Optimização e Algoritmos (2004/2005)

Instituto Superior Técnico – Engenharia Electrotécnica e de Computadores

Série de Problemas 3

Regras de Armijo e Wolfe, Introdução às funções convexas

Problema 1.[Regras de Armijo e Wolfe] Considere o problema de optimização

$$\min_{t>0} \phi(t) \tag{1}$$

onde $\phi : [0, +\infty) \to \mathbb{R}$ é uma função de classe C^1 (ou seja, diferenciável e com derivada contínua). Aqui, assume-se que $\phi'(0) < 0$ (ou seja, ϕ é inicialmente decrescente). Este programa de optimização surge como subproblema nos algoritmos de optimização que operam por pesquisa em linhas. Nesse contexto, dado que o objectivo primordial é minimizar uma função objectivo f em $\mathbb{R}^n$ (da qual ϕ é apenas uma fatia unidimensional), não se torna importante resolver *exactamente* o problema (1). É suficiente resolver (1) *aproximadamente*, ou seja, basta encontrar em $\mathbb{R}^+ = (0, +\infty)$ um ponto t que seja "aceitável". O critério de "aceitabilidade" depende da regra adoptada. Neste problema, introduzem-se 2 regras: a regra de Armijo e a regra de Wolfe. Ambas as regras estabelecem uma partição de $\mathbb{R}^+$ em 3 conjuntos disjuntos: $\mathcal{A}, \mathcal{D}, \mathcal{E}$. Esses conjuntos têm o seguinte significado: o conjunto $\mathcal{A}$ é o conjunto dos pontos "aceitáveis"; $\mathcal{D}$ é o conjunto dos pontos demasiado "à direita"; o conjunto $\mathcal{E}$ é o conjunto dos pontos demasiado "à esquerda". (Atenção: esta terminologia não significa forçosamente que, por exemplo, todos os pontos em $\mathcal{D}$ sejam estritamente maiores que os pontos em $\mathcal{A}$; veja os exemplos adiante). Mais especificamente, temos

$$\text{Armijo:} \quad \begin{array}{l} \mathcal{A} = \{\, t > 0 \,:\, \phi(t) \le \phi(0) + \sigma\phi'(0)t \,\} \\ \mathcal{D} = \{\, t > 0 \,:\, \phi(t) > \phi(0) + \sigma\phi'(0)t \,\} \\ \mathcal{E} = \emptyset \end{array} \tag{2}$$

$$\text{Wolfe:} \quad \begin{array}{l} \mathcal{A} = \{\, t > 0 \,:\, \phi(t) \le \phi(0) + \sigma\phi'(0)t \text{ e } \phi'(t) \ge \lambda\phi'(0) \,\} \\ \mathcal{D} = \{\, t > 0 \,:\, \phi(t) > \phi(0) + \sigma\phi'(0)t \,\} \\ \mathcal{E} = \{\, t > 0 \,:\, \phi(t) \le \phi(0) + \sigma\phi'(0)t \text{ e } \phi'(t) < \lambda\phi'(0) \,\} \end{array} \tag{3}$$

As constantes σ e λ são escolhidas pelo utilizador, mas devem satisfazer $0 < \sigma < \lambda < 1$. (Note que, para a regra de Armijo, nunca existem pontos demasiado "à esquerda").

(a) [Regra de Armijo: exemplo 1] Considere a função

$$\phi(t) = 8t^5 - \frac{98}{3}t^4 + 42t^3 - \frac{95}{6}t^2 - 3t$$

cujo gráfico se apresenta na figura 1. Usando a regra de Armijo, determine (aproximadamente) os conjuntos $\mathcal{A}$, $\mathcal{E}$ e $\mathcal{D}$ e represente-os em $\mathbb{R}^+$ para **(i)**

$\sigma = 0.4$ **(ii)** $\sigma = 0.6$ e **(iii)** $\sigma = 0.8$. Sugere-se a utilização do MATLAB

