Ejercicios — Funciones convexas

Universidad Nacional de Colombia, Sede Bogotá
Introducción a la Optimización
2022-I

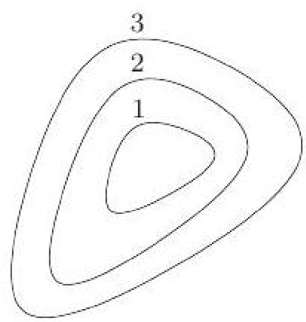
Paulina Castillo
ancastillov@unal.edu.co

Lina Simbaqueba
lmsimbaquebam@unal.edu.co

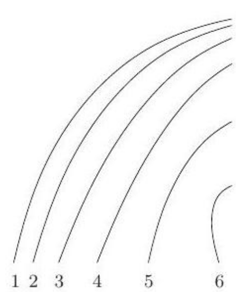
Sebastian Molina
smolinad@unal.edu.co

Exercises

3.2 Level sets of convex, concave, quasiconvex, and quasiconcave functions. Some level sets of a function f are shown below. The curve labeled 1 shows $\{x \mid f(x) = 1\}$, etc.

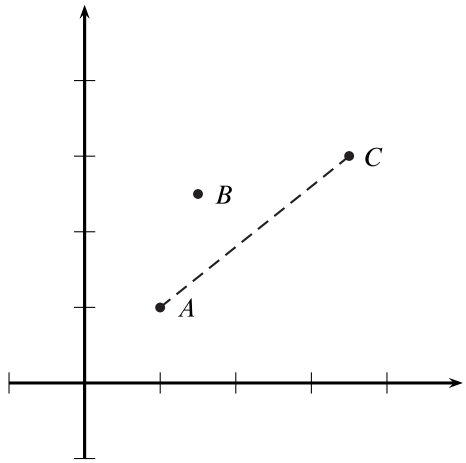


Could f be convex (concave, quasiconvex, quasiconcave)? Explain your answer. Repeat for the level curves shown below.



Solución.

La primera imagen es claramente cuasiconvexa, pero no es necesariamente convexa. Esto dado que el crecimiento en los ”lados” no es parejo, lo cual si lo graficamos en un plano con una vista perpendicular a las trazas obtenemos



Para la segunda no se puede decir mucho, pero podríamos caracterizarla en una función cóncava o cuasicóncava, no podría ser cuasiconvexa ya que las curvas de nivel no son convexas

□

3.4 Show that a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every line segment, its average value on the segment is less than or equal to the average of its values at the endpoints of the segment: For every $x, y \in \mathbb{R}^n$,

$$\int_0^1 f(x + \lambda(y - x))d\lambda \leq \frac{f(x) + f(y)}{2}.$$

Solución.

Primero supongamos que la función f es convexa. Por la desigualdad de Jensen's se concluye que, para todo $x, y \in \text{dom}(f)$ y para $\lambda \in [0, 1]$, se cumple que:

$$f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x))$$

Integramos, desde 0 a 1 con respecto a λ a ambos lados de la expresión anterior:

$$\begin{aligned}
\int_0^1 f(x + \lambda(y - x))d\lambda &\leq \int_0^1 f(x) + \lambda(f(y) - f(x))d\lambda \\
&= |f(x)\lambda|_0^1 + \left|\big(f(y) - f(x)\big)\frac{\lambda^2}{2}\right|_0^1 \\
&= \frac{f(x) + f(y)}{2}
\end{aligned}$$

Así, queda demostrado que si la función f es convexa, el valor promedio en un segmento es menor o igual que el promedio de los valores en los puntos extremos del segmento. Recíprocamente, supongamos que f no es convexa. Entonces, existen $x, y \in \text{dom}(f)$ y existe $\theta_0 \in (0, 1)$ tal que:

$$f(x) + \theta_0(f(y) - f(x)) < f(x + \theta_0(y - x))$$

Es decir, la función continua definida como $F(\theta) = f(x + \theta(y - x)) - f(x) - \theta(f(y) - f(x))$ es positiva en θ_0. Además, dado que $F(0) = F(1) = 0$; por la continuidad de F sabemos que existe un intervalo $(\alpha, \beta) \subseteq (0, 1)$ dónde F es positiva; además, $\theta_0 \in (\alpha, \beta)$. Si tomamos $u = \alpha x + (1 - \alpha)y$ y $v = \beta x + (1 - \beta)y$, tendremos que, para todo $\theta \in (0, 1)$ se cumple al siguiente desigualdad:

$$f(\theta u + (1 - \theta)v) > \theta f(u) + (1 - \theta)f(v)$$

O equivalentemente, tendremos que:

$$f(v + \theta(u - v)) > f(v) + \theta(f(u) - f(v))$$

Integrando la expresión a ambos lados en función de θ desde 0 hasta 1, tendremos un caso análogo al desarrollado en la primera parte de la demostración y se concluye que:

$$\int_0^1 f(u + \lambda(v - u))d\lambda \geq \frac{f(u) + f(v)}{2}$$

Y por tanto, si f no es convexa, existen puntos u y v en el valor promedio de la función en el segmento es mayor que el promedio de los valores en los puntos extremos del segmento.

□

3.6 *Functions and epigraphs.* When is the epigraph of a function a halfspace? When is the epigraph of a function a convex cone?

Solución.

Para que sea un subespacio es suficiente que la función sea afín, mientras que para que sea un cono convexo es necesario que la función f cumpla que $f(\alpha x) = \alpha f(x)$ para cualquier α positivo.

