Convex functions - Exercises Chapter 3

David Ricardo Pedraza Silva   Natalia Monroy Rosas

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 $x|f(x) = 1$, etc.
Could f be convex (concave, quasiconvex, quasiconcave)? Explain your answer. Repeat for the level curves shown below.

Solution:

The first function could be quasiconvex because the sublevel sets appear to be convex. It is definitely not concave or quasiconcave because the superlevel sets are not convex. It is also not convex, for the following reason. We plot the function values along the dashed line labeled I.

3.4

Show that a continuous function $f : R^n \to R$ 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, $y \in R^n$

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

Solution:

Proof. $\rightarrow$ Let's assume f is convex, therefore:

$$f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x)), \text{ for } 0 \leq \lambda \leq 1$$

Integrating both sides from 0 to 1, we obtain:

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

$\leftarrow$ Now we prove the converse by supposing f is not convex. Then there exists x, y and θ_0 such that:

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

Let's take the function g of θ defined as:

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

This function is continuous since f is. g is positive on the interval θ_0, and for $\theta = 0$ and $\theta = 1$ its value is zero. Now, consider α to be the greatest intercept of g below θ_0 and β the smallest above θ_0. Let's take $u = \alpha x + (1-\alpha)y$ and $v = \beta x + (1-\beta)y$. On the interval (α, β), since the function is not convex we have:

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

Therefore for the interval $\theta \in (0,1)$, we get:

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

Taking the integral on both sides from 0 to 1 gives:

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

Thus for every point on the interval $[u, v]$ the average of its image in f is always greater that the average of its values at the endpoints.

Now we have both directions of the statement proved, it is demonstrated that a continuous function $f : R^n \to R$ 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.

$\square$

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:

Halfspace

When the function is convex, and it is affine.

Convex Cone

positively homogeneous $(f(\alpha x) = \alpha f(x) for \alpha \geq 0)$

Polyhedron

when it is piecewise-affine

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

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

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

Solution:

Proof. It is known that the negative entropy is convex and differentiable on $\mathbb{R}^n_{++}$, therefore:

$$f(u) > f(v) + \nabla f(v)^T(u - v)$$

for all $u, v \in \mathbb{R}^n_{++}$ with $u \neq v$. Evaluating both sides of the inequality, we obtain:

$$\begin{aligned}\sum_{i=1}^n u_i \log u_i &> \sum_{i=1}^n v_i \log v_i + \sum_{i=1}^n (\log v_i + 1)(u_i - v_i) \\ &= \sum_{i=1}^n v_i \log v_i + 1^T(u - v)\end{aligned}$$

Thus, re-arranging the inequality to obtain the Kullback-Leibler divergence we get:

$$D_{kl}(u, v) = \sum_{i=1}^n u_i \log u_i - \sum_{i=1}^n v_i \log v_i + 1^T(u - v) > 0$$

From the previous equation it is evident that $D_{kl}(u, v) = 0$ if and only if $u = v$. □

3.16

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

(a) $f(x) = e^x - 1$ on R

Solution. This function is strictly convex, and therefore it is quasiconvex. Also quasiconcave but not concave.

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

Solution. For $f(x_1, x_2) = x_1 x_2$ the Hessian is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which is neither positive semidefinite nor negative semidefinite. Thus, f is not concave nor convex. It is quasiconcave, since its superlevel sets are convex.

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

Solution. For $f(x_1, x_2) = \frac{1}{(x_1 x_2)}$. The Hessian is:

$$\frac{1}{x_1 x_2} \begin{pmatrix} \frac{2}{x_1^2} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^2} \end{pmatrix}$$

f is convex and quasiconvex. It is not quasiconcave or concave.

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

Solution. For $f(x_1, x_2) = \frac{x_1}{x_2}$. The Hessian is:

$$\begin{pmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{pmatrix}$$

It is quasiconvex and quasiconcave since the sublevel and superlevel sets are halfspaces.

(e) $f(x_1, x_2) = \frac{x_1^2}{x_2}$ on $R \times R_{++}$

Solution. The Hessian for this function is

$$\begin{pmatrix} \frac{2}{x_2} & -\frac{2x_1}{x_2^2} \\ -\frac{2x_1}{x_2^2} & \frac{2x_1^2}{x_2^3} \end{pmatrix}$$

f is convex and quasiconvex. It is not concave or quasiconcave.

