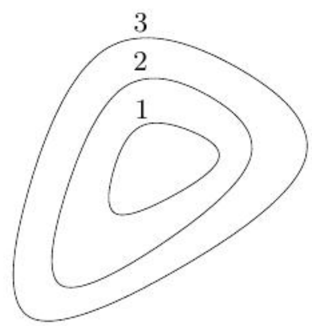
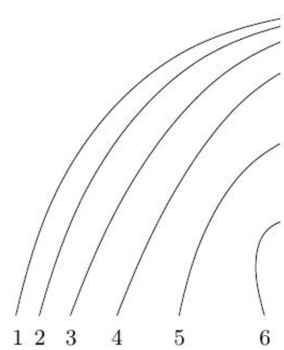
Tarea 3

Diego Romero Iregui dromeroi@unal.edu.co
Brayan Alejandro Romero brromeroc@unal.edu.co
David Felipe Mora dmorara@unal.edu.co

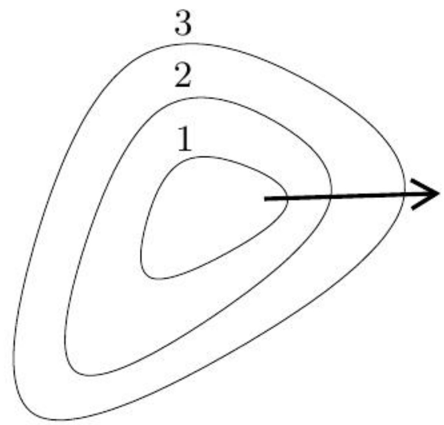
Ejercicios, libro guia

3.2 A continuación se encuentran los conjuntos de nivel de dos funciones. Puede ser convexas?(convacas, quasiconvacas, quasiconvexas).





<u>*Solución.*</u>
La primera función no puede ser convexa ya que si pasamos una linea por la parte inferior como se muestra en la siguiente figura:



La función crece rápido, pero luego crece mas lento por lo que el hessiano en esa dirección no es positivo y la función no puede ser convexa. Por otro lado, la función podría ser quasiconvexa ya que los subconjuntos de nivel son convexos, sin embargo no puede ser cóncava o quasiconcava porque los superconjuntos de nivel no son convexos. Repitiendo el mismo análisis para la segunda función se puede ver que puede ser cóncava y por lo tanto quasicóncava, pero no puede ser convexa o quasiconvexa por que los subconjuntos de nivel no son convexos

3.4 Muestre que una función continua $f : \mathbb{R}^n \to \mathbb{R}$ es convexa si y solo si para todo segmento de linea, su valor promedio es menor o igual al promedio de los valores de los puntos finales del segmento: Para todo $x, y \in \mathbb{R}^n$.

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

Solución.
Primero suponga que f es convexa. la desigualdad de jansen puede escribirse como

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

Para $0 \leq \lambda \leq 1$, integrando ambos lados entre 0 y 1 obtenemos.

$$\in_0^1 f(x+\lambda(y-x))d\lambda \leq \int_0^1 (f(x)+\lambda(f(y)-f(x)))d\lambda = \frac{f(x)+f(y)}{2}$$

Para el reciproco. Suponga que f es no convexa, entonces existen x,y y algún $\theta_0 \in (0,1)$ tal que

$$f(\theta x + (1-\theta_0)y) > \theta_0 f(x) + (1-\theta_0)f(y)$$

Considere la función de θ dada por

$$F(\theta) = f(\theta x + (1-\theta)y) - \theta f(x) - (1-\theta)f(y)$$

La cual es continua dado que f lo es. note que F es cero en $\theta = 0, 1$ y positiva en θ_0, sea α el 0 mas grande de F, detrás de θ_0 y sea β el cero mas pequeño de F por encima de θ_0, defina entonces $u = \alpha x + (1-\alpha)y$ y $v = \beta x + (1-\beta)y$, en el intervalo (α, β), nosotros tenemos que

$$F(\theta) = f(\theta x + (1-\theta)y) > \theta f(x) + (1-\theta)f(y)$$

luego para $\theta \in (0, 1)$,

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

Integrando esta expresión desde $\theta = 0$ hasta $\theta = 1$, nos da que

$$\int_0^1 f(\theta u + (1-\theta)v)d\theta > \int_0^1 \theta f(u) + (1-\theta)f(v)d\theta = \frac{f(u)+f(v)}{2}$$

En otras palabras el promedio de f sobre el intervalo $[u, v]$ excede el promedio en los valores finales.

3.6 Funciones y epígrafos. Cuándo es el epígrafo de una función un semiespacio? Cuándo es el epígrafo de una función un cono convexo? Cuándo de el epígrafo de una función un poliedro?

