é um ponto de mínimo de $f$ sobre $C$. □

Nos exemplos anteriores minimizamos uma função sem restrições. No exemplo anterior o conjunto de restrições impacta diretamente o conjunto de mínimos. Voltaremos neste ponto no próximos capítulos. Para o exemplo à seguir usaremos o seguinte resultado.

**Proposition 1.** *Seja $A$ uma matriz simétrica $2 \times 2$. Temos que $A$ é positiva semidefinida (definida) sse $\mathrm{tr}(A) \geq 0$ e $\det(A) \geq 0$ ($\mathrm{tr}(A) > 0$ e $\det(A) > 0$).*

**Example 7.** Considere a função

$$f(x_1, x_2) = 2x_1^3 + 3x_2^2 + 3x_1^2 x_2 - 24x_2,$$

[5] Aqui denotemos a norma $\|x\| = \sqrt{x^\top x}$.

sobre todo o espaço bidimensional $\mathbb{R}^2$. Acharemos e classificaremos todos os pontos estacionários de de $f$ sobre $\mathbb{R}^2$. Temos

$$\nabla f(x_1, x_2) = \left[ \begin{array}{c} 6x_1^2 + 6x_1x_2 \\ 6x_2 + 3x_1^2 - 24 \end{array} \right].$$

Os pontos estacionários são a solução do sistema:

$$\begin{aligned} 6x_1^2 + 6x_1x_2 &= 0, \\ 6x_2 + 3x_1^2 - 24 &= 0. \end{aligned}$$

A primeira equação é equivalente a $x_1 = 0$ ou $x_1 + x_2 = 0$. Se $x_1 = 0$, da segunda equação tem-se $x_2 = 4$. Se $x_1 + x_2 = 0$, substituindo-se na segunda equação tem-se $3x_1^2 - 6x_1 - 24 = 0$. Obtemos as soluções $x_1 = 4, -2$. Em conclusão, os pontos estacionários de $f$ são

$$(0, 4), (4, -4), (-2, 2).$$

A Hessiana de $f$ é

$$\nabla^2 f(x_1, x_2) = 6 \left[ \begin{array}{cc} 2x_1 + x_2 & x_1 \\ x_1 & 1 \end{array} \right].$$

Para o ponto estacionário $(0, 4)$,

$$\nabla^2 f(0, 4) = 6 \left[ \begin{array}{cc} 4 & 0 \\ 0 & 1 \end{array} \right] \succ 0,$$

portanto $(0, 4)$ é um ponto de mínimo local estrito. Ele não é um ponto de mínimo global já que $f$ não é limitada por baixo:

$$f(x_1, 0) = 2x_1^3 \to -\infty \text{ quando } x_1 \to -\infty.$$

Para o ponto estacionário $(4, -4)$,

$$\nabla^2 f(4, -4) = 6 \left[ \begin{array}{cc} 4 & 4 \\ 4 & 1 \end{array} \right].$$

Como $\det(\nabla^2(4, -4)) = -6^2 \cdot 12 < 0$, temos que $\nabla^2(4, -4)$ não é positiva semidefinida. Também, $\nabla^2(4, -4)$ possui elemento diagonal positivo, logo não é negativa semidefinida. Segue que $\nabla^2(4, -4)$ é indefinida e $(4, -4)$ é um ponto de sela.

Para o ponto estacionário $(-2, 2)$,

$$\nabla^2 f(-2, 2) = 6 \left[ \begin{array}{cc} -2 & -2 \\ -2 & 1 \end{array} \right],$$

que é indefinida já que possui elementos na diagonal positivos e negativos. Segue que $(-2, 2)$ é um ponto de sela. $\square$

**Example 8.** Retornemos ao Exemplo 2, obteremos os pontos estacionários da função

$$f(x,y) = \frac{x+y}{x^2+y^2+1}.$$

Temos

$$\nabla f(x,y) = \frac{1}{(x^2+y^2+1)^2} \begin{bmatrix} x^2+y^2+1-2(x+y)x \\ x^2+y^2+1-2(x+y)y \end{bmatrix}.$$

Os pontos estacionários são a solução do sistema:

$$\begin{aligned} -x^2 - 2xy + y^2 &= -1, \\ x^2 - 2xy - y^2 &= -1. \end{aligned}$$