□

3.13 *Kullback-Leibler divergence and the information inequality.* Let D_{kl} be the Kullback Leibler divergence, as defined in (3.17). Prove the information inequality: $D_{\mathrm{kl}}(u, v) \geq 0$ for all $u, v \in \mathbb{R}^n_{++}$. Also show that $D_{\mathrm{kl}}(u, v) = 0$ if and only if $u = v$. Hint. The Kullback-Leibler divergence can be expressed as

$$D_{\mathrm{kl}}(u, v) = f(u) - f(v) - \nabla f(v)^T (u - v),$$

where $f(v) = \sum_{i=1}^n v_i \log v_i$ is the negative entropy of v.

Solución.

Observe que

$$\begin{aligned}
f(u) - f(v) - \nabla f(v)^T (u - v) &= \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n v_i \log v_i - \sum_{i=1}^n (\log v_i + 1)(u_i - v_i) \\
&= \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n v_i \log v_i - \sum_{i=1}^n (u_i \log v_i - v_i log v_i + u_i - v_i) \\
&= \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n u_i \log v_i - \sum_{i=1}^n u_i - v_i \\
&= \sum_{i=1}^n u_i \log \frac{u_i}{v_i} - \sum_{i=1}^n u_i - v_i \\
&= \sum_{i=1}^n \left(u_i \log \frac{u_i}{v_i} - u_i + v_i \right) \\
&= D_{\mathrm{kl}}(u, v)
\end{aligned}$$

Como $f(u)$ es convexa y diferenciable en $\mathbb{R}^n_{++}$, entonces $f(u) > f(v) + \nabla f(v)^T (u - v)$ para $u, v \in \mathbb{R}^n_{++}$, con $u \neq n$. Luego

$$\sum_{i=1}^n u_i \log u_i > \sum_{i=1}^n u_i \log v_i + (u_i - v_i) \quad \therefore \quad D_{\mathrm{kl}} > 0.$$

Claramente, si $u = v$ entonces $D_{\mathrm{kl}} = 0$. □

3.16 For each of the following functions determine whether it is convex, concave, quasiconvex, or quasiconcave.

a) $f(x) = e^x - 1$ on $\mathbb{R}$.

Solución.
Es convexa por la condición de segundo orden, para todo $x \in \mathbb{R}$:

$$f'(x) = e^x = f''(x) > 0$$

Como es convexa, es cuasiconvexa. No puede ser cóncava. Sin embargo, si es cuasicóncava. Para ver que es cuasicóncava, basta observar que el conjunto $\{x \in \mathbb{R} | e^x - 1 \leq \alpha\}$ es un conjunto convexo para todo $\alpha \in \mathbb{R}$.

□

b) $f\left(x_1, x_2\right) = x_1 x_2$ on $\mathbb{R}^2_{++}$.

Solución.
Para ver si es cóncava o convexa, utilicemos la condición de segundo orden:

$$\nabla f(x_1, x_2) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$$

Y, por tanto, tendremos que el Hessiano equivale a:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Se sabe que la matriz anterior no es semidefinida positiva ni semidefinida negativa ya que, si tomamos $y \in \mathbb{R}^2$, y $x \in \mathbb{R}^2_{++}$:

$$y^T \nabla^2 f(x_1, x_2) y = \begin{pmatrix} y_1 & y_2 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = 2y_1 y_2,$$

y $2y_1 y_2$ puede ser mayor o menor que cero dependiendo de los valores de y. En conclusión, la función $f(x_1, x_2)$ no es una función cóncava ni convexa. Para ver que la función en cuasicóncava, basta probar que $g(x_1, x_2) = -f(x_1, x_2)$ es cuasiconvexa, para ello, utilizaremos la condición de segundo orden sobre $g(x_1, x_2)$. Sea $x = (x_1, x_2) \in \mathbb{R}^2_{++}$ y $y = (y_1, y_2) \in \mathbb{R}$, la función g es cuasiconvexa si $y^T \nabla g(x) = 0$ implica que $y^T \nabla^2 g(x) y \geq 0$. Si se da el caso en que $y^T \nabla g(x) = 0$, tendremos que $-y_1 x_2 - y_2 x_1 = 0$ y esto implica que $-\frac{y_1}{y_2} = \frac{x_1}{x_2}$. Como x_1 y x_2 son ambos positivos, se concluye que y_1 y y_2 deben tener signos contrarios. Ahora si suponemos que y_1 y y_2 tienen signos contrarios, $y^T \nabla^2 g(x) y = -2y_1 y_2 \geq 0$, lo cuál satisface la condición de segundo orden y, por tanto, la función g es cuasiconvexa. Así, se concluye que f es cuasicóncava.

□

c) $f\left(x_1, x_2\right) = 1/\left(x_1 x_2\right)$ on $\mathbb{R}^2_{++}$.