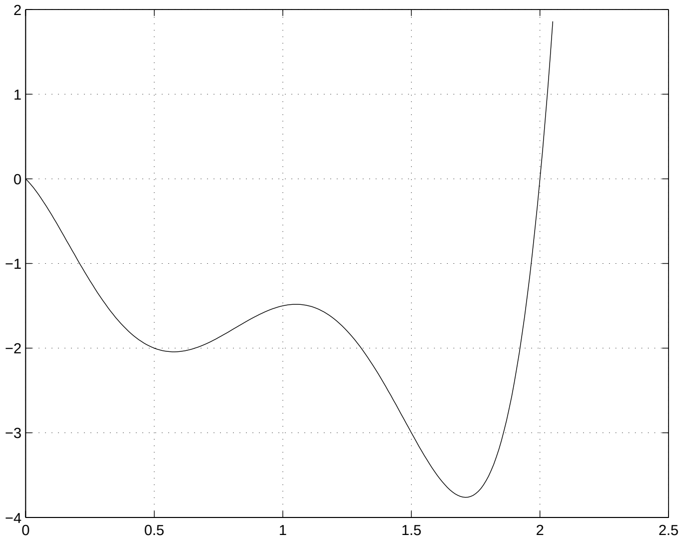

Figura 1: $\phi(t)$ [Problema 1(a)]

para uma determinação numérica mais precisa de $\mathcal{A}, \mathcal{D}, \mathcal{E}$. Em qualquer dos casos, é útil neste tipo de problemas começar por desenhar as rectas $\phi(0) + \sigma\phi'(0)t$ sobrepostas ao gráfico da função $\phi(t)$. A título de exemplo, exibe-se na figura 2 o caso referente a $\sigma = 0.4$.

(b) [Regra de Armijo: exemplo 2] Repita a alínea anterior para a função

$$\phi(t) = t^2 - 2t$$

cujo gráfico se apresenta na figura 3.

(c) [Regra de Wolfe: exemplo 1] Considere de novo a função

$$\phi(t) = 8t^5 - \frac{98}{3}t^4 + 42t^3 - \frac{95}{6}t^2 - 3t$$

cujo gráfico se apresenta na figura 1. Usando a regra de Wolfe, determine os conjuntos $\mathcal{A}$, $\mathcal{E}$ e $\mathcal{D}$ e represente-os em $\mathbb{R}^+$ para **(i)** $(\sigma, \lambda) = (0.4, 0.6)$ **(ii)** $(\sigma, \lambda) = (0.6, 0.8)$ e **(iii)** $(\sigma, \lambda) = (0.8, 0.9)$.

(d) [Regra de Wolfe: exemplo 2] Considere de novo a função

$$\phi(t) = t^2 - 2t$$

cujo gráfico se apresenta na figura 3. Usando a regra de Wolfe, determine os conjuntos $\mathcal{A}$, $\mathcal{E}$ e $\mathcal{D}$ e represente-os em $\mathbb{R}^+$ para **(i)** $(\sigma, \lambda) = (0.4, 0.6)$ **(ii)** $(\sigma, \lambda) = (0.6, 0.8)$ e **(iii)** $(\sigma, \lambda) = (0.8, 0.9)$.

Problema 2.[Regra de Armijo: propriedade básica] Considere a regra de Armijo descrita no problema 1 e uma função $\phi : [0, +\infty) \to \mathbb{R}$ de classe

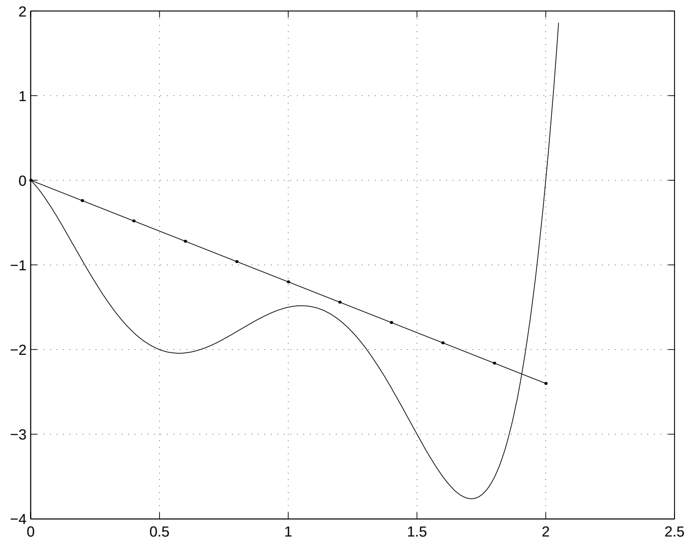

Figura 2: $\phi(t)$ e $\phi(0) + 0.4\phi'(0)t$ [Problema 1(a)]

