Taller 2 Optimización

Daniel Ricardo Rodriguez
Paola Andrea Gallegos Pinto
Raúl Mauricio Peña Losada

May 2022

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)?

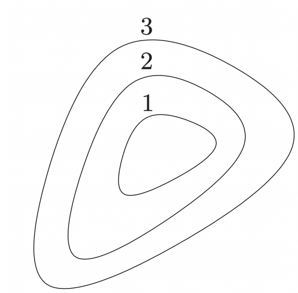
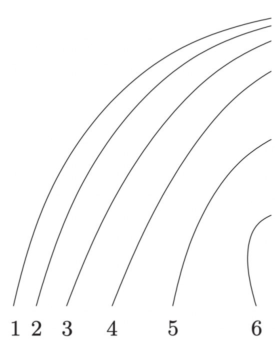
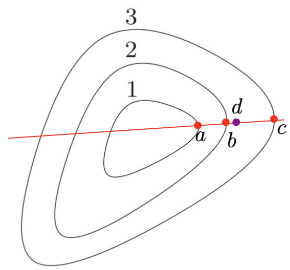


Figure 1:

Explain your answer. Repeat for the level curves shown below.



The first function can be quasiconvex because the figure of the level sets are convex sets also is not concave or queasiconcave because for $f(x) \geq 1$ the set is not convex, also the set is not convex: because the line that pass between **a**



and **c** contain **d** and **d** is not in f

making same analysis we notice that the sublevel sets are not convex and therefore f are not quasiconvex but the sublevels set are concave so the function can be concave so therefore can be also queasiconcave

$- --$

3.4 [RV73, page 15] Show that a continuous function $f : \mathbf{R}^n \to \mathbf{R}$ 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 $x, y \in \mathbf{R}^n$,

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

- First, let's show that if $f(x)$ is convex, then $\int_0^1 f(x + \lambda(y - x))d\lambda \leq \frac{f(x)+f(y)}{2}$.

  We have that since $f(x)$ is convex, then Jensen's inequality is holds: $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$

  Let's also note that:

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

  We can switch the labels of x and y without changing the equality, since x and y are just labels to call two points in the domain of the function, so we get:

$$x + \theta(y - x)$$

  Using this on Jensen's inequality, we have that:

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

  Now, if we integrate with respect to theta from 0 to 1 in both sides, we get:

$$
\begin{aligned}
\int_0^1 f(x+\theta(y-x))d\theta &\leq \int_0^1 f(x)+\theta(f(y)-f(x))d\theta \\
&= (\theta f(x) + \frac{\theta^2(f(y)-f(x))}{2})\Big|_0^1 \\
&= (1)f(x) + \frac{(1)^2(f(y)-f(x))}{2} - ((0)f(x) + \frac{(0)^2(f(y)-f(x))}{2}) \\
&= \frac{2f(x)+f(y)-f(x)}{2} - (0+\frac{0}{2}) \\
&= \frac{f(x)+f(y)}{2}
\end{aligned}
$$

If we do a simple substitution $\lambda = \theta$ on the left-hand side, we get:

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

- Now, to prove that if $\int_0^1 f(x+\lambda(y-x))d\lambda \leq \frac{f(x)+f(y)}{2}$ then $f(x)$ is convex, we can prove that if $f(x)$ is not convex, then $\int_0^1 f(x+\lambda(y-x))d\lambda > \frac{f(x)+f(y)}{2}$

  First, assuming that $f(x)$ is not convex, then we know that there's at least a θ such that Jensen's inequality doesn't hold, let's call it a, then we have:

  $$f(ax+(1-a)y) > af(x)+(1-a)f(y)$$

  Now, since $f(x)$ is not convex, we can say there's a "mountain" in the plot of $f(x)$. If we now call θ_0 and θ_1 the start and end of this "mountain", or more formally, the nearest points both before and after in the convex combination around a on which Jensen's inequality holds. In the worst case scenario (the function is completely concave in the interval (x,y)) we know that it will hold at least for x and y, or for $\theta_0 = 0$ and $\theta_1 = 1$.