Solución.
Utilicemos la condición de segundo orden:

$$\nabla f(x_1, x_2) = \begin{pmatrix} -\frac{1}{x_2 x_1^2} \\ -\frac{1}{x_1 x_2^2} \end{pmatrix}$$

El Hessiano equivale a:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \frac{2}{x_2 x_1^3} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_2^2 x_1^2} & \frac{2}{x_1 x_2^3} \end{pmatrix} = \frac{1}{x_1 x_2} \begin{pmatrix} \frac{2}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^2} \end{pmatrix}$$

Veamos que $\nabla^2 f(x_1, x_2)$ es definida semipositiva y, por tanto, la función f es convexa. Sea $(x, y) \in \mathbb{R}^2$.

$$
\begin{aligned}
\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{2}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{2x}{x_1^2} + \frac{y}{x_1 x_2} & \frac{x}{x_1 x_2} + \frac{2y}{x_2^2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \frac{2x^2}{x_1^2} + \frac{2xy}{x_1 x_2} + \frac{2y^2}{x_2^2} \\
&= \frac{2x^2 x_2^2 + 2x_1 x_2 xy + 2x_1^2 y^2}{x_1^2 x_2^2} \\
&\geq \frac{x^2 x_2^2 + 2x_1 x_2 xy + x_1^2 y^2}{x_1^2 x_2^2} \\
&= \frac{(x x_2 + x_1 y)^2}{x_1^2 x_2^2} \geq 0
\end{aligned}
$$

Lo anterior implica que f es convexa y, por tanto cuasiconvexa. No ess cuasicóncavo porque el conjunto $\{(x_1, x_2) \in \mathbb{R}^2_{++} | \frac{1}{x_1 x_2} \geq \alpha\}$ no es un conjunto convexo ya que la función $\frac{1}{x} - \alpha$ no es convexa en el primer cuadrante si $\alpha > 0$ □

d) $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$.

Solución.
Para estudiar la convexidad o concavidad, calculamos el Hessiano:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{1}{x_2} \\ -\frac{x_1}{x_2^2} \end{pmatrix}$$

Así, el Hessiano corresponde a la expresión:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{pmatrix}$$

La cuál no es semidefinida positiva ni semidefinida negativa:

$$
\begin{aligned}
\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} -\frac{y}{x_2^2} & -\frac{x}{x_2^2} + \frac{2yx_1}{x_2^3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= -\frac{xy}{x_2^2} - \frac{xy}{x_2^2} + \frac{2y^2 x_1}{x_2^3} \\
&= -\frac{2xy}{x_2^2} + \frac{2y^2 x_1}{x_2^3}
\end{aligned}
$$

Dado que la expresión anterior es positiva si $yx_1 > xx_2$ y negativa en caso contrario, se concluye que la función no es cóncava ni convexa. Los conjuntos de nivel $\{(x_1, x_2) \in \mathbb{R}^2_{++} | \frac{x_1}{x_2} \leq \alpha\}$ y $\{(x_1, x_2) \in \mathbb{R}^2_{++} | \frac{x_1}{x_2} \geq \alpha\}$ son convexos en el primer cuadrante ya que son semiespacios.

□

e) $f(x_1, x_2) = x_1^2/x_2$ on $\mathbb{R} \times \mathbb{R}_{++}$.

Solución.
Calculamos el Hessiano de la función:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \frac{2x_1}{x_2} \\ -\frac{x_1^2}{x_2^2} \end{pmatrix}$$

El cuál corresponde a la expresión:

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{pmatrix}$$

A continuación vemos que la matriz es demidefinida positiva:

$$\begin{aligned}
\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} \frac{2x}{x_2} - \frac{2x_1 y}{x_2^2} & -\frac{2x_1 x}{x_2^2} + \frac{2yx_1^2}{x_2^3} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= \frac{2x^2}{x_2} - \frac{2x_1 xy}{x_2^2} - \frac{2x_1 xy}{x_2^2} + \frac{2y^2 x_1^2}{x_2^3} \\
&= \frac{2x^2 x_2^2 - 4x_1 xyx_2 + 2y^2 x_1^2}{x_2^3} \\
&= \frac{2(xx_2 - yx_1)^2}{x_2^3} \geq 0
\end{aligned}$$

Así, f es convexa y, por tanto, cuasiconvexa.

Por otro lado, la función no es cóncava y tampoco es cuasicóncava ya que el conjunto $\{(x_1, x_2) | \frac{x_1^2}{x_2} \geq \alpha$ no es un conjunto convexo. □

f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \leq \alpha \leq 1$, on $\mathbb{R}^2_{++}$.

Solución.
Estudiemos el Hessiano de f:

$$\nabla f(x_1, x_2) = \begin{pmatrix} \alpha x_1^{\alpha-1} x_2^{1-\alpha} \\ (1-\alpha) x_2^{-\alpha} x_1^\alpha \end{pmatrix}$$

El Hessiano equivale a:

$$\begin{aligned}
\nabla^2 f(x_1, x_2) &= \begin{pmatrix} \alpha(\alpha-1) x_1^{\alpha-2} x_2^{1-\alpha} & (1-\alpha)\alpha x_2^{-\alpha} x_1^{\alpha-1} \\ \alpha(1-\alpha) x_1^{\alpha-1} x_2^{-\alpha} & (1-\alpha)(-\alpha) x_2^{-\alpha-1} x_1^\alpha \end{pmatrix} \\
&= \alpha(1-\alpha) x_1^\alpha x_2^{1-\alpha} \begin{pmatrix} -x_1^{-2} & x_1^{-1} x_2^{-1} \\ x_1^{-1} x_2^{-1} & -x_2^{-2} \end{pmatrix}
\end{aligned}$$

Veamos que $\nabla^2 f$ es definida seminegativa.

$$\begin{aligned}
\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} -x_1^{-2} & x_1^{-1} x_2^{-1} \\ x_1^{-1} x_2^{-1} & -x_2^{-2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} &= \begin{pmatrix} -xx_1^{-2} + yx_1^{-1} x_2^{-1} & xx_1^{-1} x_2^{-1} - yx_2^{-2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\
&= -x^2 x_1^{-2} + 2xyx_1^{-1} x_2^{-1} - y^2 x_2^{-2} \\
&= -(xx_1^{-1} - yx_2^{-1})^2 \leq 0
\end{aligned}$$

Con lo anterior se concluye que la función es cóncava y, por tanto, cuasicóncava. Además, la función es cuasiconvexa sólo si $\alpha = 0, 1$ ya que el conjunto $\{(x_1, x_2) | x_1^\alpha x_2^{1-\alpha} \leq t\}$ es un conjunto convexo en este caso debido a que corresponde a un semiespacio.