C^1 e com $\phi'(0) < 0$. Mostre que, para qualquer $d \in \mathcal{D}$, existe um ponto "aceitável" no intervalo $(0, d)$. [**Pista:** note que $\phi'(0) < \sigma\phi'(0)$. Agora, dado que ϕ' é contínua, existe $t^* > 0$ tal que $\phi'(t) < \sigma\phi'(0)$ para todo o $t \in [0, t^*]$. Explorando a identidade

$$\phi(t^*) = \phi(0) + \int_0^{t^*} \phi'(t)dt$$

conclua que t^* é "aceitável"]

Problema 3.[Regra de Wolfe: propriedade básica] Considere a regra de Wolfe descrita no problema 1 e uma função $\phi : [0, +\infty) \to \mathbb{R}$ de classe C^1 e com $\phi'(0) < 0$. O objectivo deste problema é mostrar que, para quaisquer $e \in \mathcal{E}$ e $d \in \mathcal{D}$ satisfazendo $e < d$, existe um ponto "aceitável"no intervalo (e, d).

(a) Defina o conjunto $S = (e, +\infty) \cap \mathcal{D}$ e seja $m = \inf S$. Mostre que $m > e$. [**Pista:** assuma que $m = e$. Então, mostre que é possível sintetizar uma sequência $t_k > e$, que converge para e e satisfaz

$$\frac{\phi(t_k) - \phi(e)}{t_k - e} \geq \sigma\phi'(0).$$

Tomando o limite $k \to +\infty$ na expressão acima, conclua que $\phi'(e) \geq \sigma\phi'(0)$ e que esta última desigualdade contradiz $e \in \mathcal{E}$.]

(b) Note que, pela alínea anterior, já sabemos que $\phi(t) \leq \phi(0) + \sigma\phi'(0)t$ para todo o $t \in (0, m)$. Agora, mostre que

$$\phi(m) \geq \phi(0) + \sigma\phi'(0)m.$$

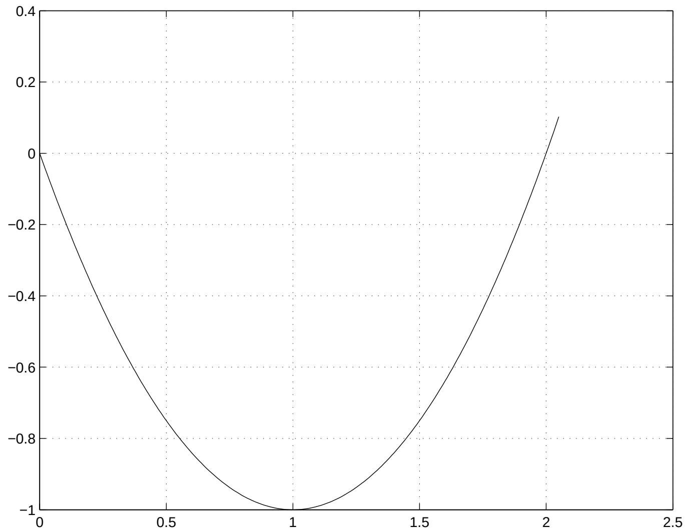

Figura 3: $\phi(t)$ [Problema 1(b)]

[**Pista:** considere uma sequência $t_k \in \mathcal{D}$ que converge para m.]

(c) Use o resultado da alínea anterior para mostrar a existência de um ponto $t^* \in (e, m) \subset (e, d)$ que satisfaz $\phi'(t^*) > \lambda\phi'(0)$. [**Pista:** aplique o teorema do valor intermédio à função ϕ no intervalo $[e, m]$. Recorde que este teorema garante a existência de um ponto $t^* \in (e, m)$ que satisfaz

$$\phi'(t^*) = \frac{\phi(m) - \phi(e)}{m - e}.$$

Mostre que a expressão acima implica $\phi'(t^*) \geq \sigma\phi'(0)$ e note que $0 < \sigma < \lambda < 1$.]

(d) Mostre que t^* é um ponto "aceitável".