Adicionando as duas equações obtemos $xy = \frac{1}{2}$. Subtraindo-as, obtemos $x^2 = y^2$, implicando que $x = y$. Portanto, $x^2 = \frac{1}{2}$. Concluímos que os pontos estacionários de $f$ são

$$\left(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).$$

Para o ponto estacionário $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, note que $f(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = \frac{1}{\sqrt{2}}$. Em seguida,

$$f(x,y) = \frac{x+y}{x^2+y^2+1} \leq \sqrt{2}\frac{\sqrt{x^2+y^2}}{x^2+y^2+1} \leq \sqrt{2}\max_{t\geq 0} \frac{t}{t^2+1},$$

onde usamos Cauchy-Schwarz. Para todo $t \geq 0$, $t^2+1 \geq 2t$, implicando que $f(x,y) \leq \frac{1}{\sqrt{2}}$ para todo $(x,y) \in \mathbb{R}^2$. Segue que $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ atinge o ponto de máximo sendo portanto um ponto de máximo global.

Um argumento análogo mostra que $(\frac{-1}{\sqrt{2}}, \frac{-1}{\sqrt{2}})$ é ponto de mínimo global. □

# 7 Condições para soluções globais

As condições anteriores referem-se apenas a pontos de ótimos locais: temos informação do sinal da Hessiana na vizinhança de um ponto. Quando a Hessiana é positiva semidefinida em todo ponto, então os pontos estacionários são pontos de mínimo globais. De fato, veremos futuramente que esta propriedade implica convexidade da função.

**Theorem 9.** *Seja $f : \mathbb{R}^n \to \mathbb{R}$ função duas vezes continuamente diferenciável. Suponha que*

$$\nabla^2 f(x) \succeq 0, \quad \forall x \in \mathbb{R}^n.$$

*Então, todo ponto estacionário $x^* \in \mathbb{R}^n$ de $f$ é um ponto de mínimo global.*

*Prova.* Pelo Teorema 1, para todo $x \in \mathbb{R}^n$, existe $\xi_x \in [x^*, x]$ tal que

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)\nabla^2 f(\xi_x)(x - x^*).$$

Para todo $x \in \mathbb{R}^n$ $\nabla^2 f(\xi_x) \succeq 0$. Temos que

$$\forall x \in \mathbb{R}^n, \quad f(x) \geq f(x^*),$$

e, portanto, $x^*$ é ponto de mínimo global de $f$. □

# 8 Funções quadráticas

Um caso clássico é quando temos uma função *quadrática*, isto é, da forma

$$f(x) = x^\top A x + 2b^\top x + c,$$

onde $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. O gradiente e Hessiana podem ser expressos como (veja Ex. 5)

$$\begin{aligned} \nabla f(x) &= 2Ax + 2b \\ \nabla^2 f(x) &= 2A. \end{aligned}$$

Usando estas fórmulas, podemos caracterizar os pontos estacionários e pontos ótimos de uma função quadrática.

**Lemma 1** (Pontos estacionários e ótimos de função quadrática)**.** *Seja uma função como àcima com $A$ simétrica. Então*

(i) *$x$ é ponto estacionário sse $Ax = -b$.*

(ii) *Suponha que $A \succeq 0$. Então $x$ é ponto de mínimo global sse $Ax = -b$.*

(iii) *Suponha que $A \succ 0$. Então $x = -A^{-1}b$ é ponto de mínimo global estrito.*

*Prova.* (i) Segue imediatamente da fórmula do gradiente.

(ii) Suponha que $A \succeq 0$. Da ́formula da Hessiana, segue que $\nabla^2 f(x) \succeq 0$ para todo $x \in \mathbb{R}^n$. O resultado segue então do Teorema 9 e item (i).

(iii) Suponha que $A \succ 0$. Então $x = -A^{-1}b$ é a única solução de $Ax = -b$. Segue do item (ii) que $x = -A^{-1}b$ é o único ponto de mínimo global de $f$ e, portanto, mínimo global estrito. $\square$

Concluímos com a caracterização da coercividade de funções quadráticas.

**Lemma 2** (Coercividade de funções quadráticas)**.** *Seja função $f(x) = x^\top A x + 2b^\top x + c$, onde $A \in \mathbb{R}^{n \times n}$ é uma matriz simétrica, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$. Então $f$ é coerciva sse $A \succ 0$.*

# 9 Exercícios

**Exercise 1.** Considere a função

$$f(x) = 3x^4 - 20x^3 + 42x^2 - 36x.$$

Justificando, ache seus pontos de mínimo locais e globais. *Dica*: mostre que $\lim_{|x| \to \infty} f(x) = \infty$, implicando que $f$ tem mínimo global.

