Convexity

Exercise 1. Let A be an $m \times n$ matrix, let $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Solve the optimization problem: minimize $c^T x$ subject to the constraints $Ax = b$.

Solução:

Primeiro, vamos excluir uma possibilidade trivial que é o caso $c = (0, \ldots, 0)$ que claramente possui mínimo igual a 0.

Das aulas de Álgebra Linear, sabemos há três possibilidades para o número de soluções do sistema $Ax = b$.

Se o sistema não tiver nenhuma solução, o conjunto de soluções (*Feasible Set*) é vazio e o problema de otimização também não tem solução (*infeasible*). Se o sistema tiver solução única, chamemos esta solução de x^*. Então valor ótimo do problema é $c^T x^*$.

Se o sistema possuir infinitas soluções, certamente o espaço nulo $N(A)$ não é trivial. Basta pegar duas soluções x_1 e x_2, ou seja, $Ax_1 = b$ e $Ax_2 = b$ e subtrair ambas as equações, obtendo $A(x_1 - x_2) = 0$. Logo $N(A)$ possui este elemento que é não-nulo pois $x_1 \neq x_2$.

Vou separar a questão em mais dois casos finais: c é ortogonal a $N(A)$ ou não. Isso pois podemos representar a solução agora como $x_p + x_N$, a soma de uma solução particular mais um elemento de $N(A)$. Como $N(A)$ é um espaço vetorial $x_N \in N(A) \Rightarrow \alpha x_N \in N(A)$ para $\alpha \in \mathbb{R}$. Se $c^T x_N \neq 0$ podemos tomar $\alpha = \pm\infty$ e um desses casos resultará num mínimo $-\infty$ (*Unbounded*).

Caso contrário, c é um elemento do espaço-linha de A. Sejam a_i as linhas de A. Então $c = \alpha_1 a_1 + \ldots \alpha_m a_m$. $c^T x$ fica pré determinado então. $c^T x = (\alpha_1 a_1 + \ldots \alpha_m a_m)x = \alpha_1 b_1 + \ldots \alpha_m b_m$, pois $Ax = b \Rightarrow a_i^T x = b_i$. Assim verificamos todos os casos para este problema de otimização.

Exercise 2. Show that the following sets are convex:

a) $\{\sum_{i=1}^{n} \lambda_i x_i : \lambda \geq 0, \sum_{i=1}^{n} \lambda_i = 1\}$ where $x_1, \ldots, x_n$ are given in $\mathbb{R}^n$.

b) $\{x \in \mathbb{R}^n : ||x - x_0|| \leq r\}$ where $r > 0$, $x_0 \in \mathbb{R}^n$, and $||.||$ is a norm on $\mathbb{R}^n$

c) $\{x \in \mathbb{R}^n : a^T x = b\}$ e $\{x \in \mathbb{R}^n : a^T x \leq b\}$ where $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

d) $\{x \in \mathbb{R}^n : Ax = b, Cx \leq d\}$

Solução:

Item a)

Tome dois elementos deste conjunto $a = \sum_{i=1}^{n} \lambda_i x_i$ e $b = \sum_{i=1}^{n} \kappa_i x_i$. Então para $0 \leq t \leq 1$

$$at + (1-t)b = \sum_{i=1}^{n} x_i(\lambda_i t + \kappa_i(1-t))$$

Como

$$\sum_{i=1}^{n} \lambda_i t + \kappa_i(1-t) = \sum_{i=1}^{n} \lambda_i t + \sum_{i=1}^{n} \kappa_i(1-t) = 1$$

a combinação convexa de dois elementos quaisquer pertence ao conjunto. Logo ele é convexo.

Item b)

Tome dois elementos x, y dentro desta bola. Queremos verificar se, para $0 < t < 1$, $xt + y(1-t)$ pertence à bola. Para isso

$$\begin{aligned} ||xt + y(1-t) - x_0|| &= ||xt + y(1-t) - x_0(t + 1 - t)|| \\ &= ||(x - x_0)t + (y - x_0)(1-t)|| \\ &\leq ||x - x_0||t + ||y - x_0||(1-t) \leq rt + r(1-t) = r, \end{aligned}$$

onde a primeira desigualde é a desigualdade triangular e a segunda pois x, y pertence à bola. O conjunto é convexo.