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

Solution.

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

f is not convex or quasiconvex.

3.20

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

(a) $f(x) = \|Ax - b\|$, where $A \in R^{mn}$, $b \in R^m$, and $\|\cdot\|$ is a norm on R^m.
Solution. In this case, f is the composition of a convex norm and an affine function, therefore is convex.

(b) $f(x) = -(\det(A_0 + x_1A_1 + \cdots + x_nA_n))^{1/m}$, on $\{x|A_0 + x_1A_1 + \cdots + x_nA_n \prec 0\}$, where $A_i \in S_m$.
Solution. f is the composition of the convex function $g(X) = -(\det X)^{\frac{1}{m}}$ and an affine transformation. Let's verify that $g(X)$ is convex on S^m_{++}. We can restrict it to a line and prove that $h(t) = -\det(Z + tV)^{\frac{1}{m}}$ is convex:

$$\begin{aligned}
h(t &= -(\det(Z + tV))^{\frac{1}{m}} \\
&= -(\det(Z))^{\frac{1}{m}}(\det(I + tZ^{-\frac{1}{2}}VZ^{-\frac{1}{2}}))^{\frac{1}{m}} \\
&= -(\det(Z))^{\frac{1}{m}}(\prod_{i=1}^{m}(1 + t\lambda_i))
\end{aligned}$$

Here all λ_i represent the eigenvalues of $Z^{-\frac{1}{2}}VZ^{-\frac{1}{2}}$. As h is expressed as the product of a negative constant and the geometric mean of $1 + \lambda_i$, $i = 1, \cdots, m.$, we can conlude g is convex, therefore f is convex as well.

(c) $f(X) = \mathbf{tr}(A_0 + x_1A_1 + \cdots + x_nA_n)^{-1}$, on $\{x|A_0 + x_1A_1 + \cdots + x_nA_n \prec 0\}$, where $A_i \in S^m$.
Solution. f is the composition of $\mathbf{tr}(X^{-1})$ and an affine transformation:

$$x \to A_0 + x_1A_1 + \cdots + x_nA_n$$

3.21

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

a.

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

b.

$$f(x) = \sum |x|_i = \max |x_{i1}| + \cdots + |x_{ir}|$$

this is the point-wise maximum of $\binom{n}{r}$ convex functions.

3.22

Composition rules. Show that the following functions are convex.

a.

As $\log(\sum_{i=1}^{n} e^{y_i})$ is convex, $g(x) = \log(\sum_{i=1}^{m} e^{a_i^t})$ is also convex since it is a composition of logarithmic, sum and exponential functions, with an affine mapping. Therefore, we have that g is concave. The function $h(y) = -\log y$ is convex and decreasing. Hence $f(x) = h(-(x))$ is also convex.

b.

We can take f as $f(x, u, v) = -\sqrt{u(v - x^T \frac{x}{u})}$. The function $h(x_1, x_2) = -\sqrt{x_1 x_2}$ is convex on R^2_{++}. The functions $g_1(u, v, x) = u$ and $g2(u, v, x) = v - x^T \frac{x}{u}$ are concave. Therefore $f(x, u, v) = h(g(x, u, v))$ is convex.

c.

f can be expressed as $f(x, u, v) = -\log u - \log(v - x^T x \frac{x}{u})$, we know that $-logu$ is conevex. The function $v - x^T x \frac{x}{u}$ is concave because v is linear and $x^T x \frac{x}{u}$ is convex on $(x, u)|u > 0$. Then, the second term of f is convex and f is convex in consequence.

d.

We can express f as:

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

Which is the composition of the convex and decreasing function $h(y_1, y_2) = -y_1^{\frac{1}{p}} y_2^{1-\frac{1}{p}}$ and two concave functions: $g_1(x, 1) = t^{1-\frac{1}{p}}$ and $g_2(x, t) = t - \frac{\|x\|_p^p}{t^{p-1}}$.

e.

Let's take f as

$$\begin{aligned} f(x, t) &= -\log t^{p-1} - \log(t - \frac{\|x\|_p^p}{t^{p-1}}) \\ &= -(p - 1)\log t - \log(t - \frac{\|x\|_p^p}{t^{p-1}}). \end{aligned}$$

This way it is evident that the first term is convex and the second term is the composition of a decreasing convex function and a concave function, and is also convex. Therefore, f is convex.