□

3.20 *Composition with an affine function.* Show that the following functions $f : \mathbb{R}^n \to \mathbb{R}$ are convex.

a) $f(x) = \|Ax - b\|$, where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $\|\cdot\|$ is a norm on $\mathbb{R}^m$.

Solución.
$f(x) = (g \circ h)(x)$, donde $h(x) = Ax + b$ es una función afín —y por lo tanto convexa— y $g(x) = \|x\|$ es una función convexa. Luego f es una composición de funciones convexas.

□

b) $f(x) = -(\det(A_0 + x_1A_1 + \cdots + x_nA_n))^{1/m}$, on $\{x \mid A_0 + x_1A_1 + \cdots + x_nA_n \succ 0\}$, where $A_i \in \mathbf{S}^m$.

Solución.
Por el **Ejercicio 3.18** del texto guía, la función $h(x) = (\det(x))^{1/m}$ es cóncava, luego $-(\det(x))^{1/m}$ es convexa. Por tanto la composición de $-h$ con una función afín, como lo es $\sum_{i=0} A_i x_i$, resulta en una función convexa.

$\square$

c) $f(X) = \operatorname{tr}(A_0 + x_1A_1 + \cdots + x_nA_n)^{-1}$, on $\{x \mid A_0 + x_1A_1 + \cdots + x_nA_n \succ 0\}$, where $A_i \in \mathbf{S}^m$. (Use the fact that $\operatorname{tr}\left(X^{-1}\right)$ is convex on $\mathbf{S}^m_{++}$; see exercise 3.18).

Solución.
La función traza, luego, de manera análoga al literal anterior esta composición con una función afín resulta convexa.

$\square$

3.21 *Pointwise maximum and supremum.* Show that the following functions $f : \mathbb{R}^n \to \mathbb{R}$ are convex.

a) $f(x) = \max_{i=1,\dots,k} \left\|A^{(i)}x - b^{(i)}\right\|$, where $A^{(i)} \in \mathbb{R}^{m\times n}$, $b^{(i)} \in \mathbb{R}^m$ and $\|\cdot\|$ is a norm on $\mathbb{R}^m$.

Solución.
Observe que f es la función de máximo punto a punto de k funciones de la forma $g_i(x) = \|A^{(i)}x - b^{(i)}\|$. Sin embargo, cada función g_i es la composición de las funciones convexas $A^{(i)}x + b^{(i)}$ —una función afín— y la norma $\|\cdot\|$, la cual también es convexa. $\square$

b) $f(x) = \sum_{i=1}^r |x|_{[i]}$ on $\mathbb{R}^n$, where $|x|$ denotes the vector with $|x|_i = |x_i|$ (i. *e.*, $|x|$ is the absolute value of x, componentwise), and $|x|_{[i]}$ is the i th largest component of $|x|$. In other words, $|x|_{[1]}, |x|_{[2]}, \dots, |x|_{[n]}$ are the absolute values of the components of x, sorted in nonincreasing order.

Solución.
Observe que

$$f(x) = \sum_{i=1}^{r} |x|_{[i]} = \max_{1\le i_j \le n} \sum_{j=1}^{r} |x_{i_j}|,$$

por lo que f es la función de máximo punto a punto de $\binom{n}{r}$ funciones convexas, definidas por la selección de las r normas $\{|x_{i_j}|\}_{1\le j\le r}$ del vector $|x|$.

$\square$

3.22 *Composition rules.* Show that the following functions are convex.

a) $f(x) = -\log\left(-\log\left(\sum_{i=1}^m e^{a_i^T x + b_i}\right)\right)$ on $\operatorname{dom} f = \left\{x \mid \sum_{i=1}^m e^{a_i^T x+b_i} < 1\right\}$. You can use the fact that $\log\left(\sum_{i=1}^n e^{y_i}\right)$ is convex.

Solución.
$f(x)$ es una función convexa. Para probarlo, basta observar que $\log\left(\sum_{i=1}^m e^{a_i^T x+b_i}\right)$ es una función convexa porque es la composición de una función convexa: $\log\left(\sum_{i=1}^n e^{y_i}\right)$, con una función afín: $a_i^T x + b_i$. Esto implica que $-\log\left(\sum_{i=1}^m e^{a_i^T x+b_i}\right)$ es cóncava. Finalmente, al ser $-\log(x)$ una función convexa no creciente, tenemos que $f(x)$ es la composición de una función convexa no creciente con una función cóncava y, por tanto, $f(x)$ es convexa. $\square$

b) $f(x, u, v) = -\sqrt{uv - x^T x}$ on $\operatorname{dom} f = \left\{(x, u, v) \mid uv > x^T x, u, v > 0\right\}$. Use the fact that $x^T x/u$ is convex in (x, u) for $u > 0$, and that $-\sqrt{x_1x_2}$ is convex on $\mathbb{R}^2_{++}$.