Solución.
En un primer lugar, recuérdese que el epígrafo de una función real $f : R^n \to R$ es el conjunto de puntos situados en o sobre su gráfica. Si se quisiera que este conjunto fuera un semiespacio, la función que lo delimita tendría que ser un hiperplano. Por lo tanto, f tendría que ser una función lineal.
Si se quisiera que el epígrafo de f fuera un cono convexo, hay dos primeros requisitos que debería cumplir, pasar por el origen y ser convexa. Además, si se toma como nulo el valor de todas las variables menos una, la función resultante siempre deberá ser definida a trozos, siendo un función lineal de $-\infty$ a 0, y una función lineal, posiblemente diferente, de 0 a ∞.
Finalmente, el epígrafo de f será un poliedro si es la intersección entre finitos semiespacios. Esto se obtiene cuando la función es convexa, y definida a trozos entre un número finito de funciones lineales.

3.13 La divergencia Kullback Leibler D_{KL} esta definida por:

$$D_{KL} = \sum_{i=1}^{n} u_i \log\left(\frac{u_i}{v_i}\right) - u_i + v_i \tag{1}$$

para $u, v \in \mathbb{R}^n_{++}$. Demostrar que $D_{KL}(u, v) \geq 0$ para todo $u, v \in \mathbb{R}^n_{++}$ y que $D_{KL}(u, v) = 0$ si y solo si $u = v$.

<u>*Solución.*</u>

Primero note que $D_{KL}(u, v) = f(u) - f(v) - \nabla f(v)^T(u-v)$ con $f(u) = \sum_{i=1}^{n} u_i \log u_i$.
En efecto como $\nabla f(v) = (1 + \log v_1, \cdots, 1 + \log v_n)$:

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

Como $f''(x) = \frac{1}{x} > 0$ para $x > 0$, entonces por la condición de segundo orden, f es convexa. Por lo tanto cumple la condición de primero orden, es decir que para todo $u, v \in \mathbb{R}^n_{++}$:

$$f(u) \geq f(v) - \nabla f(v)^T(u - v) \tag{2}$$

Reorganizando los términos obtenemos que $D_{KL} \geq 0$.

Por otro lado, tenemos que si $u = v$ entonces:

$$D_{KL}(u, v) = f(u) - f(u) - \nabla f(u)^T(u - u) = 0$$

3.16 Para cada una de las siguientes funciones determine si es convexa, concava, quasiconvexa o quasiconcava.

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

<u>*Solución.*</u>

Esta función es estrictamente convexa (es siempre creciente y con doble derivada mayor a 0), y por tanto también cuasiconvexa. Además es cuasiconcava pero no es concava.

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

<u>*Solución.*</u>
El Hessiano de f es:

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

El cual no es ni definido positivo ni definido negativo. Por tanto f como f es 2 veces diferenciable, se sigue que no es ni convexa ni concava. Si es quasiconcava dado que su superconjuntos de nivel son:

$$\{(x_1, x_2) \in \mathbb{R}^2_{++} : x_1 x_2 \geq \alpha\}$$

son convexos. Esta función no es quiasiconvexa pues no siempre se tiene que el conjunto

$$\{(x_1, x_2) \in \mathbb{R}^2_{++} : x_1 x_2 \leq 0\}$$

sea convexo, basta tomar $\alpha = 1$.

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

<u>*Solución.*</u>
Note que el hessiano de f es:

$$\nabla^2 f(x) = \frac{1}{x_1 x_2} \begin{bmatrix} 2/x_1^2 & 1/x_1 x_2 \\ 1/x_1 x_2 & 2/x_2^2 \end{bmatrix} \succeq 0$$

Es decir esta matriz es semidefinida positiva. Dado que esta función no es lineal entonces no es concava, y los conjuntos

$$\{(x_1, x_2) \in \mathbb{R}^2_{++} : 1/x_1 x_2 \geq \alpha\}$$

no son necesariamente convexos basta tomar $\alpha = 1$, así esta función no es quasiconvexa.

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

<u>*Solución.*</u>
El Hessiano de f es

$$\nabla^2 f(x) = \frac{1}{x_1 x_2} \begin{bmatrix} 0 & -1/x_1 x_2 \\ -1/x_2^2 & 2x_1/x_2^3 \end{bmatrix} \succeq 0$$

El cual no es definido positivo ni negativo. por tanto f no es convexa o concava. si es quasilineal puesto que su superconjunto de nivel son semiespacios.

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

<u>*Solución.*</u>
Notemos que f es doblemente diferenciable, luego aplica el criterio del Hessiano, notemos que su hessiano esta dado por:

$$\nabla^2 f(x) = \frac{1}{x_1x_2}\begin{bmatrix} 0 & -1/x_1x_2 \\ -1/x_2^2 & 2x_1/x_2^3 \end{bmatrix} = (2/x_2)\begin{bmatrix} 1 \\ -2x_1/x_2 \end{bmatrix}\begin{bmatrix} 1 \\ -2x_1/x_2 \end{bmatrix}^T \succeq 0$$

Por lo tanto, f es convexa y quasiconvexa. Además esta función fue presentada en el libro y su gráfica se muestra acontinuación.