  Now let $l_0 = \theta_0 x + (1-\theta_0)y$ and $l_1 = \theta_1 x + (1-\theta_1)y$ be the points for which $f(\theta x + (1-\theta)y) = \theta f(x) + (1-\theta)f(y)$. Then, for λ in $(0,1)$ we have that $f(\lambda l_0 + (1-\lambda)l_1) > \lambda f(l_0) + (1-\lambda)f(l_1)$

Finally, if we integrate on both sides from $\lambda = 0$ to $\lambda = 1$:

$$\begin{aligned}
\int_0^1 f(\lambda l_0 + (1-\lambda)l_1)d\lambda &> \int_0^1 \lambda f(l_0) + (1-\lambda)f(l_1)d\lambda \\
&= \int_0^1 \lambda f(l_0) + f(l_1) - \lambda f(l_1)d\lambda \\
&= \int_0^1 f(l_1) + \lambda(f(l_0) - f(l_1))d\lambda \\
&= \lambda f(l_1) + \frac{\lambda^2(f(l_0) - f(l_1))}{2}\Big|_0^1 \\
&= (1)f(l_1) + \frac{(1)^2(f(l_0) - f(l_1))}{2} - \Big((0)f(l_1) - \frac{(0)^2(f(l_0) - f(l_1))}{2}\Big) \\
&= f(l_1) + \frac{(f(l_0) - f(l_1))}{2} - \Big(0 - \frac{0}{2}\Big) \\
&= \frac{2f(l_1) + f(l_0) - f(l_1)}{2} \\
&= \frac{f(l_1) + f(l_0)}{2}
\end{aligned}$$

Then, we have:

$$\int_0^1 f(\lambda l_0 + (1-\lambda)l_1)d\lambda > \frac{f(l_1) + f(l_0)}{2}$$

And with this we can prove that if $f(x)$ is not convex, then $\int_0^1 f(x+\lambda(y-x))d\lambda > \frac{f(x)+f(y)}{2}$, thus proving by contraposition that if $\int_0^1 f(x+\lambda(y-x))d\lambda \leq \frac{f(x)+f(y)}{2}+$, then $f(x)$ is convex.

$-----------------------------------$

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

Solution: Assume f is a real-valued affine fuction then
$(x,t) \in epi(f) \Leftrightarrow f(x) \leq t \Leftrightarrow c^T x + d \leq t \Leftrightarrow c^T x \leq t - d$.
The last inequality shows $c^T x \leq t - d$ that means $epi(f)$ is a halfspace.
considering that when $epi(f)$ is a halfspace:
$(x,t) \in epi(f) \Leftrightarrow (b, c^t) * (x,t)^T \leq t \Leftrightarrow c^T * x + bt \leq t \Leftrightarrow f(x) \leq t$
which shows that f is equivalent to an affine fuction.
Then, If the function is positive homogenous, then by just checking definitions, we see that its epigraph is a cone. That is, for all $a > 0$, we have:
$(x,t) \in epi(f) \Leftrightarrow f(x) \leq t \Leftrightarrow af(x) = f(ax) \leq at \Leftrightarrow (ax, at) \in epi(f)$

suppose the epigraph is a cone. This means for all $a < 0$, if $(x, t) \in epi(f)$ then $(ax, at) \in epi(f)$. Clearly $(x, f(x)) \in epi(f)$, so $(ax, af(x)) \in epi(f)$, which means that $f(ax) \leq af(x)$ .We get: $f(x) \leq f(ax)/a \leq f(x)$ which means $f(ax) = af(x)$, ie, its positive homogenous.

And last but not least, we say that a fuction $f : R^n \to R$ is polyhedral id its epigraph is a polyhedron.

That is, a function $f : R^n \to R$ is polyhedral if and only if the epigraph $epi(f) := \{(x, t) \in R^{n+1} : t \geq f(x)\}$ can be expressed as a polyhedron: there exist a matrix $C \in R^{mx(n+1)}$ and a vector $d \in R^m$ such that $epi(f) = \{(x, t) \in R^{n+1} : C(x, t) \leq d\}$

in resume If the function is affine, positively homogeneous ($f(\alpha x) = \alpha f(x)$) for $\alpha \geq 0$, and piecewise-affine, respectively.

$- -$

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

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

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

Solution:

We know that the negative entropy is strictly convex and differentiable on $\mathbf{R}^n_{++}$ notice that for $u \neq v$:

$f(u) - f(v) \geq 1^T(u - v)$ but $1^T(u - v) = \sum_{i=1}^n (log(v_i + 1))(u - v)$ therefore $f(u) - f(v) - \nabla f(v)^T(u - v) = D_{\text{kl}}(u, v) \geq 0$

if $u = v$ therefore: $f(u) - f(u) - \nabla f(u)^T(u - u) = 0 - \nabla f(u)^T(0) = 0$

$- -$

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

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

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

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

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

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

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

Solution: a) $f(x) = e^x - 1$ in $\mathbf{R}$. $\operatorname{dom} f = \mathbf{R}$ so $\operatorname{dom} f$ is convex. f is twice differentiable as follows.