3.25

Let $V_c \in R^n$ such that $V_{ci} = 1$ if $i \in C$ and $V_{ci} = 0$ otherwise. Node that:

$$d_{mp}(p, q) = \max \|V_c^T p - V_c^T q\|_1 = \max \|V_i\|_1 \|p - q\|_1$$

Which means:

$$d_{mp}(p, q) = \max |C| \|p - q\|_1 = n\|p - q\|_1$$

Let x, y be probability distributions as specified. Observe that:

$$d_{mp}(\lambda x + (1 - \lambda)y) = n\|\lambda(x_1 - x_2) - (1 - \lambda x)(y_1 - y_2)\|$$

note that:

$$n\|\lambda(x_1 - x_2) - (1 - \lambda x)(y_1 - y_2)\| \leq n\lambda\|x_1 - x_2\|_1 + n(1 - \lambda)\|y_1 - y_2\|_1$$

since:

$$n\lambda\|x_1 - x_2\|_1 + n(1 - \lambda)\|y_1 - y_2\|_1 = \lambda d_{mp}(x) + (1 - \lambda)d_{mp}(y)$$

We conclude d_{mp} is convex.

3.29

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

$$R^n = X_1 \cup X_2 \cup \cdots X_L$$

where $\mathbf{int} X_i \neq \emptyset$ and $\mathbf{int} X_j \neq \emptyset$ for $i \neq j$, and a family of affine functions $a_i^T x + b_1, ..., a_L^T x + b_L$ such that $f(x) = a_i^T x + b_i$ for $x \in X_i$.
Show that this means that $f(x) = max a_i^T x + b_1, ..., a_L^T x + b_L$.

Solution. Taking Jensen's inequality, we have:

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

Therefore

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

Now we assume that $x \in X_i$. Let's take any $y \in \mathbf{int} X_j$, for some j, and some t small enough such that $y + t(x-y) \in X_j$. The above inequality reduces to

$$a_i^T x + b_i \geq a_j^T y + b_j + \frac{(a_j^T(y+t(x-y)) + b_j - a_j^T y - b_i)}{t} = a_j^T x + b_j$$

Since this is true for any j, so that $a_i^T x + b_i \geq \max_{j=1,\ldots,L}(a_j^T x + b_j)$. We can conclude that:

$$a_i^T x + b_i = \max_{j=1,\ldots,L}(a_j^T x + b_j).$$

3.32

a.

let f and g be convex, let $0 \leq \lambda \leq 1$

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

And since:

$$(\lambda f(x)+(1-\lambda)f(y))(\lambda g(x)+(1-\lambda)g(y)) = \lambda f(x)g(x)+(1-\lambda)f(y)g(y)+\lambda(1-\lambda)(f(y)-f(x))(g(x)-g(y))$$

We conclude fg is convex

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

b.

Analogous to a

c.

Note that $\frac{1}{g}$ is positive and convex. It follows from a. that $\frac{f}{g}$ if convex.

3.36

Derive the conjugates of the following functions.

(a) *Max function.* $f(x) = \max_{i=1,\ldots,n} x_i$ on $\mathbf{R}^n$.
Solution. It will be proved that

$$f^*(y) = \begin{cases} 0 & \text{if } y \succeq 0, \quad 1^T y = 1 \\ \infty & \text{otherwise.} \end{cases}$$

First, we will verify what is the domain of f^*. Let's suppose y has a negative component $y_k < 0$. If we choose a vector x with $x_k = -t, x_i = 0$ for $i \neq k$, and let t go to infinity, we see that

$$x^T y - \max_i x_i = -ty_k \to \infty,$$

so y is not in **dom** f^*. Next, assume $y \succeq 0$ but $\mathbf{1}^T y > 1$. We choose $x = t\mathbf{1}$ and let t go to infinity, to show that

$$x^T y - \max_i x_i = t\mathbf{1}^T y - t$$

is unbounded above. Similarly, when $y \succeq 0$ and $\mathbf{1}^T y < \mathbf{1}$, we choose $x = -t\mathbf{1}$ and let t go to infinity. The remaining case for y is $y \succeq 0$ and $\mathbf{1}^T y = \mathbf{1}$. Here we have

$$x^T y \leq \max_i x_i$$

for all x, and therefore $x^T y - \max_i x_i \leq 0$ for all x, with equality for $x = 0$. Therefore $f^*(y) = 0$.