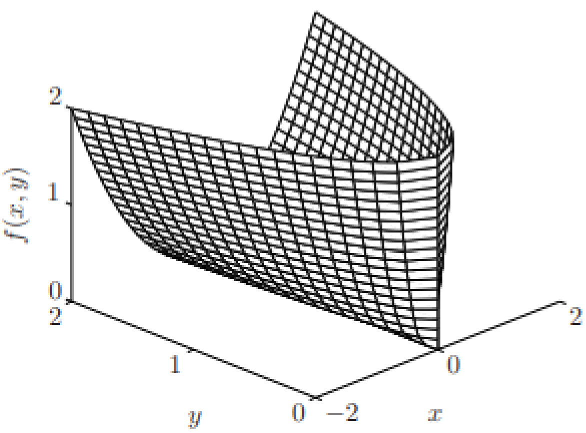


Figure 3.3 Graph of $f(x, y) = x^2/y$.

Como se ve esta función no es ni quasiconcava ni concava.

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

<u>*Solución.*</u>
Notemos que f es concava y quasiconcava, ya que f es diferenicable y por tanto aplica el critrerio del hessiano.

$$\begin{aligned}
\nabla^2 f(x) &= \begin{bmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(\alpha-1)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(\alpha-1)x_1^{\alpha-1}x_2^{-\alpha} & -\alpha(1-\alpha)x_1^{\alpha}x_2^{-\alpha-1} \end{bmatrix} \\
&= \alpha(\alpha-1)x_1^{\alpha}x_2^{1-\alpha}\begin{bmatrix} -1/x_1^2 & 1/x_1x_2 \\ 1/x_1x_2 & -1/x_2^2 \end{bmatrix} \\
&= -\alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha}\begin{bmatrix} -1/x_1 \\ -1/x_2 \end{bmatrix}\begin{bmatrix} -1/x_1 \\ -1/x_2 \end{bmatrix}^T \\
&\preceq 0.
\end{aligned}$$

Además como esta función no es lineal, se sigue que no es convexa, y como sus superconjuntos de nivel no son subespacios (solo son concavos por lo anterior) se sigue que f no es quasiconvexa.

3.20 *Composición con función afín.* Muestre que las siguientes funciones $f : R^n \to R$ con convexas.

Solución.

(a) $f(x) = ||Ax - b||$, con $A \in R^{m \times n}$, $b \in R^m$, y $|| \cdot ||$ es una norma en R^m.

Se trata de la composición entre funciones convexas, una norma y una función afín. Por lo tanto f es convexa.

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

Nótese que la función f es la composición entre la función $-det(X)^{1/m}$ y una transformación afín. Como X es una matriz de $m \times m$, la función que se le aplica puede ser vista como el producto de una constante por un promedio geométrico. Por lo tanto esta última función es constante y f también lo sería

(c) $f(X) = \mathbf{tr}(A_0 + x_1A_1 + ... + x_nA_n)^{-1}$, en $\{x \mid A_0 + x_1A_1 + ... + x_nA_n \succ 0\}$, donde $A_i \in S^m$. (Use el hecho de que $\mathbf{tr}(X^{-1})$ es convexo en S^m_{++}).

Esta vez, f es la composición entre la función $\mathbf{tr}X^{-1}$, que sabemos es convexa, y una transformación afín. Se puede entonces afirmar que f es a su vez convexa.

3.21 Mostrar que las siguientes funciones son convexas:

a. $f(x) = \max_{1 \leq i \leq n} ||A^{(i)}x - b^{(i)}||$ donde $A^{(i)} \in \mathbb{R}^{m \times n}$ y $b^{(i)} \in \mathbb{R}^m$

b. $\sum_{i=1}^{r} |x|_{[i]}$ donde $|x|$ es el resultado de aplicar valor absoluto componente a componente y $|x|_{[i]}$ es la i-esima componente mas grande de x.

Solución.

a. Las funciones $A^{(i)}x - b^{(i)}$ son afines, como las normas son convexas entonces $||A^{(i)}x - b^{(i)}||$ es convexa. Luego como el máximo punto a punto preserva convexidad entonces $\max_{1 \leq i \leq n} ||A^{(i)}x - b^{(i)}||$ es convexa.

b. $|x_{i_1}| + \cdots + |x_{i_n}|$ para $i_1, \cdots, i_r \leq n$, es una función convexa pues es la suma de funciones convexas $f_k(x) = |x_k|$. Luego $\sum_{i=1}^{r} |x|_{[i]} = \max\{|x_{i_1}| + \cdots + |x_{i_n}|, i_1, \cdots, i_r \leq n\}$ y como el máximo punto a punto de funciones convexas es convexa entonces la función es convexa.

3.22 *reglas de composición.* Muestre que las siguientes funciones son convexas.