$$f'(x) = e^x, \quad f''(x) = e^x$$

Since $f''(x)$ is positive for all $x \in \operatorname{dom} f$ it follows that f is strictly convex and therefore quasiconvex. From the graph you can see that it is quasiconcave but not concave.

b) $f(x_1, x_2) = x_1 x_2$ in $\mathbf{R}^2_{++}$. Since f is twice differentiable, we can compute the Hessian of f as,

$$H = \begin{bmatrix} \mathrm{f}''_{x_1 x_1} & \mathrm{f}''_{x1 x_2} \\ \mathrm{f}''_{x2 x_1} & \mathrm{f}''_{x2 x_2} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Since the eigenvalues of H are not all nonnegative or all nonpositive, H is neither positive semidefinite nor negative semidefinite, so f is neither convex nor concave. Let S_α, with $\alpha \in \mathbf{R}$, be the level sets of f defined as:

$$S_\alpha = \left\{(x_1, x_2) \in \mathbf{R}{+}{+}^2 \mid x_1 x_2 \geq \alpha\right\}.$$

S_α is always convex for all α so f is quasiconcave. f is not quasiconvex.

c) $f(x_1, x_2) = 1/(x_1 x_2)$ in $\mathbf{R}^2_{++}$. Since f is twice differentiable, we compute the Hessian of f as,

$$H = \begin{bmatrix} \mathrm{f}''_{x_1x_1} & \mathrm{f}x_1x_2'' \\ \mathrm{f}''_{x_2x_1} & \mathrm{f}x_2x_2'' \end{bmatrix} = \begin{bmatrix} \frac{2}{x_1^3 x_2} & \frac{1}{x_1^2 x_2^2} \\ \frac{1}{x_1^2 x_2^2} & \frac{2}{x_1 x_2^3} \end{bmatrix} \succeq 0$$

So f is convex. The level sets defined as $\left\{x \in \mathbf{R}{+}{+}^2 \mid 1/x^1 \geq \ \alpha x_2\right\}$ are convex, so f is quasiconvex.

d) $f(x_1, x_2) = x_1/x_2$ in $\mathbf{R}^2_{++}$. The Hessian of f is

$$H = \begin{bmatrix} \mathrm{f}''_{x_1x_1} & \mathrm{f}x_1x_2'' \\ \mathrm{f}''_{x_2x_1} & \mathrm{f}x_2x_2'' \end{bmatrix} = \begin{bmatrix} 0 & \frac{-1}{x_2^2} \\ \frac{-1}{x_2^2} & \frac{2x_1}{x_2^3} \end{bmatrix}$$

Since H is neither positive semidefinite nor negative semidefinite, f is neither convex nor concave. The level sets defined as $\left\{x \in \mathbf{R}{+}{+}^2 \mid x^1 \geq \alpha x_2\right\}$ and $\left\{x \in \mathbf{R}{+}{+}^2 \mid x^1 \leq \ \alpha x_2\right\}$ represent the intersection of two hyperplanes, the line $x_1 = \alpha x_2$ and the ortants, so f is quasiconvex and quasiconcave.

e) $f(x_1, x_2) = x_1^2/x_2$ in $\mathbf{R} \times \mathbf{R}_{++}$. We compute the Hessian of f as

$$H = \begin{bmatrix} \mathrm{f}''_{x_1x_1} & \mathrm{f}''_{x_1x_2} \\ \mathrm{f}''_{x_2x_1} & \mathrm{f}''_{x_2x_2} \end{bmatrix} = \begin{bmatrix} \frac{2}{x_2} & \frac{-2x_1}{x_2^2} \\ \frac{-2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{bmatrix} = \frac{2}{x^2} \begin{bmatrix} 1 \\ \frac{-2x_1}{x_2} \end{bmatrix} \begin{bmatrix} 1 & \frac{-2x_1}{x_2} \end{bmatrix} \succeq 0$$

Thus, f is convex and quasiconvex.

f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \leq \alpha \leq 1$, in $\mathbf{R}^2_{++}$. The Hessian of f is

$$H = \begin{bmatrix} \mathrm{f}''_{x_1x_1} & \mathrm{f}''_{x_1x_2} \\ \mathrm{f}''_{x_2x_1} & \mathrm{f}''_{x_2x_2} \end{bmatrix} = \begin{bmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} & (-\alpha)(1-\alpha)x_1^\alpha x_2^{-\alpha-1} \end{bmatrix}$$
$$= \alpha(1-\alpha)x_1^\alpha x_2^{1-\alpha} \begin{bmatrix} \frac{-1}{x_1^2} & \frac{1}{x_1x_2} \\ \frac{1}{x_1x_2} & \frac{-1}{x_2^2} \end{bmatrix} = -\alpha(1-\alpha)x_1^\alpha x_2^{1-\alpha} \begin{bmatrix} \frac{1}{x_1} \\ \frac{-1}{x_2} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} & \frac{-1}{x_2} \end{bmatrix}$$

