Convexidade

Fernanda Costa, Sofia Lopes

Departamento de Matemática
Universidade do Minho

Outline

Conjuntos convexos

O conceito de convexidade é fundamental na otimização. Muitos problemas práticos possuem esta propriedade, o que geralmente os torna mais fáceis de resolver, tanto na teoria como na prática. O termo "convexo" pode ser aplicado tanto a conjuntos como a funções.

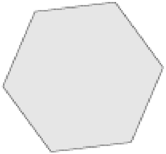
Definição (Conjunto Convexo)

Um conjunto $\mathcal{D} \subseteq \mathbb{R}^d$ é convexo se o segmento de reta entre quaisquer dois pontos de $\mathcal{D}$ está contido em $\mathcal{D}$. Ou seja, $\forall w, z \in \mathcal{D}$ e $\forall t \in [0, 1]$

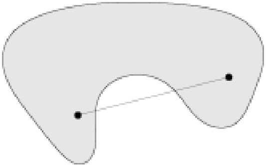
$$t\, w + (1 - t)z \in \mathcal{D}.$$

- Um ponto da forma $t\, w + (1 - t)z \in \mathcal{D}$ é chamado uma combinação convexa de w e z.
- Notar que quando $t = 0$, estamos em z, quando $t = 1$ estamos em w, e para valores intermédios de t, estamos no segmento de recta que liga w e z.

Exemplos:



convexo



não convexo

não convexo

Definição (Combinação Convexa)

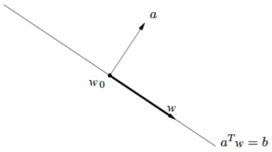
Sejam $w_1, w_2, \ldots, w_n \in \mathbb{R}^d$ *e* $t_1, t_2, \ldots, t_n \geq 0$ *com* $\sum_{i=1}^{n} t_i = 1$. *Os pontos da forma*

$$t_1 w_1 + t_2 w_2 + \ldots + t_n w_n$$

são chamados combinações convexas de $w_1, w_2, \ldots, w_n \in \mathbb{R}^d$.

Exemplos de conjuntos convexos

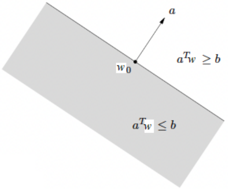
1. $\mathbb{R}^d$
2. $\emptyset$ (por convenção);
3. Conjuntos unitários $\{w_0 \in \mathbb{R}^d\}$
4. Subespaços de $\mathbb{R}^d$
5. Rectas: $R = \{w \in \mathbb{R}^2 : a^T w = b\}$ (com $a \in \mathbb{R}^2 \backslash \{0\}, b \in \mathbb{R}$);
6. Hiperplanos: $H = \{w \in \mathbb{R}^d : a^T w = b\}$ (com $a \in \mathbb{R}^d \backslash \{0\}, b \in \mathbb{R}$, $d > 2$);



Nota:

- $a \neq 0$ é o vetor normal ao hiperplano, i.e. normal a qualquer vetor contido no hiperplano definido por $a^T(w - w_\mathbf{0}) = 0$.
- Um hiperplano pode ser interpretado como dividindo o espaço em dois semi-espaços: $\{w \in \mathbb{R}^d : a^T w \leq b\}$ e $\{w \in \mathbb{R}^d : a^T w \geq b\}$

6. Semi-espaços: $S = \{w \in \mathbb{R}^d : a^T w \leq b\}$;



7. Bolas Euclidianas:
$B(w_c, r) = \{w \in \mathbb{R}^d : \|w - w_c\|_2 \leq r\}$ (com $w_c \in \mathbb{R}^d, r \in \mathbb{R}^+$)

8. Conjunto das matrizes simétricas semi-definidas positivas:

$$S = \{P \in \mathcal{M}_{\{n \times n\}} : w^T P w \geq 0, \forall w \in \mathbb{R}^n\}.$$

■ Estas demonstrações ficam como Exercícios.