Problema 4.[Localização de pontos aceitáveis: parte I] Considere o contexto do problema 1 que visa a resolução *aproximada* do problema de optimização

$$\min_{t>0} \phi(t)$$

onde $\phi : [0, +\infty) \to \mathbb{R}$ é uma função de classe C^1 com $\phi'(0) < 0$. Mais precisamente, pretende-se localizar um ponto "aceitável" usando a regra de Armijo ou de Wolfe. Para localizar um ponto aceitável pode-se utilizar o algoritmo na tabela 1.

Comentários ao algoritmo da tabela 1:

- o algoritmo testa sucessivamente pontos até aprovar um (termina quando encontra um ponto "aceitável");

1.	**inicialização**	▶ escolher $t_0 > 0$
2.		▶ inicializar $e = 0$, $d = +\infty$, $k = 0$
3.	**ciclo**	▶ testar t_k:
4.		• se $t_k \in \mathcal{A}$: terminar
5.		• se $t_k \in \mathcal{D}$: $d = t_k$
6.		• se $t_k \in \mathcal{E}$: $e = t_k$
7.		▶ testar d:
8.		• se $d = +\infty$: escolher $t_{k+1} > t_k$
9.		• se $d < +\infty$: escolher $t_{k+1} \in (e, d)$
10.		▶ $k \leftarrow k + 1$ e retornar ao ciclo

Tabela 1: Algoritmo para localização de um ponto aceitável

- na linha 8 do algoritmo assuma (para simplificar) que se escolhe $t_{k+1} > t_k$ fazendo $t_{k+1} = \alpha\, t_k$ (onde $\alpha > 1$ é uma constante previamente escolhida);
- na linha 9 do algoritmo assuma (para simplificar) que se escolhe $t_{k+1} \in (e, d)$ fazendo $t_{k+1} = (e + d)/2$;
- as variáveis internas e e d definem um intervalo (e, d) que, à luz dos problemas 2 e 3, sabemos que contém um ponto “aceitável”;
- cada modificação em d (linha 5), resulta na diminuição de d e cada modificação em e (linha 6), resulta no aumento de e;
- portanto, cada modificação em e ou d, “aperta”o intervalo (e, d);
- em cada iteração k, é executada <u>apenas uma</u> das linhas 4, 5 ou 6;
- em cada iteração k, é executada <u>apenas uma</u> das linhas 8 ou 9.

O objectivo deste problema é provar que este algoritmo localiza um ponto aceitável num número finito de iterações, independentemente da regra (Armijo ou Wolfe) adoptada. Note que esta propriedade de terminação finita é crucial: no contexto dos algoritmos de optimização que operam por pesquisas em linha, este algoritmo de localização de pontos aceitáveis é invocado em cada iteração; se entrar em ciclo infinito, o algoritmo primário fica bloqueado.

Neste problema, assuma que ϕ é limitada inferiormente, ou seja, existe m tal que $\phi(t) \geq m$ para todo o $t \geq 0$.

(a) Comece por provar que a linha 8 não pode ser executada infinitamente. [**Pista:** suponha que a linha 8 é executada infinitamente. Observe que, necessariamente, a linha 5 nunca é executada. Logo, os pontos t_k gerados pelo algoritmo são sempre considerados demasiado "à esquerda". Mostre então que t_k pertence a $\mathcal{E}$ e verifica $t_k \to +\infty$. Conclua que $\phi(t_{m_k}) \to -\infty$, contradizendo assim o facto de ϕ ser limitada inferiormente.]

(b) Pela alínea anterior, sabemos que a linha 8 deixa de ser executada pelo algoritmo após um número K de iterações. Isto implica que é a linha 9 que passa a ser executada em cada iteração $k > K$. Suponha que o algoritmo corre infinitamente. Seja e_k e d_k o valor de e e d ao final da iteração k. Mostre que e_k é uma sucessão crescente, d_k é uma sucessão decrescente e ambas convergem para um mesmo ponto t^*. [**Pista:** cada alteração em e ou d "aperta" o intervalo (e, d) para metade do comprimento.] Mostre que

$$\frac{\phi(d_k) - \phi(t^*)}{d_k - t^*} \geq \sigma\phi'(0)$$

e, passando ao limite, conclua que $\phi'(t^*) \geq \sigma\phi'(0) > \lambda\phi'(0)$. Por outro lado, observe que $\phi'(e_k) \leq \lambda\phi'(0)$ o que, passando ao limite, resulta $\phi'(t^*) \leq \lambda\phi'(0)$ e gera uma contradição.