Solución.
Dado que la función $-v$ es convexa, y la función $\frac{x^T x}{u}$ es convexa, suma de convexas es convexas y, así, la función $v - \frac{x^T x}{u}$ es cóncava. Adicionalmente, la función u es una función cóncava. Dado que la función $-\sqrt{x_1x_2}$ es convexa y no creciente en cada argumento, composición de convexa con cóncava es convexa y, por tanto, $f(x, u, v) = -\sqrt{u \cdot \left(v - \frac{x^T x}{u}\right)}$ es una función convexa. $\square$

c) $f(x, u, v) = -\log\left(uv - x^T x\right)$ on $\operatorname{dom} f = \left\{(x, u, v) \mid uv > x^T x, u, v > 0\right\}$.

Solución.
Análogo al anterior, la función $-\log(x_1 x_2)$ es una función convexa sobre $\mathbb{R}^2_{++}$ y no creciente sobre cada argumento; como se mostró anteriormente, u es cóncava y $v - \frac{x^T x}{u}$ es cóncava, en conclusión, la función $f(x,u,v) = -\log(u \cdot \left(v - \frac{x^T x}{u}\right)$ es convexa al ser composición de convexa no creciente con cóncava. $\square$

d) $f(x,t) = -\left(t^p - \|x\|_p^p\right)^{1/p}$ where $p > 1$ and $\mathbf{dom}\, f = \left\{(x,t) \mid t \geq \|x\|_p\right\}$. You can use the fact that $\|x\|_p^p / u^{p-1}$ is convex in (x,u) for $u > 0$ (see exercise 3.23), and that $-x^{1/p} y^{1-1/p}$ is convex on $\mathbb{R}^2_+$ (see exercise 3.16).

Solución.
Basta observar que $t^p - \|x\|_p^p = t^p u^{1-p} - \frac{\|x\|_p^p}{u^{p-1}}$. Dado que $-t^p u^{1-p}$ y $\frac{\|x\|_p^p}{u^{p-1}}$ son funciones convexas, la suma de ellas es convexa y, se concluye que $t^p - \|x\|_p^p$ es cóncava. Dado que $-(x)^{1/p}$ es una función convexa no creciente, tenemos que $f(x,t)$ es convexa. $\square$

e) $f(x,t) = -\log\left(t^p - \|x\|_p^p\right)$ where $p > 1$ and $\mathbf{dom}\, f = \left\{(x,t) \mid t > \|x\|_p\right\}$. You can use the fact that $\|x\|_p^p / u^{p-1}$ is convex in (x,u) for $u > 0$ (see exercise 3.23).

Solución.
Basta observar que $t^p - \|x\|_p^p = t^p u^{1-p} - \frac{\|x\|_p^p}{u^{p-1}}$. Dado que $-t^p u^{1-p}$ y $\frac{\|x\|_p^p}{u^{p-1}}$ son funciones convexas, la suma de ellas es convexa y, se concluye que $t^p - \|x\|_p^p$ es cóncava. Dado que $-\log(x)$ es una función convexa no creciente, tenemos que $f(x,t)$ es convexa. $\square$

3.25 *Maximum probability distance between distributions.* Let $p, q \in \mathbb{R}^n$ represent two probability distributions on $\{1,\ldots,n\}$ (so $p, q \succeq 0$, $\mathbf{1}^T p = \mathbf{1}^T q = 1$). We define the maximum probability distance $d_{\mathrm{mp}}(p,q)$ between p and q as the maximum difference in probability assigned by p and q, over all events:

$$d_{\mathrm{mp}}(p,q) = \max\{|\,\mathbf{prob}(p,C) - \mathbf{prob}(q,C)| \mid C \subseteq \{1,\ldots,n\}\}.$$

Here $\mathbf{prob}(p,C)$ is the probability of C, under the distribution p, i.e., $\mathbf{prob}(p,C) = \sum_{i \in C} p_i$. Find a simple expression for d_{mp}, involving $\|p-q\|_1 = \sum_{i=1}^n |p_i - q_i|$, and show that d_{mp} is a convex function on $\mathbb{R}^n \times \mathbb{R}^n$. (Its domain is $\left\{(p,q) \mid p, q \succeq 0, \mathbf{1}^T p = \mathbf{1}^T q = 1\right\}$, but it has a natural extension to all of $\mathbb{R}^n \times \mathbb{R}^n$).

Solución.

$$\mathrm{prob}(p,C) - \mathrm{prob}(q,C) = -(\mathrm{prob}(p,\tilde{C}) - \mathrm{prob}(q,\tilde{C})),$$

Con $\dot{C} = \{1,\ldots,n\} \backslash C$, por lo que la función d_{mp} puede reescribirse como

$$d_{\mathrm{mp}}(p,q) = \max\{\mathrm{prob}(p,C) - \mathrm{prob}(q,C) \mid C \subseteq \{1,\ldots,n\}\}.$$