Item c)

Vou fazer apenas o caso ($\leq$) pois contempla a igualdade e as contas são análogas.

Tome dois elementos x, y neste conjunto e t tal que $0 \leq t \leq 1$. Então:

$$a^T(xt + y(1-t)) = (a^T x)t + (a^T y)(1-t) \leq bt + b(1-t) = b$$

A combinação convexa de dois elementos do conjunto pertence ao conjunto, portanto ele é convexo.

Item d)
Este conjunto é a interseção dos conjuntos

$$\bigcap_i \{x \in \mathbb{R}^n : a_i^T x = b_i\} \cap \bigcap_j \{x \in \mathbb{R}^n : c_j^T x \leq d_j\},$$

onde a_i são as linhas de A e c_j são as linhas de C. Provamos que cada conjunto desse é convexo no item anterior. Como a interseção de conjuntos convexos é convexo (será provado no próximo exercício) o conjunto resultante é convexo.

Exercise 3. Show that if a nonempty set is an intersection of convex sets then it is convex. Deduce that the set of semi-definite positive matrices is convex.

Seja $(A_\alpha)_{\alpha \in I}$ uma coleção arbitrária de conjuntos convexos. Tome dois elementos $x, y \in A = \bigcap_{\alpha \in I} A_\alpha$ (temos que esses elementos existem pois o conjunto é não vazio). Os elementos podem ser iguais. Para $0 \leq t \leq 1$ a combinação $x \cdot t + y \cdot (1 - t)$ pertence a A_α para todo $\alpha \in I$ pois cada um destes conjuntos é convexo. Pela definição da interseção, esta combinação também pertence a A. Logo A é convexo.

Para as matrizes semidefinidas positivas, vamos fixar um valor $x \in \mathbb{R}^n$. Considere os conjuntos $A_x = \{A \in \mathbb{R}_{n \times n} | x^T A x \geq 0\}$. Fazendo a combinação convexa de duas matrizes B, C neste conjunto obtemos

$$x^T(t \cdot B + (1 - t) \cdot C)x = t \cdot x^T(B)x + (1 - t) \cdot x^T(C)x \geq 0$$

Como a combinação convexa continua no conjunto, para cada x, A_x é convexo. Portanto, $S = \bigcap_x A_x$ também é convexo, e S é justamente o conjunto das matrizes semidefinidas positivas.

Exercise 4. Let $f : \mathbb{R}^n \to R$ be a function. Show that f is convex iff $g : \mathbb{R} \to \mathbb{R}$ given by $g(t) = f(x + td)$ is convex for every $x, d \in \mathbb{R}^n$. Deduce that the following functions are convex:

a) $f(X) = -\ln(\det(X))$ where X is a definite positive matrix.

b) $f(x) = x^T Ax + p^T x + r$ for $x \in \mathbb{R}^n$ where A is semi-definite positive

Preliminar

($\Rightarrow$)

Seja f uma função convexa e, fixando x e d, defina $g(t) = f(x + td)$. Queremos mostrar que, para $0 < \lambda < 1$ e a, b reais:

$$g(\lambda a + (1-\lambda)b) \leq \lambda g(a) + (1-\lambda)g(b)$$

Vamos trabalhar a expressão $g(\lambda a + (1-\lambda)b)$.

$$\begin{aligned} g(\lambda a + (1-\lambda)b) &= f(x + (\lambda a + (1-\lambda)b) \cdot d) \\ &= f(x + \lambda ad + (1-\lambda)bd) \\ &= f(x\lambda + x(1-\lambda) + \lambda ad + (1-\lambda)bd) \\ &= f(\lambda(x + ad) + (1-\lambda)(x + bd)) \\ &\leq \lambda f(x + ad) + (1-\lambda)f(x + bd) \\ &= \lambda g(a) + (1-\lambda)g(b) \end{aligned}$$

Onde a desigualdade vem da convexidade de f ∎

($\Leftarrow$)