Problema 5.[Localização de pontos aceitáveis: parte II] Considere o algoritmo descrito na tabela 1.

(a) [Regra de Armijo] Considere a função

$$\phi(t) = 8t^5 - \frac{98}{3}t^4 + 42t^3 - \frac{95}{6}t^2 - 3t$$

cujo gráfico se apresenta na figura 1. Faça uma implementação em MATLAB do algoritmo na tabela 1, para a regra de Armijo. **Nota:** não se pretende aqui que utilize os resultados obtidos no problema 1 para a função ϕ; o seu programa em MATLAB deve testar cada ponto t_k usando as definições em (2).

Mostre a sequência de pontos t_k gerados pelo algoritmo quando usa **(i)** $\sigma = 0.4$ **(ii)** $\sigma = 0.6$ e **(iii)** $\sigma = 0.8$. Para qualquer dos casos, faça $t_0 = 1.5$ e $\alpha = 3$.

(b) [Regra de Wolfe] Para a mesma função ϕ da alínea anterior, faça uma implementação em MATLAB do algoritmo na tabela 1, utilizando agora a regra de Wolfe.

Mostre a sequência de pontos t_k gerados pelo algoritmo quando usa **(i)** $(\sigma, \lambda) = (0.4, 0.6)$ e $t_0 = 2$ **(ii)** $(\sigma, \lambda) = (0.4, 0.6)$ e $t_0 = 1.5$ **(iii)** $(\sigma, \lambda) =$

(0.6, 0.8) e $t_0 = 1.5$. Para qualquer dos casos, faça $\alpha = 3$.

Problema 6.[Conjuntos convexos: propriedades gerais] Um conjunto $S \in \mathbb{R}^n$ diz-se convexo se $(1-\lambda)x_1 + \lambda x_2 \in S$ para quaisquer $x_1, x_2 \in S$ e $\lambda \in [0,1]$. Ou seja, S é convexo se para quaisquer $x_1, x_2 \in S$ o segmento de recta que une x_1 a x_2 também pertence a S.

(a) Mostre que a intersecção de conjuntos convexos é um conjunto convexo.

(b) Prove ou exiba um contra-exemplo: a união de conjuntos convexos é um conjunto convexo.

(c) Suponha que S é um conjunto convexo, limitado e aberto em $\mathbb{R}$. Mostre que S é um intervalo (a,b). [**Pista:** mostre que $S = (m, M)$ onde $m = \inf S$ e $M = \sup S$]

(d) Mostre que $S \subset \mathbb{R}^n$ é convexo se e só se

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \in S,$$

para quaisquer $x_1, x_2, \ldots, x_k \in S$ e $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$, $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$. [**Pista**: para provar a necessidade da condição, use indução em k.]

Problema 7.[Conjuntos convexos elementares] Usando a definição, mostre que os seguintes conjuntos são convexos em $\mathbb{R}^n$:

(a) [Semi-espaço]

$$S = \left\{x \in \mathbb{R}^n \,:\, a^T x \leq b\right\},$$

onde $a \in \mathbb{R}^n$ $(a \neq 0)$ e $b \in \mathbb{R}$

(b) [Hiperplano]

$$S = \left\{x \in \mathbb{R}^n \,:\, a^T x = b\right\},$$

onde $a \in \mathbb{R}^n$ $(a \neq 0)$ e $b \in \mathbb{R}$

(c) [Bola]

$$S = \{x \in \mathbb{R}^n \,:\, \|x - x_0\| < R\},$$

onde $x_0 \in \mathbb{R}^n$ e $R > 0$

(d) [Cone]

$$S = \{a_1\mu_1 + a_2\mu_2 + \cdots + a_k\mu_k \,:\, \mu_1, \mu_2, \ldots, \mu_k \geq 0\} \subset \mathbb{R}^n,$$

onde $a_1, a_2, \ldots, a_k \in \mathbb{R}^n$. Esboçe o conjunto $S = \{a_1\mu_1 + a_2\mu_2 \,:\, \mu_1, \mu_2 \geq 0\}$ em $\mathbb{R}^2$, onde $a_1 = (1,1)$ e $a_2 = (0,1)$.