It has to

$$\begin{bmatrix} \frac{1}{x_1} \\ \frac{-1}{x_2} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} & \frac{-1}{x_2} \end{bmatrix} \succeq 0$$

and since $0 \leq \alpha \leq 1$, and $x_1, x_2 \in \mathbf{R}^2_{++}$,

$$-\alpha(1-\alpha)x_1^\alpha x_2^{1-\alpha} \leq 0.$$

That is, $H \preceq 0$, so f is concave and quasiconcave.

— —

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

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

Solution:
f is the composition of a norm, which is convex, and an affine function.

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

Solution:
f is the composition of the convex function $h(X) = (detX)^{1/m}$ and an affine transformation. To see that h is convex on S^m_{++}, we restrict h to a line and prove that $g(t) = det(Z + tV)^{1/m}$ is convex: $g(t) = -(detZ)^{1/m}(\prod_{i=1}^m(1 + t\lambda_i))^{1/m}$
where $\lambda_1, ..., \lambda_m$ denote the eigenvalues of $Z^{1/2}VZ^{1/2}$. We have expressed g as the product of a negative constant and the geometric mean of $1 + t\lambda_i$, $i = 1, ..., m$. Therefore g is convex.

(c) $f(X) = \operatorname{tr}\left(A_0 + x_1A_1 + \cdots + x_nA_n\right)^{-1}$, 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.)

Solution:
f is the composition of $\mathbf{tr}X^1$ and an affine transformation

— —

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

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

Solution:
f is the pointwise maximum of k functions $\left\|A^{(i)}x - b^{(i)}\right\|$. Each of those functions is convex because it is the composition of an affine transformation and a norm.

(b) $f(x) = \sum_{i=1}^r |x|_{[i]}$ on $\mathbf{R}^n$, 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, $|x|_{[1]}, |x|_{[2]}, \ldots, |x|_{[n]}$ are the absolute values of the components of x, sorted in nonincreasing order.

Solution:
write f as

$$f(x) = \sum_{i=1}^{r} |x|_{[i]} = max_{1 \leq i_1 < i_2 < \cdots < i_r \leq n} |x_{i_1}| + \cdots + |x_{i_r}|$$

which is the pointwise maximum of $n!/(r!(n-r)!)$ convex functions.

– –

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

Solution:

$g(x) = \log\left(\sum_{i=1}^{m} e^{a_i^T x + b_i}\right)$ is convex because of the composition of log-sum-exp function and an affine mapping so $-g$ is concave. $h(y) = -\log y$ is convex and decreasing. So $f(x) = (g(x))$ is convex.

(b) $f(x, u, v) = -\sqrt{uv - x^T x}$ on $\mathbf{dom}\, f = \left\{(x, u, v) \mid uv > x^T x, u, v > 0\right\}$. Use the fact that $x^T x/u$ is convex in (x, u) for $u > 0$, and that $-\sqrt{x_1 x_2}$ is convex on $\mathbf{R}_{++}^2$.

Solution:

we can express f as $f(x, u, v) = -\sqrt{u(v - x^T x/u)}$ the function $h(x_1, x_2) = -\sqrt{x_1 x_2}$ is convex on $\mathbf{R}_{++}^2$, adn decreasing in each argument. The functions $g_1(u, v, x) = u$ and $g_2(u, v, x) = v - x^T x/u$ are concave. Therefore $f(u, v, x) = h(g(u, v, x))$ is convex

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

Solution:

we can express f as

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

the first term is convex. The function $vx^T x/u$ is concave because v is linear and $x^T x/u$ is convex on $\{(x, u) \mid u > 0\}$. Therefore the second term in f is convex: it is the composition of a convex decreasing functionlog t and a concave function.

(d) $f(x, t) = -\left(t^p - \|x\|_p^p\right)^{1/p}$ where $p > 1$ and $\mathbf{dom}\, f = \{(x, t) \mid t \geq \|x\|p\}$. 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 $\mathbf{R}+^2$ (see exercise 3.16).

Solution:

we can express f as

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

This is the composition of $h(y_1, y_2) = y_1^{1/p} y_2^{1-1/p}$ (convex and decreasing in each argument) and two concave functions

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

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

Solution:
Express f as

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

the first term is convex. The second term is the composition of a decreasing convex function and a concave function, and is also convex.