Vamos abrir a condição de convexidade para um x e d quaisquer e fazer escolhas espertas desses valores para obter a convexidade de f.Então tome $0 < \lambda < 1$ e $a, b \in \mathbb{R}$. Temos que

$$\begin{aligned} g(\lambda a + (1-\lambda)b) &\leq \lambda g(a) + (1-\lambda)g(b) \\ f(x + (\lambda a + (1-\lambda)b)d) &\leq \lambda f(x + ad) + (1-\lambda)f(x + bd) \\ f(\lambda(x + ad) + (1-\lambda)(x + bd)) &\leq \lambda f(x + ad) + (1-\lambda)f(x + bd). \end{aligned}$$

Se conseguirmos que $x + ad$ e $x + bd$ sejam vetores $v, w \in \mathbb{R}^n$ quaisquer obtemos a convexidade de f. E de fato conseguimos isso, basta tomar, por exemplo, $a = 0, x = v, d = (w - v)/b$ e b um real qualquer não nulo ∎.

Item a)

Considere $g(t) = f(X + tD)$, X e D matrizes definidas positivas quaisquer.

Veja que

$$
\begin{aligned}
g(t) &= -\ln\det(X + tD) \\
&= -\ln\det(X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}}DX^{-\frac{1}{2}})X^{\frac{1}{2}}) \\
&= -\ln\det(I + tX^{-\frac{1}{2}}DX^{-\frac{1}{2}}) - \ln\det X
\end{aligned}
$$

Esses passos dependem da propriedade de determinantes $det(AB) = det(A)det(B)$ e de que matrizes positivas definidas possuem inversa definida positiva e possuem raiz quadrada definida positiva. Como o produto de matrizes definidas positivas é definido positivo, e matrizes definidas positivas possuem decomposição espectral com todos os autovalores reais positivos, podemos escrever $X^{-\frac{1}{2}}DX^{-\frac{1}{2}}$ como um produto $Q^T\Lambda Q$, onde Q é ortogonal e Λ uma diagonal contendo os autovalores. Lembrando que para matrizes ortogonais vale $I = Q^TQ$, podemos continuar as equações com

$$
\begin{aligned}
g(t) &= -\ln\det(I + Q^T\Lambda Q) - \ln\det X \\
&= -\ln\det(Q^TQ + Q^T\Lambda Q) - \ln\det X \\
&= -\ln\det Q^T(I + t\Lambda)Q - \ln\det X \\
&= -\ln\det Q^TQ(I + t\Lambda) - \ln\det X \\
&= -\ln\det(I + t\Lambda) - \ln\det X
\end{aligned}
$$

Como a matrix $I + t\Lambda$ é diagonal seu determinante é o produto dos elementos da diagonal.

$$
\begin{aligned}
g(t) &= -\ln\prod_{i=1}^{n}(1 + t\lambda_i) - \ln\det X \\
&= -\sum_{i=1}^{n}\ln(1 + t\lambda_i) - \ln\det X
\end{aligned}
$$

Como somar uma constante não altera convexidade, e soma de funções convexas é convexa, pela convexidade da função $-\ln$ finalizamos o resultado. g é convexa e basta agora usar a volta do resultado preliminar.

Item b)

Definindo $g(t) = f(x + td)$ obtemos

$$
\begin{aligned}
g(t) &= f(x + td) \\
&= (x + td)^TA(x + td) + p^T(x + td) + r \\
&= x^TAx + t^2d^TAd + p^T + tp^Td + r
\end{aligned}
$$

Então $g(t)$ é uma função quadrática em t, portanto convexa. Usando a volta do resultado preliminar ganhamos a convexidade de f.

Exercise 5. Show that the set of optimal solutions of a convex optimization problem is convex.

