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.

Along this line the function passes through the points marked as black dots in the figure below. Clearly along this line segment, the function is not convex.
If we repeat the same analysis for the second function, we see that it could be concave (and therefore it could be quasiconcave). It cannot be convex or quasiconvex, because the sublevel sets are not convex.

3.3 Inverse of an increasing convex function. Suppose \( f: \mathbb{R} \to \mathbb{R} \) is increasing and convex on its domain \((a, b)\). Let \( g \) denote its inverse, i.e., the function with domain \((f(a), f(b))\) and \( g(f(x)) = x \) for \( a < x < b \). What can you say about convexity or concavity of \( g \)?

Solution. \( g \) is concave. Its hypograph is

\[
\text{hypo} g = \{(y, t) \mid t \leq g(y)\}
= \{(y, t) \mid f(t) \leq f(g(y))\} \quad \text{(because } f \text{ is increasing)}
= \{(y, t) \mid f(t) \leq y\}
= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \text{epi } f.
\]

For differentiable \( g, f \), we can also prove the result as follows. Differentiate \( g(f(x)) = x \) once to get

\[ g'(f(x)) = 1/f'(x). \]

so \( g \) is increasing. Differentiate again to get

\[ g''(f(x)) = -\frac{f''(x)}{f'(x)^3}, \]

so \( g \) is concave.

3.4 [RV73, page 15] 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 \( x, y \in \mathbb{R}^n \),

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

Solution. First suppose that \( f \) is convex. Jensen’s inequality can be written as

\[ f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x)) \]

for \( 0 \leq \lambda \leq 1 \). Integrating both sides from 0 to 1 we get

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

Now we show the converse. Suppose \( f \) is not convex. Then there are \( x \) and \( y \) and \( \theta_0 \in (0, 1) \) such that

\[ f(\theta_0 x + (1 - \theta_0)y) > \theta_0 f(x) + (1 - \theta_0)f(y). \]
Consider the function of $\theta$ given by
\[ F(\theta) = f(\theta x + (1 - \theta)y) - \theta f(x) - (1 - \theta)f(y), \]
which is continuous since $f$ is. Note that $F$ is zero for $\theta = 0$ and $\theta = 1$, and positive at $\theta_0$.
Let $\alpha$ be the largest zero crossing of $F$ below $\theta_0$ and let $\beta$ be the smallest zero crossing of $F$ above $\theta_0$. Define $u = \alpha x + (1 - \alpha)y$ and $v = \beta x + (1 - \beta)y$. On the interval $(\alpha, \beta)$, we have
\[ F(\theta) = f(\theta x + (1 - \theta)y) > \theta f(x) + (1 - \theta)f(y), \]
so for $\theta \in (0, 1)$,
\[ f(\theta u + (1 - \theta)v) > \theta f(u) + (1 - \theta)f(v). \]
Integrating this expression from $\theta = 0$ to $\theta = 1$ yields
\[ \int_0^1 f(u + \theta(u - v)) \, d\theta > \int_0^1 (f(u) + \theta(f(u) - f(v))) \, d\theta = \frac{f(u) + f(v)}{2}. \]
In other words, the average of $f$ over the interval $[u, v]$ exceeds the average of its values at the endpoints. This proves the converse.

3.5 [RV73, page 22] \textit{Running average of a convex function.} Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is convex, with $\mathbb{R}_+ \subseteq \text{dom} \ f$. Show that its running average $F$, defined as
\[ F(x) = \frac{1}{x} \int_0^x f(t) \, dt, \quad \text{dom} \ F = \mathbb{R}_{++}, \]
is convex. You can assume $f$ is differentiable.
\textbf{Solution.} $F$ is differentiable with
\[
F'(x) = -(1/x^2) \int_0^x f(t) \, dt + f(x)/x \quad \text{and} \quad F''(x) = (2/x^3) \int_0^x f(t) \, dt - 2f(x)/x^2 + f'(x)/x.
\]
Convexity now follows from the fact that $f(t) \geq f(x) + f'(x)(t-x)$ for all $x, t \in \text{dom} \ f$, which implies $F''(x) \geq 0$.