Lema

O conjunto solução de um sistema de equações lineares $S = \{w \in \mathbb{R}^d : Aw = b\}$ *onde* $A \in \mathbb{R}^{m\times d}$ *e* $b \in \mathbb{R}^m$ *é um conjunto convexo.*

Demonstração:
Sejam quaisquer $u, z \in S$ e $t \in [0, 1]$. Então temos

$$Au = b, \qquad Az = b.$$

Queremos mostrar que $w := tu + (1 - t)z \in S$, ou seja, $Aw = b$.
Vejamos: pela linearidade da multiplicação por matriz, temos

$$Aw = A\big(tu + (1 - t)z\big) = tAu + (1 - t)Az = tb + (1 - t)b = b.$$

Portanto $w \in S$, donde se conclui que S é convexo. $\square$

Propriedades dos conjuntos convexos

Proposição (Operações que preservam a convexidade de conjuntos)

i) *Sejam* $D_1, D_2, \ldots, D_n$ *conjuntos convexos de* $\mathbb{R}^d$. *Então a intersecção* $D = D_1 \cap D_2 \cap \cdots \cap D_n$ *é um conjunto convexo.*

ii) *Sejam* D_1 *e* D_2 *dois conjuntos convexos de* $\mathbb{R}^d$. *Então a soma e a subtração destes conjuntos é um conjunto convexo:* $D_1 \pm D_2 = \{w \pm z : w \in D_1, z \in D_2\}$.

iii) *Seja* D *um conjunto convexo e* $\alpha \in \mathbb{R}$ *um escalar arbitrário. Então o conjunto* $\alpha\, D = \{\alpha w : w \in D, \alpha \in \mathbb{R}\}$ *é um conjunto convexo.*

iv) *Sejam* $D_i \subset \mathbb{R}^{d_i}$ *para* $i = 1, \ldots, k$, *conjuntos convexos. Então* $D_1 \times \ldots \times D_k$ *é um conjunto convexo de* $\mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_k}$.

- As demonstrações ficam como Exercícios.

Definição (Conjunto poliedral)

Um conjunto poliedral, ou simplesmente poliedro, é o conjunto solução de um sistema de equações e inequações lineares:

$$S = \{w \in \mathbb{R}^d : A_1 w = b_1, A_2 w \leq b_2\}$$

onde $A_1 \in \mathbb{R}^{n\times d}$, $A_2 \in \mathbb{R}^{p\times d}$, $b_1 \in \mathbb{R}^n$ *e* $b_2 \in \mathbb{R}^p$.

Lema

O conjunto poliedral $S = \{w \in \mathbb{R}^d : A_1 w = b_1, A_2 w \leq b_2\}$ *é convexo.*

Demonstração: Um conjunto poliedral é o conjunto solução de um sistema finito de equações e inequações lineares, isto é, resulta da interseção de um conjunto afim ($\{w : \mathbf{A_1} w = \mathbf{b_1}\}$) e de semi-espaços ($\{w : a_i^\top w \leq b_i\}$, onde os a_i^T correspondem às linhas de $\mathbf{A_2}$, $i = 1, \ldots, p$), cada um dos quais é convexo. Como a interseção arbitrária de conjuntos convexos é convexa, conclui-se que S é convexo. $\square$

Funções convexas

Definição (Função convexa)

Seja D um conjunto convexo não vazio. Uma função $F : D \to \mathbb{R}$ diz-se uma função convexa se, para todo $w, z \in D$ e para todo $t \in [0, 1]$ se tem que

$$F(tw + (1-t)z) \leq tF(w) + (1-t)F(z). \tag{1}$$

Exemplo:

- Geometricamente, a desigualdade (1) significa que o segmento de reta que une os pontos $(w, F(w))$ e $(z, F(z))$, encontra-se acima ao gráfico de F.
- F é estritamente convexa se a desigualdade em (1) for estrita, para todo $w \neq z$ e $t \in]0, 1[$.
- Diz-se que F é côncava se $-F$ é convexa.

Exemplos de funções convexas em $\mathbb{R}$

1. $F(w) = aw + b; \quad a, b \in \mathbb{R}; a \neq 0$ **(funções lineares/afins)**
2. $F(w) = aw^2 + bw + c; \quad a \in \mathbb{R}^+; b, c \in \mathbb{R}; a \neq 0$ **(funções quadráticas)**
3. $F(w) = e^{aw}; \quad a \in \mathbb{R}$.
4. $F(w) = -\log w; \quad$ (definida em $\mathbb{R}^+$).
5. $F(w) = w^a; \quad$ (definida em $\mathbb{R}^+$), $a \geq 1$ ou $a \leq 0$.
6. $F(w) = -w^a; \quad$ (definida em $\mathbb{R}^+$), $0 \leq a \leq 1$.
7. $F(w) = |w|^a; \quad a \geq 1$.
8. $F(w) = w \log w; \quad$ (definida em $\mathbb{R}^+$).

Represente graficamente as funções de (1) a (8) e certifique-se visualmente da sua convexidade.

Em breve, veremos algumas caracterizações de funções convexas que tornam a tarefa de verificar a convexidade um pouco mais fácil.