– –

3.25 Maximum probability distance between distributions. Let $p, q \in \mathbf{R}^n$ represent two probability distributions on $\{1, \ldots, n\}$ (so $p, q \succeq 0, 1^T p = 1^T q = 1$). 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:

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

Here prob (p,C) is the probability of C, under the distribution p, i.e., 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 $\mathbf{R}^n \times \mathbf{R}^n$. (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 $\mathbf{R}^n \times \mathbf{R}^n$.)

First, we can see that d_{mp} is a convex function since it is the max of lineal functions.

Now, for $\|p-q\|_1 = \sum_{i=1}^n |p_i - q_i|$, let's search for a way to first maximize:

$$prob(p,C) - prob(q,C)$$

First, we can see that:

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

Now, let $A = \{1, ..., n\}$ and let $C^* = \{i \in A | p_i > q_i\}$
First, we can ignore any $i \in A$ for which $p_i = q_i$, since for these $i, |p_i - q_i| = 0$
Then, for any $i \in A$ for which $p_i < q_i, p_i - q_i < 0$, so it would be against our objective to maximize the sum to include it.

Then we have that:

$$\sum_{i \in A} p_i - q_i = \sum_{i \in C^*} (p_i - q_i) + \sum_{i \in A \setminus C^*} (p_i - q_i)$$

Now, since

$$\sum_{i \in A} p_i - q_i = \sum_{i \in A} p_i - \sum_{i \in A} q_i = 1 - 1 = 0$$

then:

$$\sum_{i\in C^*}(p_i - q_i) = -(\sum_{i\in A\setminus C^*}(p_i - q_i))$$

Finally:

$$\begin{aligned}
d_{mp}(p,q) &= prob(p,C^*) - prob(q,C^*)\\
&= \sum_{i\in C^*} p_i - q_i\\
&= (1/2)(2 * \sum_{i\in C^*} p_i - q_i)\\
&= (1/2)(\sum_{i\in C^*} p_i - q_i + \sum_{i\in C^*} p_i - q_i)\\
&= (1/2)(\sum_{i\in C^*} p_i - q_i - \sum_{i\in A\setminus C^*} p_i - q_i)\\
&= (1/2)(\sum_{i\in A} |p_i - q_i|)\\
&= (1/2)\|p-q\|_1
\end{aligned}$$

And since it a constant multiple of a norm, we know that it is convex. We know that it's domain is $\mathbf{R}^n \times \mathbf{R}^n$ since it take an ordered pair of two elements in $\mathbf{R}^n$

$- -$

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

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

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

Solution.

By Jensen's inequality, we have for all $x, y \in \operatorname{dom} f$, and $\theta \in [0,1]$,

$$f(y + \theta(x-y)) \le f(y) + \theta(f(x) - f(y)),$$

and hence

$$f(y + \theta(x-y)) - f(y) \le \theta(f(x) - f(y)),$$

then

$$\frac{f(y + \theta(x-y)) - f(y)}{\theta} \le f(x) - f(y),$$

hence

$$f(x) \ge f(y) + \frac{f(y + \theta(x-y)) - f(y)}{\theta}.$$

Now suppose $x \in X_i$. Choose any $y \in \operatorname{int} X_j$, for some j, and take θ sufficiently small so that $y + \theta(x - y) \in X_j$. The above inequality reduces to

$$a_i^T x + b_i \geq a_j^T y + b_j + \frac{\left(a_j^T (y + \theta(x - y)) + b_j - a_j^T y - b_j\right)}{\theta} = a_j^T x + b_j.$$

This is true for any j, so $a_i^T x + b_i \geq \max j = 1, \ldots, L \left(aj^T x + b_j\right)$. We conclude that

$$a_i^T x + b_i = \max j = 1, \ldots, L \left(aj^T x + b_j\right).$$

— —

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 **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.

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

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

Solution:

a) Let f and g be positive, convex functions and let $0 \leq \theta \leq 1$,

$$\begin{aligned}
f(x + (1-\theta)y) \cdot g(x + (1-\theta)y) \leq & (\theta f(x) + (1-\theta) f(y)) \cdot (\theta g(x) + (1-\theta) g(y)) \\
\leq & \theta^2 f(x) g(x) + (1-\theta)^2 f(y) g(y) \\
& + \theta(1-\theta) f(x) g(y) + \theta(1-\theta) f(y) g(x) \\
\leq & (\theta - \theta(1-\theta)) f(x) g(x) + ((1-\theta) - \theta(1-\theta)) f(y) g(y) \\
& + \theta(1-\theta) f(x) g(y) + \theta(1-\theta) f(y) g(x) \\
\leq & \theta f(x) g(x) - \theta(1-\theta) f(x) g(x) \\
& + (1-\theta) f(y) g(y) - \theta(1-\theta) f(y) g(y) \\
& + \theta(1-\theta) f(x) g(y) + \theta(1-\theta) f(y) g(x) \\
\leq & \theta f(x) g(x) + (1-\theta) f(y) g(y) \\
& + (\theta(1-\theta) f(y) - \theta(1-\theta) f(x))(g(x) - g(y)) \\
\leq & \theta f(x) g(x) + (1-\theta) f(y) g(y) \\
& + \theta(1-\theta)(f(y) - f(x))(g(x) - g(y))
\end{aligned}$$