(a) $f(x) = -\log(-\log(\sum_{i=1}^{m} e^{a_i^T x + b_i}))$ en $dom f = \{x : \sum_{i=1}^{m} e^{a_i^T x + b_i} < 1\}$, tu puedes utilizar el hecho de que $log(\sum_{i=1}^{n} e_i^y)$ es convexa

<u>*Solución.*</u>
Sea $g(x) = -\log(-\log(\sum_{i=1}^{m} e^{a_i^T x + b_i}))$ es convexa (composición de los sum exp con una función afín). luego -g es concava.La función $h(y) = -logy$ es convexa y decreciente. Por lo tanto $f(x) = h(-g(x))$ es convexa.

(b) $f(x, u, v) = -\sqrt{uv - x^T x}$, en $dom f = \{(x, u, v) : uv > x^T x, u, v > 0\}$. Use el hecho de que $x^T x/u$ es convexa en (x, u) para $u > 0$, y que $-\sqrt{x_1 x_2}$ es convexa en R^2_{++}.

<u>*Solución.*</u>
Podemos expresar a f como $f(x, u, v) = -\sqrt{u(v - x^T x/u)}$.La función $h(x_1, x_2) = -\sqrt{x_1 x_2}$ es convexa en $\mathbb{R}^2_{++}$ y decreciente en cada argumento. Las funciones $g(u, v, x) = u$ y $g_2(u, v, x) = v - x^T x/u$ son concavas. Por lo tanto $f(u, v, x) = h(g(u, v, x))$ es convexa.

(c) $f(x, u, v) = -\log(uv - x^T x)$ en $dom f = \{(x, u, v) : uv > x^T x, u, v > 0\}$.

<u>*Solución.*</u>
Nosotros podemos expresar a f como

$$f(x, u, v) = -\log u - \log(v - x^T x/u)$$

El primer termino es convexo. La función $v - x^T x/u$ es concava dado que v es lineal y $x^T x/u$ es convexo para $u > 0$. Por lo tanto el segundo termino de f es convexo, es la composición de una función decreciente $-logt$ y una función concava.

(d) $f(x, t) = -(t^p - ||x||_p^p)^{1/p}$ donde $p > 1$ y $dom f = \{(x, t) : t \geq ||x||_p\}$, puedes usar el hecho de que $||x||_p^p/u^{p-1}$ es convexa en (x, u) para $u > 0$ y que $-x^{1/p} y^{1-1/p}$ es convexa en R^2_+

<u>*Solución.*</u>
Nosotros podemos expresar f como:

$$f(x, t) = -(t^{p-1}(t - \frac{||x||_p^p}{t^{p-1}}))^{1/p} = -t^{1-1/p}(t - \frac{||x||_p^p}{t^{p-1}})^{1/p}$$

Esta es la composición de $h(y_1, y_2) = -y_1^{1/p} y_2^{1-1/p}$ (convexa y decraciente en cada argumento) y las dos funciones concavas.

$g_1(x, t) = t^{1-1/p}$ $g_2(x, t) = t - \frac{||x||_p^p}{t^{p-1}}$

(e) $f(x, t) = -\log(t^p - ||x||_p^p)$ donde $p > 1$ y $dom f = \{(x, t) : t > ||x||_p\}$. tu puedes usar el hecho de que $||x||_p^p/u^{p-1}$ es convexa en (x, u) para $u > 0$.

<u>*Solución.*</u>
Exprese f como

$$f(x,t) = -\log t^{p-1} - \log(t - ||x||_p^p/t^{p-1}) = -(p-1)\log t - \log(t - ||x||_p^p/t^{p-1})$$

El primer termino es convexo. El segundo es la composición de una fución decreciente convexa y una función concava, por tanto es convexa.

3.25 *Maximum probability distance between distributions.* Sean $p, q \in \mathbb{R}^n$ dos distribuciones de probabilidad en 1,...,n (Entonces $p, q \succeq 0, 1^T p = 1^T q = 1$). Se define la *máxima distancia de probabilidad* $d_{mp}(p,q)$ entre p y q como la diferencia máxima en probabilidad asignada por p y q sobre todos los eventos:

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

Acá $\mathbf{prob}(p,C)$ es la probabilidad de C, bajo la distribución o, i.e., $\mathbf{prob}(p,C) = \sum_{i \in C} p_i$.
Encuentre una expresión simple para d_{mp}, que utilice $||p-q||_1 = \sum_{i=1}^n |p_i - q_i|$, y muestre que d_{mp} es una función convexa en $\mathbb{R}^n \times \mathbb{R}^n$. (Su dominio es $\left\{(p,q)|p,q \succeq 0, 1^T p = 1^T q = 1\right\}$, pero es una extensión natural de todo $\mathbb{R}^n \times \mathbb{R}^n$.)