Siendo así d_{mp} convexa. Buscaremos un C que maximice la función

$$\mathrm{prob}(p,C) - \mathrm{prob}(q,C) = \sum_{i \in C} (p_i - q_i)$$

Un posible candidato es

$$C^\star = \{i \in \{1,\ldots,n\} \mid p_i > q_i\}.$$

Sin perdida de generalidad asumimos $p_{>}q_i$ o $p_i < q_i$. Supongamos existe otro C. Si existe un elemento k in $C^\star$pero no en C, añadimos k a C e incrementamos $\mathrm{prob}(p,C) - \mathrm{prob}(q,C)$ ya que $p_k - q_k > 0$, así C podría no se optimo. Por otro lado, sea $k \in C \backslash C^\star$, así $p_k - q_k < 0$. Elimando k de C, incrementamos $\mathrm{prob}(p,C) - \mathrm{prob}(q,C)$ by $q_k - p_k > 0$, Estando en el mismo punto. Así, tenemos $d_{\mathrm{mp}}(p,q) = \sum_{p_i > q_i} (p_i - q_i)$. Expresemos esto es términos de $\|p-q\|_1$. A partir

$$\sum_{p_i > q_i} (p_i - q_i) + \sum_{p_i \leq q_i} (p_i - q_i) = \mathbf{1}^T p - \mathbf{1}^T q = 0,$$

Tenemos

$$\sum_{p_i > q_i} (p_i - q_i) = -\left(\sum_{p_i \leq q_i} (p_i - q_i)\right)$$

$$\begin{aligned}
d_{\mathrm{mp}}(p,q) &= (1/2)\sum_{p_i>q_i}(p_i-q_i)-(1/2)\sum_{p_i\le q_i}(p_i-q_i)\\
&= (1/2)\sum_{i=1}^{n}|p_i-q_i|\\
&= (1/2)\|p-q\|_1.
\end{aligned}$$

Siendo así la función convexa.

$\square$

3.29 *Representation of piecewise-linear convex functions.* A convex function $f:\mathbb{R}^n\to\mathbb{R}$, with $\operatorname{dom} f=\mathbb{R}^n$, is called piecewise-linear if there exists a partition of $\mathbb{R}^n$ as

$$\mathbb{R}^n = X_1\cup X_2\cup\cdots\cup X_L,$$

where $\mathbf{int}\, X_i\neq\emptyset$ and $\mathbf{int}\, X_i\cap\mathbf{int}\, X_j=\emptyset$ for $i\neq j$, and a family of affine functions $a_1^Tx+b_1,\ldots,a_L^Tx+b_L$ such that $f(x)=a_i^Tx+b_i$ for $x\in X_i$. Show that such a function has the form $f(x)=\max\{a_1^Tx+b_1,\ldots,a_L^Tx+b_L\}$.

Solución.

Sean $x,y\in\operatorname{dom} f$. Por la desigualdad de Jensen se tiene que

$$f(tx+(1-t)y)=f(tx+y-ty)=f\big(t(x-y)+y\big)\le f(y)+t\big(f(x)-f(y)\big).$$

Por lo tanto, $\frac{f\left(t(x-y)+y\right)-f(y)}{t}+f(y)\le f(x)$. Sea $x\in X_i$, $y\in X_j$ y $t\in[0,1]$ tal que $x+t(x-y)\in X_j$. Por lo tanto, al aplicar la desigualdad anterior a la familia de funciones afines, se tiene que

$$\begin{aligned}
\frac{a_j^T\big(t(x-y)+y\big)+b_j-(a_j^Ty+b_j)}{t}+a_j^Ty+b_j &= \frac{ta_j^Tx-ta_j^Ty+a_j^Ty+b_j-(a_j^Ty+b_j)}{t}+a_j^Ty+b_j\\
&= a_j^Tx-a_j^Ty+a_j^Ty+b_j\\
&= a_j^Tx+b_j\\
&\le a_i^Tx+b_i.
\end{aligned}$$

Como lo anterior se cumple para culaquier X_j arbitrario, obtenemos que $a_i^Tx+b_i\ge\max_{j\in\{1,\ldots,L\}}a_j^Tx+b_j$, por lo que para $x\in X_i$, $f(x)=\max_{j\in\{1,\ldots,L\}}a_j^Tx+b_j$.

$\square$

3.32 *Products and ratios of convex functions.* In general the product or ratio of two convex functions is not convex. However, there are some results that apply to functions on $\mathbb{R}$. Prove the following.

a) If f and g are convex, both nondecreasing (or nonincreasing), and positive functions on an interval, then fg is convex.

Solución.

f y g son convexas y positivas. Así para $0\le\theta\le1$,

$$\begin{aligned}
f(\theta x+(1-\theta)y)g(\theta x+(1-\theta)y) \le&(\theta f(x)+(1-\theta)f(y))(\theta g(x)+(1-\theta)g(y))\\
=&\theta f(x)g(x)+(1-\theta)f(y)g(y)\\
&+\theta(1-\theta)(f(y)-f(x))(g(x)-g(y))
\end{aligned}$$

Como f y g son ambas crecientes o decrecientes $(f(y)-f(x))(g(x)-g(y))<0$, así

$$f(\theta x+(1-\theta)y)g(\theta x+(1-\theta)y)\le\theta f(x)g(x)+(1-\theta)f(y)g(y).$$

$\square$

b) If f, g are concave, positive, with one nondecreasing and the other nonincreasing, then fg is concave.

Solución.
Es completamente análogo al literal anterior variando el signo □

c) If f is convex, nondecreasing, and positive, and g is concave, nonincreasing, and positive, then f/g is convex.

Solución.
Considerando que f/g es convexa esto es un resultado inmediato del literal a). □

3.36 Derive the conjugates of the following functions.

a) *Max function.* $f(x) = \max_{i=1,\ldots,n} x_i$ on $\mathbb{R}^n$.