3.6 \textit{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?
\textbf{Solution.} If the function is affine, positively homogeneous $(f(\alpha x) = \alpha f(x)$ for $\alpha \geq 0)$, and piecewise-affine, respectively.

3.7 Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex with $\text{dom} \ f = \mathbb{R}^n$, and bounded above on $\mathbb{R}^n$. Show that $f$ is constant.
\textbf{Solution.} Suppose $f$ is not constant, i.e., there exist $x, y$ with $f(x) < f(y)$. The function
\[ g(t) = f(x + t(y-x)) \]
is convex, with $g(0) < g(1)$. By Jensen’s inequality
\[ g(1) \leq \frac{t-1}{t} g(0) + \frac{1}{t} g(t) \]
for all $t > 1$, and therefore
\[ g(t) \geq t g(1) - (t-1) g(0) = g(0) + t (g(1) - g(0)), \]
so $g$ grows unboundedly as $t \rightarrow \infty$. This contradicts our assumption that $f$ is bounded.
(a) The Hessian of \( \tilde{f} \) must be positive semidefinite everywhere:
\[

\nabla^2 \tilde{f}(z) = F^T \nabla^2 f(Fz + \hat{x})F \succeq 0.
\]

(b) The condition in (a) means that \( v^T \nabla^2 f(Fz + \hat{x})v \geq 0 \) for all \( v \) with \( Av = 0 \), i.e.,
\[ v^T A^T A v = 0 \implies v^T \nabla^2 f(Fz + \hat{x})v \geq 0. \]

The result immediately follows from the hint.

3.10 An extension of Jensen’s inequality. One interpretation of Jensen’s inequality is that randomization or dithering hurts, i.e., raises the average value of a convex function: For \( f \) convex and \( v \) a zero mean random variable, we have \( \mathbf{E} f(x_0 + v) \geq f(x_0) \). This leads to the following conjecture. If \( f_0 \) is convex, then the larger the variance of \( v \), the larger \( \mathbf{E} f(x_0 + v) \).

(a) Give a counterexample that shows that this conjecture is false. Find zero mean random variables \( v \) and \( w \), with \( \text{var}(v) > \text{var}(w) \), a convex function \( f \), and a point \( x_0 \), such that \( \mathbf{E} f(x_0 + v) < \mathbf{E} f(x_0 + w) \).

(b) The conjecture is true when \( v \) and \( w \) are scaled versions of each other. Show that \( \mathbf{E} f(x_0 + tv) \) is monotone increasing in \( t \geq 0 \), when \( f \) is convex and \( v \) is zero mean.

Solution.

(a) Define \( f: \mathbb{R} \rightarrow \mathbb{R} \) as
\[
f(x) = \begin{cases} 
0, & x \leq 0 \\
x, & x > 0,
\end{cases}
\]
x_0 = 0, and scalar random variables
\[
w = \begin{cases} 
1 & \text{with probability } 1/2 \\
-1 & \text{with probability } 1/2
\end{cases} \quad v = \begin{cases} 
4 & \text{with probability } 1/10 \\
-4/9 & \text{with probability } 9/10.
\end{cases}
\]
w and \( v \) are zero-mean and
\[
\text{var}(v) = 16/9 > 1 = \text{var}(w).
\]
However,
\[
\mathbf{E} f(v) = 2/5 < 1/2 = \mathbf{E} f(w).
\]

(b) \( f(x_0 + tv) \) is convex in \( t \) for fixed \( v \), hence if \( v \) is a random variable, \( g(t) = \mathbf{E} f(x_0 + tv) \) is a convex function of \( t \). From Jensen’s inequality,
\[
g(t) = \mathbf{E} f(x_0 + tv) \geq f(x_0) = g(0).
\]

Now consider two points \( a, b \), with \( 0 < a < b \). If \( g(b) < g(a) \), then
\[
\frac{b-a}{b} g(b) + \frac{a}{b} g(a) < \frac{b-a}{b} g(a) + \frac{a}{b} g(a) = g(a)
\]
which contradicts Jensen’s inequality. Therefore we must have \( g(b) \geq g(a) \).