<u>*Solución.*</u>
Nótese que se desea maximizar $|\sum_{i \in C}(p_i - q_i)|$. Se puede afirmar que esto sucederá cuando todos los p_i sean mayores a los q_i, o cuando todos los q_i sean mayores al p_i correspondiente. Sin pérdida de generalidad, asúmase que ocurre cuando para cada índice, $p_i > q_i$. Se tendría entonces que

$$d_{mp}(p,q) = \sum_{p_i > q_i} (p_i - q_i)$$

Ahora bien, como

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

se tiene que

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

Entonces,

$$d_{mp}(p,q) = \frac{1}{2}\sum_{p_i > q_i} (p_i - q_i) - \frac{1}{2}\sum_{p_i \leq q_i} (p_i - q_i)$$

$$= \frac{1}{2}\sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2}||p - q||_1$$

Con esta última expresión se puede ver que d_{mp} es convexo.

3.29 Una función convexa $f : \mathbb{R}^n \to \mathbb{R}$ con $Dom(f) = \mathbb{R}^n$ es lineal por trozos si existe una partición de $\mathbb{R}^n$:

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

Donde $Int(X_i) \neq \emptyset$ y $Int(X_i) \cap Int(X_j) \neq \emptyset$ para todo $i \neq j$ y una familia de funciones afines $a_1^T x + b_1, \cdots, a_L^T x + b_L$ tal que $f(x) = a_i^T + b_i$ si $x \in X_i$. Mostrar que tal función tiene la siguiente forma:

$$f(x) = \text{máx}\{a_1^T x + b_1, \cdots, a_L^T x + b_L\}$$

Solución.
Como f es convexa, entonces cumple la desigualdad que para $0 \leq t \leq 1$:

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

Reorganizando y dividiendo por t obtenemos lo siguiente:

$$f(x) \geq f(y) + \frac{f(y + t(x - y)) - f(y)}{t}$$

Fije $x \in X_i$ y $y \in X_j$ para cualquier $j \neq i$, además tome t lo suficientemente pequeño para que $y + t(x - y) \in X_j$, entonces:

$$\begin{aligned} a_i^T x + b_i &\geq a_j^T + b_j + \frac{a_j^T(y + t(x - y)) + b^j - (a_j^T y + b^j)}{t} \\ &= a_j^T x + b_j \end{aligned}$$

Por lo tanto y como se demostró para i, j arbitrarios se tiene que:

$$f(x) = \text{máx}\{a_1^T x + b_1, \cdots, a_L^T x + b_L\}$$

3.32 *Productos y radios de funciones convexas.* En general el producto o radio de dos funciones convexas es no convexo. Sin embargo, hay algunos resultados aplicables a funciones en $\mathbb{R}$. Pruebe lo siguiente.

(a) Si f y g son funciones convexas, ambas no decrecientes (no crecientes), y positivas en un intervalo, entonces fg es convexa.

Solución.
Probemos la desigualdad de jensen. f y g son positivas y convexas, por lo tanto para $0 \leq \theta \leq 1$.

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

El tercer termino es menor o igual a 0 si f y g son ambas crecientes o ambas decrecientes, por lo tanto

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

(b) Si f, g son concavas, positivas, con una no-decreciente y la otra no creciente, entonces fg es concava.

Solución.
Probemos la desigualdad de jensen. f y g son positivas y concavas, por lo tanto para $0 \leq \theta \leq 1$.

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

El tercer termino es menor o igual a 0 si f y g son ambas crecientes o ambas decrecientes, por lo tanto

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

(c) Si f es convexa, no decreciente, y positiva, y g es concava, no creciente y positiva, entonces f/g es convexa.

Solución.
Basta ver que $1/g$ es convexa, positiva y creciente. Luego el resultado se sigue de la parte a

3.36 Derive el conjugado de las siguientes funciones.

(a) *Función Max.* $f(x) = max_{i=1,\dots,n}x_i$ en R^n.

Solución.
En un primer lugar, verifiquemos el dominio de f^*. Supóngase que y tiene un componente negativo. Sea entonces $y_k < 0$. Escojamos además un vector x tal que $x_k = -t$ y $x_i = 0$ para $i \neq k$. Véase entonces que si t va a infinito $x^T y - \max\, x_i = -ty_k$, que a su vez tiende a infinito. Por lo tanto y no está en el dominio.
Ahora bien, supongamos que $y \succeq 0$ y $1^T y > 1$. Si se toma $x = t1$ y se deja

que t vaya a infinito, $x^T - \max x_i = t1^T y - t$. En este caso, y cuando se supone que $1^T < 1$ y se toma $x = -t1$ el resultado no tendrá cota superior. Finalmente, si $y \succeq 0$ y $1^T y = 1$ se tiene $x^T y \leq \max x_i$, y por lo tanto $x^T y - \max x_i \leq 0$ para todo x, entonces se tiene que $f^*(y) = 0$