Seja f a função objetivo. Tome duas soluções de um problema de otimização convexa x, y, ou seja, $f(x)$ que é igual a $f(y)$ (caso contrário um deles seria menor, invalidando a otimiladade do outro) são menores ou iguais a qualquer ponto do conjunto viável, Seja $c = f(x) = f(y)$ o valor ótimo. Estes dois pontos pertencem a um conjunto viável X convexo, logo para $0 \leq t \leq 1$

$$tx + (1-t)y \in X.$$

Logo, pelo ótimo de $f(x)$ e $f(y)$

$$f(tx + (1-t)y) \leq f(x)$$
$$f(tx + (1-t)y) \leq f(y)$$

multiplicando a primeira inequação por t e a segunda por $(1-t)$ e somando ambas obtemos

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) = tc + (1-t)c = c.$$

Portanto $tx + (1-t)y$ também atinge o valor ótimo. O conjunto das soluções ótimas é fechado por combinações convexas, logo convexo.

Exercise 6. Consider the function $g : \mathbb{R} \to \mathbb{R}$ satisfying $g(x) \leq g(y)$ for $x \leq y$. If $f : \mathbb{R}^n \to \mathbb{R}$ and g are convex show that $g \circ f$ is convex.

Solução:

Tome $x, y \in \mathbb{R}$. Para $0 \leq t \leq 1$ temos que:

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y).$$

Logo

$$g(f(tx + (1-t)y)) \leq g(tf(x) + (1-t)f(y)).$$

Como g é convexa, temos que:

$$g(tf(x) + (1-t)f(y)) \leq tg(f(x)) + (1-t)g(f(y))$$

Juntando as últimas duas inequações temos a convexidade de $f \circ g$.

Exercise 7. Let A be a set of indices. Define

$$f(x) = \max_{\alpha \in A} f_\alpha(x)$$

where for all α, the function $f_\alpha : \mathbb{R}^n \to \mathbb{R}$ is convex. Show that f is convex. Deduce that the maximal singular value $\sigma(A) = \max_{||x||_2} ||Ax||_2$ of matrix A is a convex function of A.

Solução:

Tome $x, y \in \mathbb{R}^n$ e $0 \leq t \leq 1$. Temos que para todo α

$$f_\alpha(tx + (1 - ty)) \leq t f_\alpha(x) + (1 - t) f_\alpha(y) \leq t f(x) + (1 - t) f(y)$$

Como a desigualdade vale para todo $\alpha \in A$, vale também para o máximo de f para $\alpha \in A$ (caso contrário, seria possível encontrar α tal que a desigualdade não fosse valida), provando a convexidade de f.

Para mostrar a propriedade dos valores singulares basta tomar $f_\alpha : \mathbb{R}^{n \times n} \to \mathbb{R}$, $f_\alpha(A) = ||A\alpha||_2$, e o conjunto de α's os valores de $\mathbb{R}^n$ de norma 1 e usar a propriedade inicialmente demonstrada (a propriedade demonstrada foi feita para o $\mathbb{R}$ mas é direto estabelecer a relação entre $\mathbb{R}^{n^2}$ e $\mathbb{R}^{n \times n}$).

Exercise 8.

a) Show that if X is convex its interior is convex too.

b) We define the ε-enlargement of set X by

$$X^{\varepsilon} = \{y : \text{dist}(y, X) \leq \varepsilon\}$$

for $\varepsilon > 0$ where dist is the distance function:

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

Show that if X is convex then the set X^{ε} is convex too.

c) Let $B_F(0, \varepsilon) = \{x : ||x|| \leq \varepsilon\}$. Show that if X is convex and closed then $X^{\varepsilon} = X + B_F(0, \varepsilon)$. If X is closed and convex, deduce another proof of the convexity of X^{ε}.

Item a)

Tome $x, y \in \text{int}\, X$. Para x, como está no interior, existe $\varepsilon > 0$ tal que a bola aberta de centro em x e raio ε está contida em X. Seja $B(x, \varepsilon)$ esta bola. Tome também $0 \leq t \leq 1$ e considere o elemento $z = tx + (1-t)y$. Devemos provar que este elemento está no interior de X para mostrar que o interior é convexo, i.e, encontrar um raio r, tal que a bola de centro em z e raio r tal que esta bola esteja contida em X.