3.11 Monotone mappings. A function \( \psi: \mathbb{R}^n \rightarrow \mathbb{R}^n \) is called monotone if for all \( x, y \in \text{dom} \psi \),
\[
(\psi(x) - \psi(y))^T (x - y) \geq 0.
\]
(Note that ‘monotone’ as defined here is not the same as the definition given in §3.6.1. Both definitions are widely used.) Suppose \( f: \mathbb{R}^n \rightarrow \mathbb{R} \) is a differentiable convex function. Show that its gradient \( \nabla f \) is monotone. Is the converse true, i.e., is every monotone mapping the gradient of a convex function?)
Exercises

3.15 A family of concave utility functions. For \( 0 < \alpha \leq 1 \) let
\[
u_\alpha(x) = \frac{x^\alpha - 1}{\alpha},
\]
with \( \text{dom} \ u_\alpha = \mathbb{R}_+ \). We also define \( u_0(x) = \log x \) (with \( \text{dom} \ u_0 = \mathbb{R}_{++} \)).

(a) Show that for \( x > 0 \), \( u_0(x) = \lim_{\alpha \to 0} u_\alpha(x) \).
(b) Show that \( u_\alpha \) are concave, monotone increasing, and all satisfy \( u_\alpha(1) = 0 \).

These functions are often used in economics to model the benefit or utility of some quantity of goods or money. Concavity of \( u_\alpha \) means that the marginal utility (i.e., the increase in utility obtained for a fixed increase in the goods) decreases as the amount of goods increases. In other words, concavity models the effect of satiation.

Solution.

(a) In this limit, both the numerator and denominator go to zero, so we use l'Hopital's rule:
\[
\lim_{\alpha \to 0} u_\alpha(x) = \lim_{\alpha \to 0} \left( \frac{d}{d\alpha}(x^\alpha - 1) \right) = \lim_{\alpha \to 0} \frac{x^\alpha \log x}{1} = \log x.
\]
(b) By inspection we have
\[
u_\alpha(1) = \frac{1^\alpha - 1}{\alpha} = 0.
\]
The derivative is given by
\[
u'_\alpha(x) = x^{\alpha - 1},
\]
which is positive for all \( x \) (since \( 0 < \alpha < 1 \)), so these functions are increasing. To show concavity, we examine the second derivative:
\[
u''_\alpha(x) = (\alpha - 1)x^{\alpha - 2}.
\]
Since this is negative for all \( x \), we conclude that \( u_\alpha \) is strictly concave.

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

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

Solution. Strictly convex, and therefore quasiconvex. Also quasiconcave but not concave.

(b) \( f(x_1, x_2) = x_1x_2 \) on \( \mathbb{R}_{++}^2 \).

Solution. The Hessian of \( f \) is
\[
\nabla^2 f(x) = \begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix},
\]
which is neither positive semidefinite nor negative semidefinite. Therefore, \( f \) is neither convex nor concave. It is quasiconcave, since its superlevel sets
\[
\{(x_1, x_2) \in \mathbb{R}_{++}^2 \mid x_1x_2 \geq \alpha \}
\]
are convex. It is not quasiconvex.

(c) \( f(x_1, x_2) = 1/(x_1x_2) \) on \( \mathbb{R}_{++}^2 \).

Solution. The Hessian of \( f \) is
\[
\nabla^2 f(x) = \frac{1}{x_1x_2} \begin{bmatrix}
2/(x_1^2) & 1/(x_1x_2) \\
1/(x_1x_2) & 2/x_2^2
\end{bmatrix} \succeq 0
\]
Therefore, \( f \) is convex and quasiconvex. It is not quasiconcave or concave.
3.17 Suppose $f(x_1, x_2) = x_1/x_2$ on $\mathbb{R}^2_{++}$.

**Solution.** The Hessian of $f$ is

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

which is not positive or negative semidefinite. Therefore, $f$ is not convex or concave. It is quasiconvex and quasiconcave (i.e., quasilinear), since the sublevel and superlevel sets are halfspaces.