Solución.
Es claro que esta función diverge, por lo que $f^* = \infty$ □

b) *Sum of largest elements.* $f(x) = \sum_{i=1}^{r} x_{[i]}$ on $\mathbb{R}^n$.

Solución.
Análogo al caso anterior □

c) *Piecewise-linear function on* $\mathbb{R}$. $f(x) = \max_{i=1,\ldots,m}(a_i x + b_i)$ on $\mathbb{R}$. You can assume that the a_i are sorted in increasing order, i.e., $a_1 \leq \cdots \leq a_m$, and that none of the functions $a_i x + b_i$ is redundant, i.e., for each k there is at least one x with $f(x) = a_k x + b_k$.

Solución.

$$f^*(y) = \sup_x \left(xy - \max_{i=1,\ldots,m}(a_i x + b_i) \right)$$

□

d) *Power function.* $f(x) = x^p$ on $\mathbb{R}_{++}$, where $p > 1$. Repeat for $p < 0$.

Solución.
Tomamos $q = p/(p-1)$. Para $p > 1$. x^p es convexa en $\mathbb{R}_+$. Con $y < 0$ la función $yx - x^p$ tiene máximo en $x = 0$, así $f^*(y) = 0$. Para $y > 0$ la función tiene máximo en $x = (y/p)^{1/(p-1)}$, con

$$y(y/p)^{1/(p-1)} - (y/p)^{p/(p-1)} = (p-1)(y/p)^q.$$

así

$$f^*(y) = \begin{cases} 0 & y \leq 0 \\ (p-1)(y/p)^q & y > 0 \end{cases}$$

Analogo al caso anterior, para $p < 0$, $f^*(y) = \frac{-p}{q}(-y/p)^q$.

□

e) *Negative geometric mean.* $f(x) = -\left(\prod x_i\right)^{1/n}$ on $\mathbb{R}^n_{++}$.

Solución.

$$f^*(y) = \begin{cases} 0 & \text{if } y \preceq 0, \quad \left(\prod_i (-y_i)\right)^{1/n} \geq 1/n \\ \infty & \text{otherwise.} \end{cases}$$

□

f) *Negative generalized logarithm for second-order cone.* $f(x,t) = -\log\left(t^2 - x^T x\right)$ on $\{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid \|x\|_2 < t\}$.

Solución.

$$f^*(y,u) = -2 + \log 4 - \log\left(u^2 - y^T y\right), \quad \operatorname{dom} f^* = \{(y,u) \mid \|y\|_2 < -u\}$$

□

3.49 Show that the following functions are log-concave.

a) *Logistic function:* $f(x) = e^x/(1+e^x)$ with dom $f = \mathbb{R}$.

Solución.

$$\ln(f(x)) = ln(e^x/(1+e^x)) = x - \ln(1+e^x)$$

Donde x es lineal, por tanto cóncava y $\ln(1+e^x)$ es convexa, por lo que $-\ln(1+e^x)$ es cóncava y así $f(x)$ es log-cóncava □

b) *Harmonic mean:*

$$f(x) = \frac{1}{1/x_1 + \cdots + 1/x_n}, \quad \text{dom } f = \mathbb{R}^n_{++}.$$

Solución.
Consideremos la función $h(x)$ y sus primeras y segundas derivadas

$$h(x) = \log f(x) = -\log\left(1/x_1 + \cdots + 1/x_n\right)$$

$$\begin{aligned}
\frac{\partial h(x)}{\partial x_i} &= \frac{1/x_i^2}{1/x_1 + \cdots + 1/x_n} \\
\frac{\partial^2 h(x)}{\partial x_i^2} &= \frac{-2/x_i^3}{1/x_1 + \cdots + 1/x_n} + \frac{1/x_i^4}{(1/x_1 + \cdots + 1/x_n)^2} \\
\frac{\partial^2 h(x)}{\partial x_i \partial x_j} &= \frac{1/\left(x_i^2 x_j^2\right)}{(1/x_1 + \cdots + 1/x_n)^2} \quad (i \neq j).
\end{aligned}$$

Además de esto, sabemos que $y^T \nabla^2 h(x) y \prec 0$ para cualquier $y \neq 0$, así gracias a la desigualdad de Cauchy-Schwarz tenemos que

$$\left(\sum_{i=1}^n y_i/x_i^2\right)^2 < 2\left(\sum_{i=1}^n 1/x_i\right)\left(\sum_{i=1}^n y_i^2/x_i^3\right)$$

Lo cual finaliza la prueba. □

c) *Product over sum:*

$$f(x) = \frac{\prod_{i=1}^n x_i}{\sum_{i=1}^n x_i}, \quad \text{dom } f = \mathbb{R}^n_{++}$$

Solución.

$$f(x) = \sum_{i=1}^n \log x_i - \log \sum_{i=1}^n x_i$$

Consideremos la linea $x + tv$, con $x, v \in \mathbb{R}^n$ y $x \succ 0$

$$\tilde{f}(t) = \sum_i \log\left(x_i + tv_i\right) - \log \sum_i \left(x_i + tv_i\right).$$

Derivando,

$$\tilde{f}'(t) = \sum_i \frac{v_i}{x_i + tv_i} - \frac{\mathbf{1}^T v}{\mathbf{1}^T x + t\mathbf{1}^T v}.$$