Afirmativa: $r = \theta\varepsilon$ satisfaz a condição acima, e para todo $z^* \in B(z, r)$ existe $x^* \in B(x, \varepsilon)$ tal que $z^* = x^*t + (1-t)y$.

Para conseguir construir uma intuição de porque fazer esta afirmação, pense que uma combinação convexa entre dois pontos são os pontos de uma "reta" que os liga (podemos estar em um espaço que a definição de reta não é a usual, mas a ideia é similar). Então para todos os pontos dentro de uma bola centrada em x, ligue-os em y. A figura formada vai ser como um cone. Como X é convexo, todas essas retas estão em X, logo o cone está contido em X. Note que a altura deste cone e a reta que liga x a y. Se tomarmos um ponto z qualquer nesta reta, podemos inscrever uma bola no cone com um centro neste ponto, mostrando que todo ponto da reta que liga x a y está no interior de X.

Vamos para a prova da afirmativa. Fixe $z^* \in B(z, r)$ e tome x^* tal que $z^* = x^*t + (1-t)y$. Então $x^* \in B(x, \varepsilon)$. Para isto basta ver que

$$\begin{aligned} ||t(x^* - x)|| &= ||tx^* - tx|| \\ &= ||z^* - (1-t)y - (z - (1-t)y)|| \\ &= ||z^* - z|| \\ &\leq r = t\varepsilon. \end{aligned}$$

Concluimos que deve valer $||t(x^* - x)|| \leq t\varepsilon \Rightarrow ||x^* - x|| \leq \varepsilon$. Como $x^* \in B(x, \varepsilon)$, temos que $x^* \in X$. Pela convexidade de X, $z^* \in X$. Como isso foi feito para um elemento arbitrário de $B(z, r)$, toda a bola está contida em X. Portanto, uma combinação convexa de pontos do interior está no interior. O interior de um convexo é convexo.

Item b)

Tome dois pontos $x_1, x_2 \in X^\varepsilon$. Então para $0 \leq t \leq 1$

$$\begin{aligned}
\mathrm{dist}(tx_1 + (1-t)x_2) &= \inf_{y \in X} ||tx_1 + (1-t)x_2 - y|| \\
&= \inf_{y \in X} ||tx_1 + (1-t)x_2 - ty_1 + (1-t)y_2|| \\
&= \inf_{y_1, y_2 \in X} ||t(x_1 - y_1) + (1-t)(x_2 - y_2)|| \\
&\leq \inf_{y_1, y_2 \in X} (t||x_1 - y_1|| + (1-t)||x_2 - y_2||) \\
&= t \inf_{y_1 \in X} ||x_1 - y_1|| + (1-t) \inf_{y_2 \in X} ||x_2 - y_2|| \\
&\leq \varepsilon. \blacksquare
\end{aligned}$$

Item c)

Se $x \in X^\varepsilon$ então para todo $\varepsilon_n = \frac{1}{n}$ é possível encontrar um ponto $x_n \in X$ tal que $||x - x_n|| \leq \varepsilon + \varepsilon_n$. Como X é fechado, esta sequencia de pontos converge para um ponto x_c em X e este ponto é tal que $||x - x_c|| \leq \varepsilon$. Então todo ponto de X^ε está em uma bola de raio ε com centro em um ponto de $X \quad \blacksquare$. Tanto a bola como X são conjuntos convexos. Como para X fechado a extensão X^ε pode ser escrita como a soma de conjuntos convexos(que é convexo), a extensão é convexa

Exercise 9. Let C be a convex set and let $\alpha_1, \alpha_2 > 0$. Show that

$$(\alpha_1 + \alpha_2)C = \alpha_1 C + \alpha_2 C.$$

Find a set C for which the above relation does not hold.