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

**Solution.** $f$ is convex, as mentioned on page 72. (See also figure 3.3). This is easily verified by working out the Hessian:

$$\nabla^2 f(x) = \begin{bmatrix} p \frac{1}{x_1^p} & -q \frac{1}{x_2^q} \\ -q \frac{1}{x_2^q} & 2 \frac{1}{x_1 x_2^q} \end{bmatrix} = \begin{bmatrix} 1 & -2p \frac{1}{x_2} \\ -2p \frac{1}{x_1} & 0 \end{bmatrix} \geq 0.$$

Therefore, $f$ is convex and quasiconvex. It is not concave or quasiconcave (see the figure).

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

**Solution.** Concave and quasiconcave. The Hessian is

$$\nabla^2 f(x) = \begin{bmatrix} \alpha(\alpha - 1)x_1^p x_2^q & \alpha(1 - \alpha)x_1^{p-1} x_2^{q-1} \\ \alpha(1 - \alpha)x_1^{p-1} x_2^{q-1} & (1 - \alpha)(-\alpha)x_1^{p-2} x_2^{q-1} \end{bmatrix}$$

$$= \alpha(1 - \alpha)x_1^{p-1} x_2^{q-1} \begin{bmatrix} x_1 & x_2 \\ x_2 & x_1 \end{bmatrix} = -\alpha(1 - \alpha)x_1^{p-1} x_2^{q-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \begin{bmatrix} 1/x_1 \\ -1/x_2 \end{bmatrix}$$

$$\leq 0.$$

$f$ is not convex or quasiconvex.
3.23 Perspective of a function.

(a) Show that for \( p > 1 \),

\[
f(x, t) = \frac{|x_1|^p + \cdots + |x_n|^p}{t^{p-1}} = \frac{\|x\|^p_p}{t^{p-1}}
\]

is convex on \( \{(x, t) \mid t > 0\} \).

Solution. This is the perspective function of \( \|x\|^p_p = |x_1|^p + \cdots + |x_n|^p \).

(b) Show that

\[
f(x) = \frac{\|Ax + b\|^2_2}{c^Tx + d}
\]

is convex on \( \{x \mid c^Tx + d > 0\} \), where \( A \in \mathbb{R}^{m \times n} \), \( b \in \mathbb{R}^m \), \( c \in \mathbb{R}^n \), and \( d \in \mathbb{R} \).

Solution. This function is the composition of the function \( g(y, t) = y^T y/t \) with an affine transformation \( (y, t) = (Ax + b, c^T x + d) \). Therefore convexity of \( f \) follows from the fact that \( g \) is convex on \( \{(y, t) \mid t > 0\} \).

For convexity of \( g \) one can note that it is the perspective of \( x^T x \), or directly verify that the Hessian

\[
\nabla^2 g(y, t) = \begin{bmatrix}
\frac{1}{t} & -y/t^2 \\
-y/t^2 & y^T y/t^3
\end{bmatrix}
\]

is positive semidefinite, since

\[
\begin{bmatrix}
v \\
w
\end{bmatrix}^T \begin{bmatrix}
\frac{1}{t} & -y/t^2 \\
-y/t^2 & y^T y/t^3
\end{bmatrix} \begin{bmatrix}
v \\
w
\end{bmatrix} = \|tv - yw\|^2_2 / t^3 \geq 0
\]

for all \( v \) and \( w \).

3.24 Some functions on the probability simplex. Let \( x \) be a real-valued random variable which takes values in \( \{a_1, \ldots, a_n\} \) where \( a_1 < a_2 < \cdots < a_n \), with \( \text{prob}(x = a_i) = p_i \), \( i = 1, \ldots, n \). For each of the following functions of \( p \) (on the probability simplex \( \{p \in \mathbb{R}_+^n \mid 1^T p = 1\} \)), determine if the function is convex, concave, quasiconvex, or quasiconcave.

(a) \( E x \).