Exemplos de funções convexas em $\mathbb{R}^d$:

- Funções lineares: $F(w) = a^T w + b; \quad a \in \mathbb{R}^d$ e $b \in \mathbb{R}$

  Demonstração:
  $\forall u, z \in \mathbb{R}^d$ e $\forall t \in [0,1]$. Queremos mostrar que $F(tu + (1-t)z) \leq tF(u) + (1-t)F(z)$.
  Vejamos:

  $F(tu + (1-t)z) = a^T(tu + (1-t)z) + b = ta^T u + (1-t)a^T z + b =$

  $ta^T u + (1-t)a^T z + b + tb - tb = t(a^T u + b) + (1-t)(a^T z + b) = tF(u) + (1-t)F(z)$. $\square$

- Qualquer norma de $\mathbb{R}^d$, ou seja, qualquer função que satisfaça:
  a) $F(\alpha w) = |\alpha| F(w), \forall \alpha \in \mathbb{R}$
  b) $F(w + z) \leq F(w) + F(z)$
  c) $F(w) \geq 0, \forall\ w$; $F(w) = 0 \Rightarrow w = 0$.

  Demonstração:

  $\forall u, z \in \mathbb{R}^d$ e $\forall t \in [0,1]$: $F(tu + (1-t)z) \underset{b)}{\leq} F(tu) + F((1-t)z) \underset{a)}{=} tF(u) + (1-t)F(z)$. $\square$

Exemplos de normas ($\mathbb{R}^d$):

- $\|w\|_\infty = \max\{|w_1|, |w_2|, \ldots, |w_d|\}$

- $\|w\|_p = \left(\sum_{i=1}^{d} |w_i|^p\right)^{\frac{1}{p}}$, $p \geq 1$

  $p = 1$: $\|w\|_1 = |w_1| + |w_2| + \cdots + |w_d|$

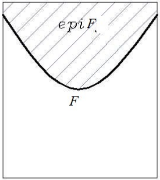
  $p = 2$: $\|w\|_2 = \sqrt{w_1^2 + w_2^2 + \cdots + w_d^2}$

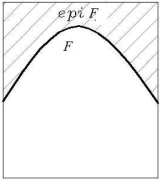
Definição

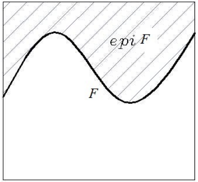
O ***epigrafo*** epi F *de uma função* $F : \mathbb{R}^d \to \mathbb{R}$ *é um subconjunto de* $\mathbb{R}^{d+1}$ *definido por*

$$\text{epi}\, F = \{(w, z) : w \in \text{dom}\, F, F(w) \leq z\}$$

Exemplos:







Teorema

Uma função $F : \text{dom}\, F \subseteq \mathbb{R}^d \to \mathbb{R}$ *é convexa se e só se o seu epigrafo é um conjunto convexo.*

Teorema 1 (condição de convexidade de 0ª ordem)

Uma função $F : \mathbb{R}^d \to \mathbb{R}$ *é convexa se e só se para todo* $w \in \operatorname{dom} F$ *e para todo* $z \in \mathbb{R}^d$, *a função* $g : \mathbb{R} \to \mathbb{R}$ *definida por* $g(\alpha) = F(w + \alpha z)$ *é convexa. (O domínio de* g *é* $\operatorname{dom} g = \{\alpha \in \mathbb{R} : w + \alpha z \in \operatorname{dom} F\}$*).*

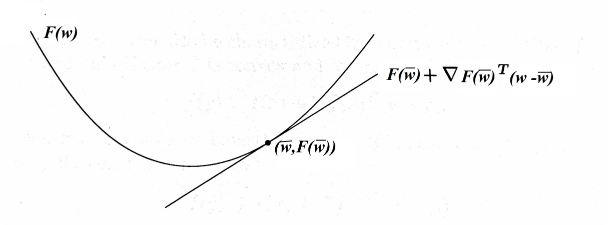
- Esta teorema é muito útil, pois permite verificar se uma função F é convexa quando restringida a qualquer reta que intersecta o seu domínio.
- Muitos dos algoritmos que iremos estudar trabalham minimizando a função F ao longo de retas. É importante reter que a restrição de uma função convexa a uma reta permanece convexa.