Note that $\theta(1-\theta)(f(y) - f(x))(g(x) - g(y))$ is negative when f and g are both increasing or decreasing, so it is fulfilled that

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

proving the convexity of fg.

b) Let f and g be positive, concave functions and let $0 \leq \theta \leq 1$. By the previous result but taking into account the concavity, we have

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

Since for f and g one non-decreasing and the other non-increasing, $\theta(1-\theta)(f(y)- f(x))(g(x)-g(y))$ is positive, so it is true that

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

Proving the concavity of fg

c) From g we know that $h = 1/g$ is convex, positive and not decreasing, and since f is convex, positive and not increasing, from the result in a) we have that $fh = f/g$ is convex.

$- -$

3.36 Derive the conjugates of the following functions.

(a) Max function. $f(x) = \max i = 1, \ldots, nxi$ on $\mathbf{R}^n$.

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

(c) Piecewise-linear function on $\mathbf{R}$. $f(x) = \max i = 1, \ldots, m\,(aix + b_i)$ on $\mathbf{R}$. 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 $a_ix + b_i$ is redundant, i.e., for each k there is at least one x with $f(x) = a_kx + b_k$. (d) Power function. $f(x) = x^p$ on $\mathbf{R}_{++}$, where $p > 1$. Repeat for $p < 0$.

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

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

Solution

(a)

$$f^*(y) = \begin{cases} 0 \text{ if } y \succeq 0, 1^Ty = 1 \\ \infty \text{ in other case} \end{cases}$$

First suppose y has a negative component, $y_j < 0$, then we chose a vector such that $x_j = -t$ and $x_i = 0$ for $i \neq j$. We can see that:

$$x^Ty - max(x) = -ty_j$$

Then, if we let t go towards infinity, we can see that this goes towards minus infinity, thus y would not be on $dom f^*(x)$

Now, if $y \succeq 0$ but $1^Ty < 0$, we choose x to be a vector of only $-t$ in all positions, we can see that:

$$x^Ty - max(x) = t1^Ty + t$$

Then, if we let t go towards infinity, we can see that this is unbounded above. Similarly, for $1^Ty > 0$ we choose x to be a vector of only t.

Now, for $y \succeq 0, 1^Ty = 1$

$$x^Ty \leq max(x)$$

for any x, thus $x^Ty - max(x) \leq 0$, with equality on $x = 0$. Thus, $f^*(y) = 0$

(b)

$$f^*(y) = \begin{cases} 0 \text{ if } 0 \preceq y \preceq 1, 1^T y = r \\ \infty \ \text{ in other case} \end{cases}$$

First suppose y has a negative component, $y_j < 0$, then we chose a vector x such that $x_j = -t$ and $x_i = 0$ for $i \neq j$. We can see that:

$$x^T y - f(x) = -ty_j$$

Then, if we let t go towards infinity, we can see that this goes towards infinity, thus y would not be on $dom f^*(x)$

Next, suppose y has a component greater than 1, $y_j > 1$, then we chose a vector x such that $x_j = t$ and $x_i = 0$ for $i \neq j$. We can see that:

$$x^T y - f(x) = -ty_j$$

Then, if we let t go towards infinity, we can see that this goes towards infinity, thus y would not be on $dom f^*(x)$

Finally, assume that $1^T x \neq r$. We choose x to be a vector of only t in all positions, we can see that:

$$x^T y - f(x) = t1^T y + tr$$

is unbounded, as t goes towards infinity or minus infinity.