Solution. \( E x = p_1 a_1 + \cdots + p_n a_n \) is linear, hence convex, concave, quasiconvex, and quasiconcave

(b) \( \text{prob}(x \geq \alpha) \).

Solution. Let \( j = \min\{i \mid a_i \geq \alpha\} \). Then \( \text{prob}(x \geq \alpha) = \sum_{i=j}^n p_i \). This is a linear function of \( p \), hence convex, concave, quasiconvex, and quasiconcave.

(c) \( \text{prob}(\alpha \leq x \leq \beta) \).

Solution. Let \( j = \min\{i \mid a_i \geq \alpha\} \) and \( k = \max\{i \mid a_i \leq \beta\} \). Then \( \text{prob}(\alpha \leq x \leq \beta) = \sum_{i=j}^k p_i \). This is a linear function of \( p \), hence convex, concave, quasiconvex, and quasiconcave.
Exercises

(d) \( \sum_{i=1}^{n} p_i \log p_i \), the negative entropy of the distribution.

Solution. \( p \log p \) is a convex function on \( \mathbb{R}_+ \) (assuming \( 0 \log 0 = 0 \)), so \( \sum p_i \log p_i \) is convex (and hence quasiconvex).

The function is not concave or quasiconcave. Consider, for example, \( n = 2, p_1 = (1, 0) \) and \( p_2 = (0, 1) \). Both \( p_1 \) and \( p_2 \) have function value zero, but the convex combination \( (0.5, 0.5) \) has function value \( \log(1/2) < 0 \). This shows that the superlevel sets are not convex.

(e) \( \text{var } x = \mathbb{E}(x - \mathbb{E}x)^2 \).

Solution. We have

\[
\text{var } x = \mathbb{E}x^2 - (\mathbb{E}x)^2 = \sum_{i=1}^{n} p_i a_i^2 - (\sum_{i=1}^{n} p_i a_i)^2,
\]

so \( \text{var } x \) is a concave quadratic function of \( p \).

The function is not convex or quasiconvex. Consider the example with \( n = 2, a_1 = 0, a_2 = 1 \). Both \( (p_1, p_2) = (1/4, 3/4) \) and \( (p_1, p_2) = (3/4, 1/4) \) lie in the probability simplex and have \( \text{var } x = 3/16 \), but the convex combination \( (p_1, p_2) = (1/2, 1/2) \) has a variance \( \text{var } x = 1/4 > 3/16 \). This shows that the sublevel sets are not convex.

(f) \( \text{quartile}(x) = \inf \{ \beta | \text{prob}(x \leq \beta) \geq 0.25 \} \).

Solution. The sublevel and the superlevel sets of \( \text{quartile}(x) \) are convex (see problem 2.15), so it is quasiconvex and quasiconcave.

\( \text{quartile}(x) \) is not continuous (it takes values in a discrete set \( \{a_1, \ldots, a_n\} \), so it is not convex or concave. (A convex or a concave function is always continuous on the relative interior of its domain.)

(g) The cardinality of the smallest set \( A \subseteq \{a_1, \ldots, a_n\} \) with probability \( \geq 90\% \). (By cardinality we mean the number of elements in \( A \).)

Solution. \( f \) is integer-valued, so it can not be convex or concave. (A convex or a concave function is always continuous on the relative interior of its domain.) \( f \) is quasiconcave because its superlevel sets are convex. We have \( f(p) \geq \alpha \) if and only if

\[
\sum_{i=1}^{k} p_{[i]} < 0.9,
\]

where \( k = \max \{i = 1, \ldots, n | i < \alpha\} \) is the largest integer less than \( \alpha \), and \( p_{[i]} \) is the \( i \)th largest component of \( p \). We know that \( \sum_{i=1}^{k} p_{[i]} \) is a convex function of \( p \), so the inequality \( \sum_{i=1}^{k} p_{[i]} < 0.9 \) defines a convex set.

In general, \( f(p) \) is not quasiconvex. For example, we can take \( n = 2, a_1 = 0 \) and \( a_2 = 1 \), and \( p^1 = (0.1, 0.9) \) and \( p^2 = (0.9, 0.1) \). Then \( f(p^1) = f(p^2) = 1 \), but \( f((p^1 + p^2)/2) = f(0.5, 0.5) = 2 \).