Solução:

Todo elemento de $S = \alpha_1 C + \alpha_2 C$ pode ser representado por $\alpha_1 x + \alpha_2 y$ para algum par x e y em C. Todo elemento de $T = (\alpha_1 + \alpha_2)C$ pode ser representado por $(\alpha_1 + \alpha_2)z$ para algum z em C. Para mostrar que existe uma equivalência entre estes dois conjuntos, basta ver que para todo elemento $\alpha_1 x + \alpha_2 y$ de S, podemos tomar

$$z = \frac{\alpha_1}{\alpha_1 + \alpha_2} x + \frac{\alpha_2}{\alpha_1 + \alpha_2} y$$

que pertence a C se x e y pertencem a C por convexidade ($\frac{\alpha_1}{\alpha_1+\alpha_2} + \frac{\alpha_2}{\alpha_1+\alpha_2} = 1$). Como $(\alpha_1+\alpha_2)z = \alpha_1 x+\alpha_2 y$, para todo elemento de S obtemos um elemento igual em T. Da mesma forma $(\alpha_1+\alpha_2)z = \alpha_1 z+\alpha_2 z$, tomando $x = z$ e $y = z$, vemos que todo elemento de T é um elemento de S.

Exercise 10. Let S, T be two nonempty convex sets in $\mathbb{R}^n$. Show that a separates S, T if and only if

$$\sup_{x\in S} a^T x \le \inf_{y\in T} a^T y$$

and

$$\inf_{x\in S} a^T x < \sup_{y\in T} a^T y.$$

Show that the separation is strong if and only if

$$\sup_{x\in S} a^T x < \inf_{y\in T} a^T y$$

Solução:

Suponha que a separa S, T, então existe c tal que

$$a^T x \le c, \forall x \in S$$

e

$$a^T y \ge c, \forall y \in T.$$

Podemos encadear as desigualdades para qualquer par x, y em S, T, obtendo que

$$a^T x \le a^T y, \quad \forall (x, y) \in S \times T.$$

obtendo como um exercíco simples de análise a primeira desigualdade $\sup_{x\in S} a^T x \le \inf_{y\in T} a^T y$ e uma versão não estrita da segunda $\inf_{x\in S} a^T x \le \sup_{y\in T} a^T y$. Afirmo que esta última não pode ser uma igualdade. Isto para não violar uma hipótese não dita no enunciado de separação de conjuntos que é eles serem disjuntos. Vou abreviar $\inf_{x\in A} a^T x$ para $\inf_A$ e o mesmo para o sup. Se $\inf_S = \sup_T$, como $\sup_S \le \inf_T$ e $\inf_S \le \sup_S$, temos que $\inf_S = \sup_T \le \sup_S \le \inf_T$. Então $\sup_T \le \inf_T \to \sup_T = \inf_T$. Chegamos a conclusão de que $\inf_S = \inf_T = \sup_S = \sup_T = c$. Então S e T fazem parte do mesmo hiperplano, que é o hiperplano separador. Isto não é de tudo um problema, a não ser que estejamos buscando uma separação *própria* dos conjuntos. Para isso o hiperplano separador não pode incluir S, T ao mesmo tempo, o que conclui a ida.

Para a volta basta tomar qualquer $c \in (\sup_S, \inf_T)$. Claramente vale que

$$\begin{aligned} a^T x &\le c, \quad \forall x \in S, \\ a^T y &\ge c, \quad \forall y \in T. \end{aligned}$$

Logo a separa os dois conjuntos.

Para a separação forte basta trocar a desigualdade por uma desigualdade estrita, todos os resultados são análogos.

Exercise 11. Show that the following functions are convex:

a) $f(x_1, \dots, x_n) = \ln(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$.

b) $\frac{1}{f}$ where f is concave and $0 < f(x) < +\infty$.

c) $f(x) = e^{\beta x^T A x}$ where $\beta \geq 0$ and A is positive semi-definite.

Item a)

Se a hessiana de uma função em várias dimensões for semidefinida positiva (supondo que podemos diferenciar a função questão duas vezes), ela é convexa. Seja $S = e^{x_1} + e^{x_2} + \dots e^{x_n}$. Então:

$$\begin{aligned}
\frac{\partial f}{\partial x_i} &= \frac{e^{x_i}}{S} \\
\frac{\partial^2 f}{\partial x_i^2} &= \frac{Se^{x_i} - e^{x_i}e^{x_i}}{S^2} \\
\frac{\partial^2 f}{\partial x_i x_j} &= -\frac{e^{x_i}e^{x_j}}{S^2}
\end{aligned}$$