If y satisfies all conditions, we have:

$$x^T y \leq f(x)$$

for all x, with equality on $x = 0$. Thus, $f^*(y) = 0$

(c)

Under the assumption, the graph of f is a piecewise-linear, with breakpoints $(b_i - b_{i+1})/(a_{i+1} - a_{i+})$ for $i = 1, ..., m-1$. Then:

$$f^*(x) = sup(xy - max(a_i x + bi))$$

We can see that $dom f^* = [a_1, a_m]$, and that it is unbounded for anything outside that range. For $a_i \leq y \leq a_{i+1}$ the sup will be on the breakpoint of the segments i and $i+1$, or $(b_i - b_{i+1})/(a_{i+1} - a_{i+})$, thus:

$$f^*(y) = -b_i - (b_{i+1} - b_i) * (\frac{y - a_i}{a_{i+1} - a_{i+}})$$

Hence, the graph of $f^*(x)$ is a piecewise-linear curve connecting the points $(a_i, -b_i)$ for $i = 1, ..., m$

(d)

Let's define $q = p/(p-1)$

If $p > 1$, x^p is convex on $\mathbb{R}_{++}$. For $y < 0$ the function $yx - x^p$ achieves its maximum for $x > 0$ at $x = 0$, so $f^*(y) = 0$. For $y > 0$, the function achieves its maximum at $x = (y/p)^{1/(p-1)}$, with:

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

Thus,

$$f^*(y) = \begin{cases} 0 \text{ if} y \leq 0 \\ (p-1)(y/p)^q & \text{in other case} \end{cases}$$

(e)

Let

$$f^*(y) = \begin{cases} 0 \text{ if } y \preceq 0, (\prod_i(-y_i))^(1/n) \geq 1/n \\ \infty & \text{in other case} \end{cases}$$

First, suppose y has a positive component, $y_j > 0$, then we chose a vector x such that $x_j = t$ and $x_i = 1$ for $i \neq j$. We can see that:

$$x^T y - f(x) = ty_j + \sum_{i \neq j} y_i - t^{1/n}$$

is unbounded when $t > 0$.

Next, suppose that $y \preceq 0$, but $(\prod_i(-y_i))^(1/n) < 1/n$. Choose $x_i = -t/y_i$, and then:

$$x^T y - f(x) = -tn - t(\prod_i (-1/y_i))^{1/n}$$

We can see that this approaches infinity as t approaches infinity. Finally, assume that both conditions are fulfilled. Then, for $x \succeq 0$, the arithmetic geometric mean inequality states that:

$$\frac{x^T y}{n} \geq (\prod_i (-y_i x_i))^{1/n} \geq 1/n * (\prod_i (x_i))^{1/n}$$

Thus, $x^T y \geq f(x)$ with equality on $x_i = -1/y_i$. Hence, $f^*(y) = 0$

(f)

Let $f^*(y, u) = -2 + log4 - log(u^2 - y^T y)$, with $dom f^* = \{(y, u) | \|y\|_2 < -u\}$

Suppose $\|y\|_2 \geq -u$. Choose $x = st$, $t = s(\|x\|_2 + 1) > s\|y\|_2 \geq -s$, for $s \geq 0$. Then,

$$y^T x + tu > sy^T y - su^2 = s(u^2 - y^T y) \geq 0$$

While, $y^x + tu$ goes to infinity, $-log(t^2 - x^T x)$ goes towards minus infinity at a slower (logaritmic) rate, therefore:

$$y^T x + tu + log(t^2 - x^T x)$$

is unbounded.

Now, suppose $\|y\|_2 < u$. If we do the cross derivate with respect to x, t and then solve for x, t, we can see that the function reaches a maximum at:

$$x = \frac{2y}{u^2 - y^T y}, \; t = -\frac{2u}{u^2 - y^T y}$$

Then,

$$\begin{aligned} f^*(y,u) &= ut + y^T x + log(t^2 - x^T x) \\ &= -2 + log4 - log(y^2 - u^{t+1}) \end{aligned}$$

−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−

3.49 Show that the following functions are log-concave.

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

Solution:
We have

$$\log\left(e^x/(1+e^x)\right) = x - \log\left(1+e^x\right).$$

The first term is linear, hence concave. Since the function $\log(1+e^x)$ is convex (it is the log-sum-exp function, evaluated at $x_1 = 0, x_2 = x$), the second term above is concave. So $e^x/(1+e^x)$ is long-concave

(b) Harmonic mean:

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

Solution:
The first and second derivatives of

$$h(x) = \log f(x) = -\log(1/x_1 + ... + 1/x_n).$$

are

$$\begin{aligned} \tfrac{\partial h(x)}{\partial x_i} &= \tfrac{1/x_i^2}{1/x_1+...+1/x_n} \\ \tfrac{\partial^2 h(x)}{\partial x_i^2} &= \tfrac{-2/x_i^3}{1/x_1+...+1/x_n} + \tfrac{1/x_i^4}{(1/x_1+...+1/x_n)^2} \\ \tfrac{\partial^2 h(x)}{\partial x_i \partial x_j} &= \tfrac{1/(x_i^2 x_j^2)}{(1/x_1+...+1/x_n)^2} \end{aligned}$$

we show that $y^T \nabla h(x) y \prec 0, y \neq 0$

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

This follows from the Cauchy-Schwarz inequality $(a^Tb)^2 \leq \|a\|_2^2\|b\|_2^2$
(c) Product over sum:

$$f(x) = \frac{\prod_{i=1}^n x_i}{\sum_{i=1}^n x_i}, \quad \mathrm{dom}\, f = \mathbf{R}_{+++}^n$$

(d) Determinant over trace:

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

– –

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 $x < 0$.