Funções convexas diferenciáveis

Teorema 2 (condição de convexidade de 1ª ordem)

Seja $D \subseteq \mathbb{R}^d$ *um conjunto convexo aberto não vazio e* $F : D \to \mathbb{R}$ *uma função diferenciável em* D*. Então* F *é convexa se, e só se, para todo* $w, \bar{w} \in D$

$$F(w) \geq F(\bar{w}) + \nabla F(\bar{w})^T (w - \bar{w}). \tag{2}$$



Nota: Uma consequência deste teorema é a seguinte: Se F é uma função diferenciável e convexa, e $\nabla F(\bar{w}) = 0$, então $\bar{w}$ é um minimizante global da função F em D (i.e, $\forall w \in D$, $F(w) \geq F(\bar{w})$).

Exemplo: Vamos mostrar que $F(w_1, w_2) = w_1^2 + w_2^2$ é uma função convexa, usando o Teorema 2:

$$F \text{ é convexa} \quad \Leftrightarrow \quad F(w) \geq F(\bar{w}) + \nabla F(\bar{w})^T (w - \bar{w}).$$

Vejamos:

$$\begin{aligned} w_1^2 + w_2^2 \geq \bar{w}_1^2 + \bar{w}_2^2 + \begin{bmatrix} 2\bar{w}_1 & 2\bar{w}_2 \end{bmatrix} \begin{bmatrix} w_1 - \bar{w}_1 \\ w_2 - \bar{w}_2 \end{bmatrix} \\ = \bar{w}_1^2 + \bar{w}_2^2 + 2\bar{w}_1 w_1 - 2\bar{w}_1^2 + 2\bar{w}_2 w_2 - 2\bar{w}_2^2 \\ = -\bar{w}_1^2 - \bar{w}_2^2 + 2\bar{w}_1 w_1 + 2\bar{w}_2 w_2. \end{aligned}$$

Ou seja,

$$w_1^2 + w_2^2 + \bar{w}_1^2 + \bar{w}_2^2 - 2\bar{w}_1 w_1 - 2\bar{w}_2 w_2 \geq 0 \quad \Leftrightarrow \quad (w_1 - \bar{w}_1)^2 + (w_2 - \bar{w}_2)^2 \geq 0. \; (1)$$

Como (1) é verdadeira para todo $w = (w_1, w_2), \bar{w} = (\bar{w}_1, \bar{w}_2) \in \mathbb{R}^2$, concluimos que F é convexa.

Teorema 3 (condição de convexidade de 2ª ordem)

Seja dom F um conjunto não vazio e convexo em $\mathbb{R}^d$ e seja $F : \mathbb{R}^d \to \mathbb{R}$ uma função duas vezes continuamente diferenciável sobre um conjunto aberto que contém dom F.

a) *Se $\nabla^2 F(w)$ é semi-definida positiva $\forall w \in$ dom F, então F é convexa em dom F.*

b) *Se $\nabla^2 F(w)$ é definida positiva $\forall w \in$ dom F, então F é estritamente convexa em dom F.*

c) *Se dom F é aberto e F é convexa em dom F, então $\nabla^2 F(w)$ é semi-definida positiva $\forall w \in$ dom F.*

Nota:

- Similarmente: F é concâva se $\nabla^2 F(w)$ é semi-definida negativa $\forall w \in dom\, F$; e é estritamente concâva se $\nabla^2 F(w)$ é definida negativa $\forall w \in dom\, F$

Recordar:

- Uma matriz quadrada $A \in \mathbb{R}^{d \times d}$ é definida positiva se $s^T A s > 0, \;\; \forall\, s \in \mathbb{R}^d$ e $s \neq 0$
- A é semi-definida positiva se $s^T A s \geq 0, \forall s \in \mathbb{R}^d$.

Exemplo: Vamos mostrar que $F(w) = w\log(w)$, $F : \mathbb{R}^+ \to \mathbb{R}$, é uma função convexa usando Teorema 3:

$$F''(w) \geq 0, \quad \forall w \in \mathbb{R}^+ \Rightarrow F \quad \text{é convexa.}$$

Vejamos:
$\mathbb{R}^+$ é um conjunto convexo, aberto e não vazio.

$$\begin{aligned} F'(w) &= \log w + 1, \\ F''(w) &= 1/w. \end{aligned}$$

Como $F'(w)$ e $F''(w)$ são contínuas em $\mathbb{R}^+$ e $F''(w) > 0$ para $\forall w \in \mathbb{R}^+$, a função F é (estritamente) convexa em $\mathbb{R}^+$.

Exemplo: Vamos mostrar que $F(w_1, w_2) = \frac{w_1^2}{w_2}$ com $w_2 > 0$, $F : \mathbb{R}^2 \to \mathbb{R}$, é uma função convexa usando o Teorema 3:

$$\nabla^2 F(w_1, w_2) \text{ é semi-definida positiva } \forall w \in dom\, F \Rightarrow F \quad \text{é convexa.}$$

Vejamos:
$dom\, F = \{(w_1, w_2) \in \mathbb{R}^2 : w_2 > 0\}$ é um conjunto convexo, aberto e não vazio.