(b) *Suma de elementos mayores.* $f(x) = \sum_{i=1}^{r} x_{[i]}$ en R^n

<u>*Solución.*</u>
De nuevo, verifiquemos el dominio de f^*. Primero supóngase que existe al menos un $y_k < 0$. Escogemos entonces un vector x con $x_k = -t$ y $x_i = 0$ para $i \neq k$. Si t va a infinito, $x^T y - f(x) = -ty_k$ también lo hará, por lo que y no está en el dominio de f^*.
Ahora supóngase que existe $y_k > 1$. Si esta vez se toma x con $x_k = t$, $x_i = 0$ para $i \neq k$, si t va a infinito, $x^T y - f(x) = ty_k - t$ también lo hace. De nuevo y no estaría en dom f^*. Finalmente, si $1^T x \neq r$ se escoge x = t1 y entonces $x^T y - f(x) = t1^T y - tr$, lo cual no está acotado por arriba. Si y cumple todas las condiciones se tendría $x^T y \leq f(x)$ para todo x, por lo tanto $f^*(y) = 0$

(c) *función lineal por partes* en R. $f(x) = max_{i=1,\ldots,m}(a_i x + b_i)$ en R. Puede asumir que los a_i están ordenados en roden ascendente, i.e., $a_1 \leq ... \leq a_m$, y que ninguna función $a_i x + b_i$ es redundante, i.e., para cada k hay al menos un x con $f(x) = a_k x + b_k$.

<u>*Solución.*</u>
Nótese que $f^*(y) = sup\,(xy - \max_{i=1,\ldots,m}(a_i x + b_i))$, y que por lo tanto dom $f^* = [a_1, a_m]$ pues si y se encuentra fuera de ese rango $(xy - \max_{i=1,\ldots,m}(a_i x + b_i)$ no tiene cota superior. Para $a_i \leq y \leq a_{i+1}$ el supremo se alcanza en el punto $\frac{b_{i+1} - b_i}{a_{i+1} - a_i}$, por lo que $f^*(y) = -b_i - \frac{(b_{i+1} - b_i)(y - ai)}{a_{i+1} - a_i}$. Por lo tanto, f^* también es lineal por partes, y conecta los puntos $(a_i, -b_i)$ para $i = 1, ..., m$.

(d) *Función potencia* $f(x) = x^p$ en R_{++}, donde $p > 1$. Repita para $p < 0$.

<u>*Solución.*</u>
Supongamos primero que p¿1. Entonces, x^p es convexo en R_+. Si $y \leq 0$, el máximo de la función $yx - x^p$ para x¿0 en x=0, es decir $f^*(y)$ sería igual a 0. Si $y > 0$, el máximo de la función estaría en $x = (\frac{y}{p})^{\frac{1}{p-1}}$, y su valor sería $y(\frac{y}{p})^{\frac{1}{(p-1)}} - (\frac{y}{p})^{\frac{p}{(p-1)}} = (p-1)(\frac{y}{p})^{\frac{p}{(p-1)}}$
Análogamente para $p < 0$ se tendrá que dom $f^* = -R_{++}$ y $f^*(y) = \frac{-p}{q}(\frac{-y}{p})^{\frac{p}{(p-1)}}$

(e) *Promedio geométrico negativo* $f(x) = -(\prod x_i)^{1/n}$ en R^n_{++}.

<u>*Solución.*</u>
Nuevamente, verifiquemos el dominio de f^*. Supongamos en un primer lugar que existe $y_k > 0$. Tómese $x_k = t$ y $x_i = 1$ para $i \neq k$, entonces $x^T y - f(x) = ty_k + \sum_{i\neq k} y_i - t^{1/n}$. Esto último no tiene cota superior.

Ahora supóngase que $0 \succeq y$ y $\left(\prod_i(-y_i)\right)^{1/n} < \frac{1}{n}$, y tómese $x_i = -t/y_i$. Se tendría $x^T y - f(x) = -tn - t\left(\prod_i(-1/y_i)\right)^{1/n}$. Puede verse que cuando t tienda a infinito, esto también lo hará. Finalmente supóngase que $0 \succeq y$ y $\left(\prod_i(-y_i)\right)^{1/n} \geq \frac{1}{n}$ y $x \succeq 0$. Por la desigualdad de promedio aritmetico-geometrico se cumpliría que $\frac{x^T y}{n} \geq \left(\prod_i(-y_i x_i)\right)^{1/n} \geq \frac{1}{n}\left(\prod_i(x_i)\right)^{1/n}$. Por lo tanto $x^T \geq f(x)$ para $x_i = -1/y_i$, y entonces $f^*(y) = 0$.

(f) *logaritmo generalizado negativo para cono de segundo orden.* $f(x,t) = -log(t^2 - x^T x)$ en $\{(x,t) \in R^n \times R \mid ||x||_2 < t\}$.