(h) The minimum width interval that contains 90% of the probability, i.e.,

\[
\inf \{ \beta - \alpha | \text{prob}(\alpha \leq x \leq \beta) \geq 0.9 \}.
\]

Solution. The minimum width interval that contains 90% of the probability must be of the form \( [a_i, a_j] \) with \( 1 \leq i \leq j \leq n \), because

\[
\text{prob}(\alpha \leq x \leq \beta) = \sum_{k=1}^{j} p_k = \text{prob}(a_i \leq x \leq a_k)
\]

where \( i = \min \{k | a_k \geq \alpha\} \), and \( j = \max \{k | a_k \leq \beta\} \).
3.25 **Maximum probability distance between distributions.** Let \( p, q \in \mathbb{R}^n \) represent two probability distributions on \( \{1, \ldots, n\} \) (so \( p, q \succeq 0, 1^T p = 1^T q = 1 \)). We define the maximum probability distance \( d_{\text{mp}}(p, q) \) between \( p \) and \( q \) as the maximum difference in probability assigned by \( p \) and \( q \), over all events:

\[
d_{\text{mp}}(p, q) = \max\{ |\text{prob}(p, C) - \text{prob}(q, C)| \mid C \subseteq \{1, \ldots, n\} \}.
\]

Here \( \text{prob}(p, C) \) is the probability of \( C \), under the distribution \( p \), i.e., \( \text{prob}(p, C) = \sum_{i \in C} p_i \).

Find a simple expression for \( d_{\text{mp}} \), involving \( \|p - q\|_1 = \sum_{i=1}^n |p_i - q_i| \), and show that \( d_{\text{mp}} \) is a convex function on \( \mathbb{R}^n \times \mathbb{R}^n \). (Its domain is \( \{(p, q) \mid p, q \succeq 0, 1^T p = 1^T q = 1 \} \), but it has a natural extension to all of \( \mathbb{R}^n \times \mathbb{R}^n \).)

**Solution.** Noting that

\[
\text{prob}(p, C) - \text{prob}(q, C) = -(\text{prob}(p, \tilde{C}) - \text{prob}(q, \tilde{C})),
\]

where \( \tilde{C} = \{1, \ldots, n\} \setminus C \), we can just as well express \( d_{\text{mp}} \) as

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

This shows that \( d_{\text{mp}} \) is convex, since it is the maximum of \( 2^n \) linear functions of \( (p, q) \).

Let’s now identify the (or a) subset \( C \) that maximizes

\[
\text{prob}(p, C) - \text{prob}(q, C) = \sum_{i \in C} (p_i - q_i).
\]

The solution is

\[
C^* = \{ i \in \{1, \ldots, n\} \mid p_i > q_i \}.
\]

Let’s show this. The indices for which \( p_i = q_i \) clearly don’t matter, so we will ignore them, and assume without loss of generality that for each index, \( p_i > q_i \) or \( p_i < q_i \). Now consider any other subset \( C \). If there is an element \( k \) in \( C^* \) but not \( C \), then by adding \( k \) to \( C \) we increase \( \text{prob}(p, C) - \text{prob}(q, C) \) by \( p_k - q_k > 0 \), so \( C \) could not have been optimal. Conversely, suppose that \( k \in C \setminus C^* \), so \( p_k - q_k < 0 \). If we remove \( k \) from \( C \), we’d increase \( \text{prob}(p, C) - \text{prob}(q, C) \) by \( q_k - p_k > 0 \), so \( C \) could not have been optimal. Thus, we have \( d_{\text{mp}}(p, q) = \sum_{p_i > q_i} (p_i - q_i) \). Now let’s express this in terms of \( \|p - q\|_1 \).

Using

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

we have

\[
\sum_{i=1}^n p_i < 0.9
\]

for all \( i, j \) that satisfy \( a_j - a_i < \gamma \). This defines a convex set.