Problema 8.[Conjuntos convexos] Usando a definição, mostre que os seguintes conjuntos são convexos em $\mathbb{R}^{n \times n}$:

(a) [Matrizes triangulares inferiores com diagonal positiva]

$$S = \left\{ X \in \mathbb{R}^{n \times n} \,:\, X_{ij} = 0, \text{ para } 1 \leq i < j \leq n, \text{ e } X_{ii} > 0 \text{ para } i = 1, \ldots, n \right\}.$$

O símbolo X_{ij} representa a entrada na linha i, coluna j da matriz X.

(b) [Matrizes simétricas]

$$S = \left\{ X \in \mathbb{R}^{n \times n} \,:\, X = X^T \right\}$$

(c) [Matrizes definidas positivas]

$$S = \left\{ X \in \mathbb{R}^{n \times n} \,:\, X \succ 0 \right\}.$$

A notação $X \succ 0$ significa que X é simétrica e $y^T X y > 0$ para todo o $y \in \mathbb{R}^n$, $y \neq 0$.

(d) [Matrizes com entradas não-negativas e traço unitário]

$$S = \left\{ X \in \mathbb{R}^{n \times n} \,:\, X_{ij} \geq 0 \text{ e } \mathsf{tr}(X) = 1 \right\}.$$

Problema 9.[Funções convexas: propriedades gerais] Seja $S \in \mathbb{R}^n$ um conjunto convexo. Uma função $f : S \to \mathbb{R}$ diz-se convexa se

$$f\left((1-\lambda)x_1 + \lambda x_2\right) \leq (1-\lambda) f(x_1) + \lambda f(x_2),$$

para quaisquer $x_1, x_2 \in S$ e $\lambda \in [0,1]$.

(a) Seja $S \subset \mathbb{R}^n$ um conjunto convexo. Para $x_0 \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$ defina-se o conjunto $I_{x_0,d} = \{t \in \mathbb{R} : x_0 + td \in S\}$. Mostre que $I_{x_0,d}$ é um conjunto convexo em $\mathbb{R}$.

(b) Seja $S \subset \mathbb{R}^n$ um conjunto convexo e $f : S \to \mathbb{R}$. Mostre que f é convexa se e só se $\phi_{x_0,d} : I_{x_0,d} \to \mathbb{R}$,

$$\phi_{x_0,d}(t) = f(x_0 + td)$$

é uma função convexa (para qualquer $x_0 \in \mathbb{R}^n$ e $d \in \mathbb{R}^n$). [**Nota:** este resultado permite reduzir o teste de convexidade de funções em $\mathbb{R}^n$ ao teste de convexidade de funções em $\mathbb{R}$]

(c) Seja $S \subset \mathbb{R}^n$ um conjunto convexo e $f : S \to \mathbb{R}$. Mostre que f é convexa se e só se

$$f(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) + \cdots + \lambda_k f(x_k),$$

para quaisquer $x_1, x_2, \ldots, x_k \in S$ e $\lambda_1, \lambda_2, \ldots, \lambda_k \geq 0$, $\lambda_1 + \lambda_2 + \cdots + \lambda_k = 1$.

(d) Seja $S \subset \mathbb{R}^n$ um conjunto convexo e $f : S \to \mathbb{R}$. Mostre: se f é convexa, então $S_\alpha = \{x \in S : f(x) \leq \alpha\}$ é um conjunto convexo, para qualquer $\alpha \in \mathbb{R}$. O reverso é verdadeiro ?

(e) Seja $g : A \subset \mathbb{R}^n \to \mathbb{R}$ uma função definida num conjunto A. O epígrafo de g é o conjunto

$$\mathsf{epi}(g) = \left\{ (x,t) \in \mathbb{R}^{n+1} \,:\, x \in A,\, g(x) \leq t \right\}.$$

Seja $S \subset \mathbb{R}^n$ um conjunto convexo e $f : S \to \mathbb{R}$. Mostre que f é convexa se e só se $\mathsf{epi}(f)$ é um conjunto convexo em $\mathbb{R}^{n+1}$.

(f) Use a alínea anterior para determinar se as seguintes funções são convexas: **(i)** $f : \mathbb{R} \to \mathbb{R}$, $f(x) = \sin(x)$; **(ii)** $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$; **(iii)** $f : \mathbb{R}^+ \to \mathbb{R}$, $f(x) = \log(x)$; **(iv)** $f : \mathbb{R}^+ \to \mathbb{R}$, $f(x) = e^x$.