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

$$f^*(y) = \begin{cases} 0 & 0 \preceq y \preceq \mathbf{1}, \quad \mathbf{1}^T y = r \\ \infty & \text{otherwise} \ , \end{cases}$$

We first verify the domain of f^*. Suppose y has a negative component, say $y_k < 0$. If we choose a vector x with $x_k = -t, x_i = 0$ for $i \neq k$, and let t go to infinity, we see that

$$x^T y - f(x) = -ty_k \to \infty,$$

so y is not in **dom** f^*. Next, suppose y has a component greater than 1 , say $y_k > 1$. If we choose a vector x with $x_k = t, x_i = 0$ for $i \neq k$, and let t go to infinity, we see that

$$x^T y - f(x) = ty_k - t \to \infty,$$

so y is not in **dom** f^*. Finally, assume that $\mathbf{1}^T x \neq r$. We choose $x = t\mathbf{1}$ and find that

$$x^T y - f(x) = t\mathbf{1}^T y - tr$$

is unbounded above, as $t \to \infty$ or $t \to -\infty$.
If y satisfies all the conditions we have

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

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

(c) *Piecewise-linear function on* $\mathbf{R}$. $f(x) = \max_{i=1,\ldots,m}(a_i x + b_i)$ 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 $a_i x + b_i$ is redundant, i.e., for each k there is at least one x with $f(x) = a_k x + b_k$.
Solution. Under the assumption, the graph of f is a piecewise-linear, with breakpoints $(b_i - b_{i+1})\,/\,(a_{i+1} - a_i)\,, i = 1, \ldots, m-1$. We can write f^* as

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

We see that $\mathbf{dom}\ f^* = [a_1, a_m]$, since for y outside that range, the expression inside the supremum is unbounded above. For $a_i \leq y \leq a_{i+1}$, the supremum in the definition of f^* is reached at the breakpoint between the segments i and $i+1$, i.e., at the point $(b_{i+1} - b_i)\,/\,(a_{i+1} - a_i)$, so we obtain

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

where i is defined by $a_i \leq y \leq a_{i+1}$. Hence the graph of f^* is also a piecewise-linear curve connecting the points $(a_i, -b_i)$ for $i = 1, \ldots, m$. Geometrically, the epigraph of f^* is the epigraphical hull of the points $(a_i, -b_i)$.

(d) *Power function.* $f(x) = x^p$ on $\mathbf{R}_{++}$, where $p > 1$. Repeat for $p < 0$.
Solution. We'll use standard notation: we define q by the equation $1/p + 1/q = 1$, i.e., $q = p/(p-1)$. We start with the case $p > 1$. Then x^p is strictly convex on $\mathbf{R}_+$. 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 $x = (y/p)^{1/(p-1)}$, where it has value

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

Therefore we have

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

For $p < 0$ similar arguments show that $\mathbf{dom}\ f^* = -\mathbf{R}_{++}$and $f^*(y) = \frac{-p}{q}(-y/p)^q$.

(e) *Geometric mean.* $f(x) = -\left(\prod x_i\right)^{1/n}$ on $\mathbf{R}^n_{++}$.
Solution. The conjugate function is

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

We first verify the domain of f^*. Assume y has a positive component, say $y_k > 0$. Then we can choose $x_k = t$ and $x_i = 1, i \neq k$, to show that

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

is unbounded above as a function of $t > 0$. Hence the condition $y \preceq 0$ is indeed required. Next assume that $y \preceq 0$, but $\left(\prod_i (-y_i)\right)^{1/n} < 1/n$. We choose $x_i = -t/y_i$, and obtain

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

as $t \to \infty$. This demonstrates that the second condition for the domain of f^* is also needed. Now assume that $y \preceq 0$ and $\left(\prod_i (-y_i)\right)^{1/n} \geq 1/n$, and $x \succeq 0$. The arithmeticgeometric mean inequality states that

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

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

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

Solution.