**Exercise 2.** Refaça o Exemplo 8, mas desta vez achando os pontos estacionários e analizando suas matrizes hessianas.

**Exercise 3.** Seja $A \in \mathbb{R}^{n\times n}$ uma matriz simétrica. Mostre que se $A$ possui um elemento diagonal negativo entao $A$ não é positiva definida. Analogamente, mostre que se $A$ possui um elemento diagonal positivo entao $A$ não é negativa definida.

**Exercise 4.** Seja $A \in \mathbb{R}^{n\times n}$ uma matriz simétrica. Mostre que se $A$ possui um elemento diagonal negativo e um elemento diagonal positivo entao $A$ é indefinida.

**Exercise 5.** Verifique que o gradiente e Hessiana da função

$$f(x) = x^\top Ax + 2b^\top x + c,$$

onde $A \in \mathbb{R}^{n\times n}$ é uma matriz simétrica, $b \in \mathbb{R}^n$ e $c \in \mathbb{R}$, são expressos por:

$$\begin{aligned} \nabla f(x) &= 2Ax + 2b \\ \nabla^2 f(x) &= 2A. \end{aligned}$$

# 10 Apêndice

## 10.1 Prova do Teorema 4

Provaremos apenas (i) já que o item (ii) segue do item (i) aplicado à função $-f$.

Sendo $x^*$ um ponto de mínimo local, existe bola $B(x^*, r) \subset U$ tal que

$$\forall x \in B(x^*, r), \quad f(x) \geq f(x^*).$$

Seja $0 \neq d \in \mathbb{R}^n$. Para todo $0 < \alpha < \frac{r}{\|d\|}$, $x^*_\alpha := x^* + \alpha d \in B(x^*, r)$ e portanto

$$f(x^*_\alpha) \geq f(x^*). \tag{2}$$

Pelo Teorema 1, existe $\xi_\alpha \in [x^*, x^*_\alpha]$ tal que

$$f(x^*_\alpha) - f(x^*) = \nabla f(x^*)^\top (x^*_\alpha - x^*) + \frac{1}{2}(x^*_\alpha - x^*)^\top \nabla^2 f(\xi_\alpha)(x^*_\alpha - x^*).$$

Sendo $x^*$ um ponto estacionário ($\nabla f(x^*) = 0$) segue que

$$f(x^*_\alpha) - f(x^*) = \frac{\alpha^2}{2} d^\top \nabla^2 f(\xi_\alpha) d. \tag{3}$$

Combinando (2)-(3) obtemos que para todo $0 < \alpha < \frac{r}{\|d\|}$,

$$d^\top \nabla^2 f(\xi_\alpha) d \geq 0.$$

Usando que $z_\alpha \to x^*$ quando $\alpha \to 0^+$ e por continuidade da Hessiana, segue que

$$d^\top \nabla^2 f(x^*) d \geq 0.$$

## 10.2 Prova do Teorema 5

Provaremos apenas (i) já que o item (ii) segue do item (i) aplicado à função $-f$.

Seja $x^*$ um ponto estacionário tal que $\nabla^2 f(x^*) \succ 0$. Sendo a Hessiana contínua, segue que existe bola $B(x^*, r) \subset U$ tal que $\nabla^2 f(x) \succ 0$ para todo $x \in B(x^*, r)$. Pelo Teorema 1 segue que, para todo $x \in B(x^*, r)$, existe $\xi_x \in [x^*, x] \subset B(x^*, r)$ tal que

$$f(x) - f(x^*) = \frac{1}{2}(x - x^*)^\top \nabla^2 f(\xi_x)(x - x^*).$$

Tem-se $\nabla^2 f(\xi_x) \succ 0$ para todo $x \in B(x^*, r)$. Segue que

$$\forall x \neq x^*, \quad f(x) > f(x^*),$$

isto é, que $x^*$ é ponto de mínimo local.

## 10.3 Prova do Teorema 6

Seja $\nabla^2 f(x^*)$ é indefinida. Portanto, $\nabla^2 f(x^*)$ possui auto-valor positivo $\lambda_1$ associado ao auto-vetor $v_1$ com norma $\|v_1\| = 1$. Sendo $U$ aberto, existe $r > 0$ tal que $x^* + \alpha v_1 \in U$ para todo $\alpha \in (0, r)$. Pelo Teorema **??** e usando que $\nabla f(x^*) = 0$, sabemos que existe função $o : \mathbb{R}_+ \to \mathbb{R}$ satisfazendo

$$\lim_{t \to 0} \frac{o(t)}{t} = 0, \tag{4}$$

tal que para todo $\alpha \in (0, r)$,

$$\begin{aligned} f(x^* + \alpha v_1) &= f(x^*) + \frac{\alpha^2}{2} v_1^\top \nabla^2 (x^*) v_1 + o(\alpha^2 \|v_1\|^2) \\ &= f(x^*) + \frac{\lambda_1 \alpha^2}{2} \|v_1\|^2 + o(\alpha^2 \|v_1\|^2) \\ &= f(x^*) + \frac{\lambda_1 \alpha^2}{2} + o(\alpha^2). \end{aligned}$$