<u>*Solución.*</u>
En un primer lugar nótese que $f^*(y,z) = -2 + log4 - log(z^2 - y^T y)$ y dom $f^* = \{(y,z) \mid ||y||_2 < -z\}$. De nuevo verificaremos el dominio. Supóngase que $||y||_2 \geq -z$ y escójase $x = sy$ y $t = s(||x||_2 + 1)$. Se tiene que $y^T x + tz > sy^T - sz^2 = s(z^2 - y^T y) \geq 0$, y se puede ver entonces que $y^T x + tz + log(t^2 - x^T x)$ no está acotado superiormente.
Ahora suponga que $||y_2|| < z$. Si se toma la derivada de $y^T x + ut + log(t^2 - x^T x)$ respecto a x y se toma $t = 0$ al resolver se obtiene que $x = \frac{2y}{z^2 - y^T y}$ y $t = -\frac{2z}{z^2 - y^T y}$. Por lo tanto $f^*(y,z) = zt + y^T x + log(t^2 - x^T x)$.

3.49 Mostrar que las siguientes funciones son log-convacas.

a. función logística

$$f(x) = \frac{e^x}{1 + e^x}$$

b. Media armónica

$$f(x) = \frac{1}{\frac{1}{x_1} + \cdots + \frac{1}{x_n}}$$

c. Producto sobre suma

$$f(x) = \frac{\prod_{i=0}^n x^i}{\sum_{i=1}^n x_i}$$

d. Determinante sobre traza

$$f(x) = \frac{Det(X)}{Tr(X)}$$

<u>*Solución.*</u>

a. Tenemos que:
$$\log\left(\frac{e^x}{1+e^x}\right) = x - \log(1+e^x)$$
Por lo tanto tenemos que x es lineal y por lo tanto cóncava. $\log(1+e^x)$ es la función log-sum-exp con $x_1 = 0$ y $x_2 = x$, por lo tanto es convexa, luego $-\log(1+e^x)$ es cóncava y por ende la función es log-cóncava.

b. Vamos a demostrar que $y^T\nabla^2 h(x)y \preceq 0$ para todo $y \neq 0$ con $h(x) = \log(f(x))$. Para ello vamos a calcular las derivadas parciales:
$$\frac{\partial h}{\partial x_i}(x) = \frac{\frac{1}{x_i^2}}{\frac{1}{x_1}+\cdots+\frac{n}{x_n}}$$
El hessiano esta conformado por:
$$\frac{\partial^2 h}{\partial^2 x_i}(x) = \frac{\frac{-2}{x_i^3}}{\frac{1}{x_1}+\cdots+\frac{n}{x_n}} + \frac{\frac{1}{x_i^4}}{(\frac{1}{x_1}+\cdots+\frac{n}{x_n})^2}$$
Para $i \neq j$
$$\frac{\partial^2 h}{\partial x_i \partial x_j}(x) = \frac{\frac{1}{x_i^2 x_j^2}}{\frac{1}{x_1}+\cdots+\frac{n}{x_n}}$$
Ahora bien, $y^T\nabla^2 h(x)y \preceq 0$, lo que es lo mismo a decir:
$$\left(\sum_{i=1}^n \frac{y_i}{x_i^2}\right)^2 < 2\left(\sum_{i=1}^n \frac{1}{x_i}\right)\left(\sum_{i=1}^n \frac{y_i^2}{x_i^3}\right)$$
Lo cual es cierto de acuerdo a la desigualdad de Cauchy-Schwarz $(xy^T)^2 \leq ||x||^2||y||^2$ tomando $x = \frac{1}{\sqrt{x_i}}$ y $y = \frac{y_i}{x_i\sqrt{x_i}}$ Por lo tanto $y^T\nabla^2 h(x)y \preceq 0$ y la función es log-concava.

c. Para verificar que la función es cóncava, debemos mostrar que:
$$\begin{aligned} h(x) &= \log(f(x)) \\ &= \log(\prod_{i=1}^n x_i) - \log(\sum_{i=1}^n x_i) \\ &= \sum_{i=1}^n \log x_i - \log\sum_{i=1}^n x_i \end{aligned}$$
es cóncava. Primero como la concavidad(y convexidad) se preservan a través de una recta arbitraria del dominio, es decir, entonces considere la recta $x + tv$ con $x, v \in \mathbb{R}^n$ y la función $g(t) = h(x+tv)$. Luego tenemos:
$$g(t) = \sum_{i=1}^n \log(x_i + tv_i) - \log\sum_{i=1}^n x_i + tv_i$$

Diferenciando obtenemos:

$$g'(t) = \sum_{i=1}^{n} \frac{v_i}{x_i + tv_i} - \frac{\sum_{i=1}^{n} v_i}{\sum_{i=1}^{n} x_i + tv_i}$$

$$g''(t) = -\sum_{i=1}^{n} \frac{v_i^2}{(x_i + tv_i)^2} - \frac{(\sum_{i=1}^{n} v_i)^2}{(\sum_{i=1}^{n} x_i + tv_i)^2}$$