Derivando nuevamente,

$$\tilde{f}''(t) = -\sum_i \frac{v_i^2}{(x_i + tv_i)^2} + \frac{\left(1^T v\right)^2}{\left(1^T x + t\mathbf{1}^T v\right)^2}.$$

Debemos probar que

$$\tilde{f}''(0) = -\sum_i \frac{v_i^2}{x_i^2} + \frac{\left(1^T v\right)^2}{\left(\mathbf{1}^T x\right)^2} \leq 0$$

se tiene para toda v, y toda$x \succ 0$. Sabemos que esto ocurre si $\mathbf{1}^T v = 0$. Si $\mathbf{1}^T v \neq 0$, sin perdida de generalidad $\mathbf{1}^T v = \mathbf{1}^T x$. Así, debemos enfocarnos en

$$\sum_i \frac{v_i^2}{x_i^2} \geq 1$$

Tenemos

$$\frac{v_i}{x_i^2} = \lambda$$

por lo que $v_i = \lambda x_i^2$. Además $\mathbf{1}^T v = \mathbf{1}^T x$ implica nuestro λ, lo cual resulta

$$v_i^\star = \frac{\sum_k x_k}{\sum_k x_k^2} x_i^2$$

Por tanto el valor mínimo de $\sum_i v_i^2/x_i^2$ con $\mathbf{1}^T v = \mathbf{1}^T x$ es

$$\sum_i \left(\frac{v_i^*}{x_i}\right)^2 = \left(\frac{\sum_k x_k}{\sum_k x_k^2}\right)^2 \sum_i x_i^2 = \left(\frac{\mathbf{1}^T x}{\|x\|_2}\right)^2 \geq 1$$

Lo cual concluye la prueba

□

d) *Determinant over trace:*

$$f(X) = \frac{\det X}{\operatorname{tr} X}, \quad \operatorname{dom} f = \mathbf{S}_{++}^n$$

Solución.

Esto es inmediato al notar que $\det(x)$ es cóncava, al igual que logaritmo, por lo que su composición lo es, al igual que con la función traza. Luego su diferencia es cóncava. Así

$$\log(\det(X)) - \log(\operatorname{tr}(x)) = \log(\det(X)/\operatorname{tr}(X))$$

es cóncava. Por tanto queda demostrado el ejercicio. □

3.54 *Log-concavity of Gaussian cumulative distribution function.* The cumulative distribution function of a Gaussian random variable,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

is log-concave. This follows from the general result that the convolution of two log-concave functions is log-concave. In this problem we guide you through a simple self-contained proof that f is log-concave. Recall that f is log-concave if and only if $f''(x)f(x) \leq f'(x)^2$ for all x.

a) Verify that $f''(x)f(x) \leq f'(x)^2$ for $x \geq 0$. That leaves us the hard part, which is to show the inequality for $x < 0$.

Solución.

Considere las siguientes derivadas de f:

$$f'(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}, \quad f''(x) = -\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}}.$$

Para $x \geq 0$, $f''(x) \leq 0$, por lo que $f''(x)f(x) \leq f'(x)^2$. □

b) Verify that for any t and x we have $t^2/2 \geq -x^2/2 + xt$.

Solución.

Observe que como $h(t) = \frac{t^2}{2}$ es convexa y diferenciable, entonces para cualesquiera x, t

$$\frac{t^2}{2} \geq -\frac{x^2}{2} + xt = \frac{x^2}{2} - x(t - x),$$

lo cual corresponde a la desigualdad $h(t) \geq h(x) - \nabla(x)^T(t - x)$. □

c) Using part *b)* show that $e^{-t^2/2} \leq e^{x^2/2-xt}$. Conclude that, for $x < 0$,

$$\int_{-\infty}^{x} e^{-t^2/2} dt \leq e^{x^2/2} \int_{-\infty}^{x} e^{-xt} dt$$

Solución.
Por el literal anterior se tiene que $-\frac{t^2}{2} \leq \frac{x^2}{2} - xt$, luego $e^{-t^2/2} \leq e^{x^2/2-xt}$. Al integrar, obtenemos la desigualdad deseada para $x < 0$:

$$\int_{-\infty}^{x} e^{-t^2/2} dt \leq \int_{-\infty}^{x} e^{\frac{x^2}{2}-xt} dt = e^{x^2/2} \int_{-\infty}^{x} e^{-xt} dt.$$

□

d) Use part *c)* to verify that $f''(x) f(x) \leq f'(x)^2$ for $x \leq 0$.

Solución.
Queremos ver que para $x \leq 0$ se cumple la siguiente desigualdad

$$-\frac{xe^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \leq \left(\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \right)^2 \Leftrightarrow -xe^{-\frac{x^2}{2}} \int_{-\infty}^{x} e^{-t^2/2} dt \leq e^{-x^2}$$

$$\Leftrightarrow \int_{-\infty}^{x} e^{-t^2/2} dt \leq -\frac{e^{-\frac{x^2}{2}}}{x}$$

Esta última desigualdad se cumple ya que por el literal anterior $\int_{-\infty}^{x} e^{-t^2/2} dt \leq e^{x^2/2} \int_{-\infty}^{x} e^{-xt} dt$, y $\int_{-\infty}^{x} e^{-xt} dt = \frac{e^{-x^2}}{-x}$. Luego

$$e^{\frac{x^2}{2}} \frac{e^{-x^2}}{-x} \leq -\frac{e^{-\frac{x^2}{2}}}{x}.$$

□