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

We first verify the domain. Suppose $\|y\|_2 \geq -u$. Choose $x = sy, t = s\left(\|x\|_2 + 1\right) > s\|y\|_2 \geq -su$, with $s \geq 0$. Then

$$y^T x + tu > s y^T y - s u^2 = s\left(u^2 - y^T y\right) \geq 0,$$

so $y^x + tu$ goes to infinity, at a linear rate, while the function $-\log\left(t^2 - x^T x\right)$ goes to $-\infty$ as $-\log s$. Therefore

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

is unbounded above. Next, assume that $\|y\|_2 < u$. Setting the derivative of

$$y^T x + ut + \log\left(t^2 - x^T x\right)$$

with respect to x and t equal to zero, and solving for t and x we see that the maximizer is

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

This gives

$$\begin{aligned} f^*(y,u) &= ut + y^T x + \log\left(t^2 - x^T x\right) \\ &= -2 + \log 4 - \log\left(y^2 - u^t u\right) \end{aligned}$$

3.49

1.

let $f(x) = \frac{e^x}{1+e^x}$ and $g = log \circ f$. Note that

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

x is concave and so is and so is $-\ln(1 + e^x)$, we conclude $g(x)$ is concave.

2.

for $x, y \in dom f$ we get that:

$$(\sum \frac{y_i}{x_i^2})^2 < 2(\sum \frac{1}{x_i})(\sum \frac{y_i^2}{x_i^3})$$

that is:

$$y^T \nabla^2 \log f(x) y < 0$$

3.

$$f(x) = \sum \log x_i - \log \sum x_i$$

Let

$$g(t) = \sum \log(xi + tvi) - \log \sum (xi + tvi)$$

$$g'(t) = \sum \frac{v_i}{x_i + tv_i} - \frac{1^T v}{1^T x + t1^T v}$$

$$g''(t) = \sum \frac{v_i^2}{(x_i + tv_i)^2} + \frac{(1^T v)^2}{(1^T x + t1^T v)^2}$$

We must show that

$$g''(0) \leq 0$$

Which is to say

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

holds whenever $1^T v = 1^T x$. The optimality condition is

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

so we have $v_i = \lambda x_i^2$. From $1^T v = 1^T x$ we can obtain λ, which gives

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

Therefore the minimum value of $\sum_i v_i^2/x_i^2$ over $1^T v = 1^T x$ is

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

because $\|x\|_2 \leq \|x\|_1$. This proves the inequality.

4.

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

Solution. We prove that

$$h(X) = \log f(X) = \log\det X - \log\operatorname{tr} X$$

is concave. Consider the restriction on a line $X = Z + tV$ with $Z \succ 0$, and use the eigenvalue decomposition $Z^{-1/2} V Z^{-1/2} = Q\Lambda Q^T = \sum_{i=1}^n \lambda_i q_i q_i^T$:

$$\begin{aligned}
h(Z+tV) &= \log\det(Z+tV) - \log\operatorname{tr}(Z+tV) \\
&= \log\det Z - \log\det\left(I + tZ^{-1/2} V Z^{-1/2}\right) - \log\operatorname{tr} Z\left(I + tZ^{-1/2} V Z^{1/2}\right) \\
&= \log\det Z - \sum_{i=1}^n \log\left(1 + t\lambda_i\right) - \log\sum_{i=1}^n \left(q_i^T Z q_i\right)\left(1 + t\lambda_i\right)\Bigg) \\
&= \log\det Z + \sum_{i=1}^n \log\left(q_i^T Z q_i\right) - \sum_{i=1}^n \log\left(\left(q_i^T Z q_i\right)\left(1 + t\lambda_i\right)\right) \\
&\quad - \log\sum_{i=1}^n \left(\left(q_i^T Z q_i\right)\left(1 + t\lambda_i\right)\right),
\end{aligned}$$

which is a constant, plus the function

$$\sum_{i=1}^n \log y_i - \log\sum_{i=1}^n y_i$$

evaluated at $y_i = \left(q_i^T Z q_i\right)\left(1 + t\lambda_i\right)$.

3.54

Solution. The derivatives of f are

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

a.

$f''(x) \leq 0$ for $x \geq 0$.

b.

Since $t^2/2$ is convex we have

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

This is the general inequality

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

which holds for any differentiable convex function, applied to $g(t) = t^2/2$.

c.

It is immediate by taking exponentials and integrating.

d.

This basic inequality reduces to

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

i.e.,

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

This follows from part (c) because

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