Ahora bien para establecer concavidad necesitamos que:

$$h''(x) = g''(0) = -\sum_{i=1}^{n} \frac{v_i^2}{x_i^2} - \frac{(\sum_{i=1}^{n} v_i)^2}{(\sum_{i=1}^{n} x_i)^2} \leq 0$$

para todo v y con $x \succeq 0$. Asumimos sin perdida de generalidad que $\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} x_i$, por lo que el problema se reduce a verificar que:

$$\sum_{i=1}^{n} \frac{v_i^2}{x_i^2} \geq 1$$

Luego el valor mínimo se da cuando $v_i = \frac{\sum_{k=1}^{n} x_k}{\sum_{k=1}^{n} x_k^2} x_i$ y tenemos lo siguiente:

$$\sum_{i=1}^{n} (\frac{v_i}{x_i})^2 = \left(\frac{\sum_{k=1}^{n} x_k}{\sum_{k=1}^{n} x_k^2} \right)^2 \sum_{i=1}^{n} x_i^2 = (\frac{\sum_{i=1} x_i}{||x||_2})^2 \leq 1$$

Por lo tanto la desigualdad se cumple y la función es log-cóncava.

d. Queremos ver que la función:

$$h(X) = (X) = \log(det(X)) - \log(Tr(X))$$

es cóncava. Para ello vamos a restringir la función a la recta $X = Z + tV$ con $Z \succ 0$ y usamos la descomposición en valores propios $Z^{-\frac{1}{2}} V Z^{-\frac{1}{2}} = QDQ^{-1} = \sum_{i=1}^{n} \lambda_i q_i q_i^T$, luego tenemos:

$$\begin{aligned} h(Z + tV) &= \log(det(Z + tV)) - \log(Tr(Z + tV)) \\ &= \log(det(Z)) - \log(det(I + tZ^{-\frac{1}{2}} V Z^{-\frac{1}{2}})) - \log(Tr(Z(I + tZ^{-\frac{1}{2}} V Z^{-\frac{1}{2}}))) \\ &= \log(det(Z)) - \sum_{i=1}^{n} \log(1 + t\lambda_i) - \log(\sum_{i=1}^{n} (q_i^T Z q_i)(1 + t\lambda_i)) \\ &= \log(det(Z)) + \sum_{i=1}^{n} \log(q_i^T Z q_i) - \sum_{i=1}^{n} \log((q_i^T Z q_i)(1 + t\lambda_i)) \\ &\quad - \log(\sum_{i=1}^{n} (q_i^T Z q_i)(1 + t\lambda_i)) \end{aligned}$$

Donde al final tenemos la constante $\log(det(Z)) + \sum_{i=1}^{n} \log(q_i^T Z q_i)$ menos una función cóncava que es la función del ejercicio anterior con $x_i = (q_i^T Z q_i)(1+ t\lambda_i)$. Por lo tanto la función es log-cóncava.

3.54 *log-concavidad de la funcion de distribución gausiana acumulativa.* La función de distribuición acumulativa de gaus de variable aleatoria.

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

Es log concava. Esto se sigue de los resultados generales de convolución de dos funciones log-concavas. En estre problema nosotros te guiaremos atravez de una simple prueba de que f es log concava. Recuerda que f es log-concava si y solo si $f''(x)f(x) \leq f'(x)^2$ para todo x.

(a) Verifique que $f''(x)f(x) \leq f'(x)^2$ para $x > 0$.

<u>*Solución.*</u>

Notemos que la primera y segunda derivada de f son:

$$f'(x) = e^{-x^2/2}/\sqrt{2\pi}, \qquad f''(x) - xe^{-x^2/2}/\sqrt{2\pi}$$

luego como la exponencial es siempre positiva, entonces para $x \geq 0$, tenemos que $f''(x) \leq 0$

(b) Verifique que para cualquier t y x, nosotros tenemos $t^2/2 \geq -x^2/2 + xt$.

<u>*Solución.*</u>
Dado que $t^2/2$ es convexo tenemos que

$$t^2/2 \geq x^2/2 + x(t-x) = xt - x^2/2$$

Esta es la inecuación general.

$$g(t) \geq g(x) + g'(x)(t-x)$$

La cual se tiene para cualquier función diferenciable convexa, aplicada a $g(t) = t^2/2$

(c) Usando la parte (b) muestre que $e^{-t^2/2} \leq e^{x^2/(2-xt)}$. concluya que para $x < 0$

<u>*Solución.*</u>

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

Por el punto anterior $-t^2/2 \leq x^2/(2-xt)$, luego como la exponencial es una funcón creciente tenemos que $e^{-t^2/2} \leq e^{x^2/(2-xt)}$, luego integrando a ambos lados de la igualdad para $x < 0$ tenemos que

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

(d) Use la parte (c) para verificar que $f''(x)f(x) \leq f'(x)^2$ para $x \leq 0$.

Solución.
Esta desigualdad se reduce a

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

Es decir.

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

Esto se sigue de la parte c dado que:

$$\int_{-\infty}^{x} e^{-xt} dt = \frac{e^{-x^2}}{-x}$$

Bibliografía.

[1] 2006 *Convex Optimization* Stephen Boyd y Lieven Vandenberghe.

[2] *Stanford lectures* https://web.stanford.edu/class/ee364a/lectures.html