Portanto a Hessiana pode ser dada por:

$$\nabla^2 f = \frac{1}{S^2}\left(\begin{bmatrix} Se^{x_1} & 0 & \dots & 0 \\ 0 & Se^{x_2} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & Se^{x_n} \end{bmatrix} - \begin{bmatrix} e^{x_1}e^{x_1} & e^{x_1}e^{x_2} & \dots & e^{x_1}e^{x_n} \\ e^{x_2}e^{x_1} & e^{x_2}e^{x_2} & \dots & e^{x_2}e^{x_n} \\ \vdots & & \ddots & \vdots \\ e^{x_n}e^{x_1} & e^{x_n}e^{x_2} & \dots & e^{x_n}e^{x_n} \end{bmatrix}\right)$$

Podemos compactar esta notação tomando $z = (e^{x_1}, \dots, e^{x_n})$ e $\mathbf{1} = (1, \dots, 1)$.

$$\nabla^2 f = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\mathbf{diag}(z) - zz^T\right)$$

Para verificar se a Hessiana é semidefinida positiva, tomemos $v \in \mathbb{R}^n$ qualquer.

$$\begin{aligned}
v^T \nabla^2 f v &= \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\mathbf{1}^T z\right) v^T \mathbf{diag}(z) v - (\langle v, z\rangle)^2\right) \\
&= \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\sum_{i=1}^n z_i\right)\left(\sum_{i=1}^n v_i^2 z_i\right) - \left(\sum_{i=1}^n v_i z_i\right)^2\right) \geq 0
\end{aligned}$$

A desigualdade vem de Cauchy-Schwarz.

Item b)

Como a função $\frac{1}{x}$ é convexa decrescente, $g(x) = -\frac{1}{x}$ é côncava crescente. Se f é côncava, podemos adaptar o Exercício 6 sobre composição de funções convexas para ter que $g \circ f$ é côncava. Portanto $-g \circ f = \frac{1}{f}$ é convexa.

Item c)

No exercício 4 mostramos a convexidade de funções quadraticas que possuem o coeficiente dominante positivo. Como e^x é uma função crescente e convexa, usando o Exercício 6 sobre composição de funções obtemos que com $f(x) = \beta x^T Ax$ e $g(x) = e^x$, temos $f \circ g$ convexo.

Exercise 12. Show that if f is differentiable and convex with

$$||\nabla f(y) - \nabla f(x)||_2 \leq M||y-x||_2 \forall x, y \in \mathbb{R}^n$$

then for all $x, y \in \mathbb{R}^n$

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{M}{2}||y-x||_2^2.$$

Solução

Defina $\varphi(t) = f(x + t(y-x))$. Esta função é diferenciável e temos $\varphi'(t) = \langle y - x, \nabla f(x + t(y-x))) \rangle$. Então

$$\begin{aligned}
\varphi'(t) - \varphi'(0) &= \langle y - x, \nabla f(x + t(y-x))) - \nabla f(x) \rangle \\
&\leq ||y-x|| ||\nabla f(x + t(y-x))) - \nabla f(x)|| \\
&\leq Mt||y-x||^2,
\end{aligned}$$

onde a primeira desigualdade vem de Cauchy-Schawrz e a segunda pela hipótese do enunciado. Então

$$\begin{aligned}
\int_0^1 \varphi'(1) - \varphi'(0) dt &= \varphi(1) - \varphi(0) - \varphi'(0) \\
&= f(y) - f(x) - \langle \nabla f(x), y - x \rangle \\
&\leq \frac{M}{2}||y-x||^2 \blacksquare
\end{aligned}$$

Exercise 13. Show that if f is differentiable and convex with

$$||\nabla f(y) - \nabla f(x)||_2 \leq M||y - x||_2 \forall x, y \in \mathbb{R}^n$$

then for all $x, y \in \mathbb{R}^n$

$$\langle y - x, \nabla f(y) - \nabla f(x) \rangle \geq \frac{1}{M}||\nabla f(y) - \nabla f(x)||_2^2.$$