Derivadas parciais de primeira ordem:

$$\begin{aligned} \frac{\partial F}{\partial w_1}(w_1, w_2) &= \frac{2w_1}{w_2} \\ \frac{\partial F}{\partial w_2}(w_1, w_2) &= -\frac{w_1^2}{w_2^2} \end{aligned}$$

Derivadas parciais de segunda ordem:

$$\begin{aligned} \frac{\partial^2 F}{\partial w_1^2}(w_1, w_2) &= \frac{2}{w_2} \\ \frac{\partial^2 F}{\partial w_1 \partial w_2}(w_1, w_2) &= -\frac{2w_1}{w_2^2} \\ \frac{\partial^2 F}{\partial w_2^2}(w_1, w_2) &= \frac{2w_1^2}{w_2^3} \end{aligned}$$

e, portanto, a matriz Hessiana é: $\nabla^2 F(w_1, w_2) = \begin{bmatrix} \frac{2}{w_2} & -\frac{2w_1}{w_2^2} \\ -\frac{2w_1}{w_2^2} & \frac{2w_1^2}{w_2^3} \end{bmatrix}$.

Vamos agora calcular o determinante $\nabla^2 F(w_1, w_2)$, usando o Critério de Sylvester:

- $|\frac{2}{w_2}| = \frac{2}{w_2}$. Como $w_2 > 0$, temos que $\frac{2}{w_2} > 0$.
- $\begin{vmatrix} \frac{2}{w_2} & -\frac{2w_1}{w_2^2} \\ -\frac{2w_1}{w_2^2} & \frac{2w_1^2}{w_2^3} \end{vmatrix} = \frac{2}{w_2} \times \frac{2w_1^2}{w_2^3} - \left(-\frac{2w_1}{w_2^2}\right) \times \left(-\frac{2w_1}{w_2^2}\right) = \frac{4w_1^2}{w_2^4} - \frac{4w_1^2}{w_2^4} = 0.$

Como $|\frac{2}{w_2}| > 0$ e $\begin{vmatrix} \frac{2}{w_2} & -\frac{2w_1}{w_2^2} \\ -\frac{2w_1}{w_2^2} & \frac{2w_1^2}{w_2^3} \end{vmatrix} = 0$, a matriz Hessiana é semidefinida positiva.

Portanto, a função F é convexa em $dom\, F$.

Propriedades das funções convexas

Proposição (Operações que preservam a convexidade de funções)

i) *Sejam* $D \subseteq \mathbb{R}^d$ *um conjunto convexo não vazio e* F_i, $i = 1, \ldots p$, *funções convexos em* D. *Então para quaisquer* $\alpha_i \in \mathbb{R}^+$, $i = 1, \ldots p$, *a função* $F(w) = \sum_{i=1}^{p} \alpha_i F_i(w)$ *é convexa em* D.

ii) *Sejam* $G : \mathbb{R}^d \to \mathbb{R}$ *uma função convexa e* $H : \mathbb{R} \to \mathbb{R}$ *uma função convexa crescente. Então a função* $F : \mathbb{R}^d \to \mathbb{R}$ *definida por* $F(w) = H \circ G(w) = H(G(w))$ *é convexa.*

iii) *Sejam* $D \subseteq \mathbb{R}^d$ *um conjunto convexo não vazio e* F_i, $i = 1, \ldots p$, *funções convexos em* D. *Então a função* $F : D \to \mathbb{R}$ *definida por* $F(w) = \sup_{i=1,\ldots,p} F_i(w)$ *é convexa.*

Os diferentes modos para provar que uma função é convexa:

- Usar a definição de função convexa.
- Mostrar que o epigráfo de função é convexo.
- $0^{\underline{a}}$ ordem: verificar que $\forall s \in \mathbb{R}^d$, $g(\alpha) = F(w + \alpha s)$ é convexa (no seu domínio $\{\alpha : w + \alpha s \in dom\ F\}$).
- $1^{\underline{a}}$ ordem: verificar que $\forall w, \bar{w} \in dom\ F,\ F(w) \geq F(\bar{w}) + (w - \bar{w})^T \nabla F(\bar{w})$.
- $2^{\underline{a}}$ ordem: verificar que $\nabla^2 F(w)$ é semi-definida positiva em todos os pontos do $dom\ F$.
- Construindo a partir de funções convexas usando operações de preservação de convexidade.