De (4), segue que existe $\epsilon_1 \in (0, r)$ tal que

$$\forall \alpha \in (0, \epsilon_1), \quad g(\alpha^2) > -\frac{\lambda_1 \alpha^2}{2}.$$

Portanto,

$$\forall \alpha \in (0, \epsilon_1), \quad f(x^* + \alpha v_1) > f(x^*).$$

Isto implica que $x^*$ não pode ser um máximo local sobre $U$.

Um argumento análogo usando que $\nabla^2 f(x^*)$ tem auto-valor negativo $\lambda_2 < 0$ associado a um auto-vetor $v_2$ de norma 1, mostra que existe $\epsilon_2 \in (0, r)$ tal que

$$\forall \alpha \in (0, \epsilon_2), \quad f(x^* + \alpha v_2) < f(x^*).$$

Portanto, $x^*$ não pode ser ponto de mínimo local sobre $U$.

## 10.4 Prova do Teorema 8

Seja $x_0 \in C$ um ponto arbitrário. Como $f$ é coerciva, segue que existe $M > 0$ tal que

$$f(x) > f(x_0) \text{ para todo } x \text{ tal que } \|x\| > M.$$

Temos que $x^*$ é um ponto de mínimo global de $f$ sobre $C$. Portanto $f(x^*) \geq f(x_0)$. Segue da afirmação em diplay que o conjunto de mínimos globais de $f$ sobre $C$ é exatamente o conjunto de mínimos globais de $f$ sobre $C \cap B[0, M]$. O conjunto $C \cap B[0, M]$ é fechado e limitado, portanto compacto. Segue do Teorema de Weierstrass que $f$ possui ponto de mínimo global sobre $C \cap B[0, M]$, e portanto, sobre $C$ também.

## 10.5 Prova do Lemma 2

Necessitaremos do seguinte lemma, que decorre to Teorema Espectral visto em Álgebra Linear.

**Lemma 3.** *Seja $A \in \mathbb{R}^{n\times n}$ simétrica. Então, para todo $x \neq 0$,*

$$\lambda_{\min}(A) \leq \frac{x^\top Ax}{\|x\|^2} \leq \lambda_{\max}(A).$$

($\Longleftarrow$). Suponha que $A \succ 0$. Denote $\alpha := \lambda_{\min}(A)$. Pelo lem1a àcima e Cauchy-Schwarz, segue que, para todo $x \in \mathbb{R}^n$,

$$\begin{aligned} f(x) &= x^\top Ax + 2b^\top x + c \\ &\geq \alpha\|x\|^2 - 2\|b\|\|x\| + c. \end{aligned}$$

Segue que $f(x) \to \infty$ quando $\|x\| \to \infty$; isto é, $f$ é coerciva.

($\Longrightarrow$). Suponha que $f$ é coerciva. Suponha que $A$ tenha auto-valores negativos. Portanto, existem $v \neq 0$ e $\lambda < 0$ tais que $Av = \lambda v$. Portanto, para todo $\alpha \in \mathbb{R}$,

$$f(\alpha v) = \lambda\|v\|^2\alpha^2 + 2(b^\top v)\alpha + c \to -\infty \text{ as } \alpha \to \infty.$$

Isto contradiz a hipótese de coercividade. Portanto, $A$ possui todos auto-valores não-negativos.

Provaremos agora que 0 não é auto-valor de $A$, provando que $A \succ 0$. Assuma que exista $v \neq 0$ tal que $Av = 0$. Então, para todo $\alpha \in \mathbb{R}$,

$$f(\alpha v) = 2(b^\top v)\alpha + c.$$

Temos que

$$f(\alpha v) \to \begin{cases} c, & \text{as } \alpha \to \infty \text{ se } b^\top v = 0, \\ -\infty, & \text{as } \alpha \to -\infty \text{ se } b^\top v > 0, \\ -\infty, & \text{as } \alpha \to \infty \text{ se } b^\top v < 0. \end{cases}$$

Em qualquer caso contradizemos a hipótese de que $f$ é coerciva. Portanto, 0 não pode ser auto-valor de $A$.