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

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

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

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

Solution: (a)

$$\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \\
f'(x) &= \frac{d}{dx} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt \\
&= \frac{1}{\sqrt{2\pi}} \frac{d}{dx} \int_{-\infty}^{x} e^{-t^2/2} dt \\
&= \frac{1}{\sqrt{2\pi}} (e^{-x^2/2}) \text{ By Calculus Fundamental Theorem} \\
f'(x)^2 &= (\frac{e^{-x^2/2}}{\sqrt{2\pi}})^2 \\
&= \frac{e^{-2x^2/2}}{2\pi} \\
&= \frac{e^{-x^2}}{2\pi} \\
f''(x) &= \frac{d}{dx} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \\
&= \frac{\frac{d}{dx} e^{-x^2/2}}{\sqrt{2\pi}} \\
&= \frac{e^{-x^2/2} * \frac{d}{dx} \frac{-x^2}{2}}{\sqrt{2\pi}} \\
&= \frac{e^{-x^2/2} * -x}{\sqrt{2\pi}} \\
&= \frac{-xe^{-x^2/2}}{\sqrt{2\pi}}
\end{aligned}$$

Now, we have:

$$f''(x) f(x) = \frac{-xe^{-x^2/2}}{\sqrt{2\pi}} * \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

Since $e^{-x^2/2} > 0$ for all $x \in \mathbb{R}$, then $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt > 0$ for all $x \in \mathbb{R}$. And since $-x < 0$ for all $x > 0$, we know that $f''(x)f(x) < 0$ for all $x > 0$

Then, since $x^2 \geq 0$ for all $x \in \mathbb{R}$ and $f'(x) \in \mathbb{R}$ for any $x \in \mathbb{R}$, it follows that $f'(x) \geq 0$ for all $x \in \mathbb{R}$.

Then, it immediately follows that $f''(x)f(x) \leq f'(x)^2$ for all $x > 0$

(b)
Let $g(t) = t^2/2$. We know that t^2 is convex, thus $g(t)$ is convex. Now, know that for convex functions:

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

Thus, we have:

$$\begin{aligned} g(t) &\geq g(x) + g'(x)(t - x) \\ t^2/2 &\geq x^2/2 + (2x/2)(t - x) \\ &\geq x^2/2 + x(t - x) \\ &\geq x^2/2 + xt - x^2 \\ &\geq -x^2/2 + xt \end{aligned}$$

Q.E.D
(c)

$$\begin{aligned} t^2/2 &\geq -x^2/2 + xt \\ -t^2/2 &\leq x^2/2 - xt \\ e^{-t^2/2} &\leq e^{x^2/2 - xt} \\ e^{-t^2/2} &\leq e^{x^2/2} e^{-xt} \\ \int_{-\infty}^{x} e^{-t^2/2} dt &\leq \int_{-\infty}^{x} e^{x^2/2} e^{-xt} dt \\ \int_{-\infty}^{x} e^{-t^2/2} dt &\leq e^{x^2/2} \int_{-\infty}^{x} e^{-xt} dt \end{aligned}$$

(d)

From (a) we know that the inequality transforms into:

$$
\begin{aligned}
\frac{-xe^{-x^2/2}}{\sqrt{2\pi}} * \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt &\leq \frac{e^{-x^2}}{2\pi} \\
\frac{-xe^{-x^2/2}}{2\pi} \int_{-\infty}^{x} e^{-t^2/2} dt &\leq \frac{e^{-x^2}}{2\pi} \\
-xe^{-x^2/2} \int_{-\infty}^{x} e^{-t^2/2} dt &\leq e^{-x^2} \\
\int_{-\infty}^{x} e^{-t^2/2} dt &\leq \frac{e^{-x^2}}{-xe^{-x^2/2}} \\
\int_{-\infty}^{x} e^{-t^2/2} dt &\leq \frac{e^{-x^2/2}}{-x} \\
\int_{-\infty}^{x} e^{-t^2/2} dt &\leq e^{x^2/2} \frac{e^{-x^2}}{-x}
\end{aligned}
$$

Now, since we know that

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

Q.E.D