The function is not convex, concave nor quasiconvex in general. Consider the example with \( n = 3 \), \( a_1 = 0 \), \( a_2 = 0.5 \) and \( a_3 = 1 \). On the line \( p_1 + p_3 = 0.95 \), we have

\[
f(p) = \begin{cases} 
0 & p_1 + p_3 = 0.95, \quad p_1 \in [0.05, 0.1] \cup [0.9, 0.95] \\
0.5 & p_1 + p_3 = 0.95, \quad p_1 \in (0.1, 0.15) \cup [0.85, 0.9] \\
1 & p_1 + p_3 = 0.95, \quad p_1 \in (0.15, 0.85)
\end{cases}
\]

It is clear that \( f \) is not convex, concave nor quasiconvex on the line.

We show that the function is quasiconcave. We have \( f(p) \geq \gamma \) if and only if all intervals of width less than \( \gamma \) have a probability less than 90%,

\[
\sum_{k=1}^j p_k < 0.9
\]
(a) Let $A = \text{conv } B$. Since $B \subseteq A$, we obviously have $S_B(y) \leq S_A(y)$. Suppose we have strict inequality for some $y$, i.e.,

$$y^Tu < y^Tv$$

for all $u \in B$ and some $v \in A$. This leads to a contradiction, because by definition $v$ is the convex combination of a set of points $u_i \in B$, i.e., $v = \sum_i \theta_i u_i$, with $\theta_i \geq 0$, $\sum_i \theta_i = 1$. Since

$$y^Tu_i < y^Tv$$

for all $i$, this would imply

$$y^Tv = \sum_i \theta_i y^Tu_i < \sum_i \theta_i y^Tv = y^Tv.$$  

We conclude that we must have equality $S_B(y) = S_A(y)$.

(b) Follows from

$$S_{A+B}(y) = \sup \{ y^T(u+v) \mid u \in A, v \in B \} = \sup \{ y^Tu \mid u \in A \} + \sup \{ y^Tv \mid u \in B \} = S_A(y) + S_B(y).$$

(c) Follows from

$$S_{A \cap B}(y) = \sup \{ y^Tu \mid u \in A \cap B \} = \max \{ \sup \{ y^Tu \mid u \in A \}, \sup \{ y^Tv \mid u \in B \} \} = \max \{ S_A(y), S_B(y) \}.$$ 

(d) Obviously, if $A \subseteq B$, then $S_A(y) \leq S_B(y)$ for all $y$. We need to show that if $A \not\subseteq B$, then $S_A(y) > S_B(y)$ for some $y$.

Suppose $A \not\subseteq B$. Consider a point $\bar{x} \in A, \bar{x} \not\in B$. Since $B$ is closed and convex, $\bar{x}$ can be strictly separated from $B$ by a hyperplane, i.e., there is a $y \neq 0$ such that

$$y^T\bar{x} > y^T \bar{x}$$

for all $x \in B$. It follows that $S_B(y) < y^T \bar{x} \leq S_A(y)$.

Conjugate functions

3.36 Derive the conjugates of the following functions.

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

Solution. We will show that

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

We first verify the domain of $f^*$. First 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

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

so $y$ is not in $\text{dom } f^*$. Next, assume $y \geq 0$ but $1^Ty > 1$. We choose $x = t1$ and let $t$ go to infinity, to show that

$$x^Ty = \max_i x_i = t1^Ty - t$$
is unbounded above. Similarly, when $y \geq 0$ and $1^T y < 1$, we choose $x = -t 1$ and let $t$ go to infinity.

The remaining case for $y$ is $y \geq 0$ and $1^T y = 1$. In this case 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 $\mathbb{R}^n$.

Solution. The conjugate is

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

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 $\text{dom } f^*$.

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 $\text{dom } f^*$.

Finally, assume that $1^T x \neq r$. We choose $x = t 1$ and find that

$$x^T y - f(x) = t 1^T y - tr$$

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 $\mathbb{R}$. $f(x) = \max_{i=1, \ldots, m} (a_i x + b_i)$ 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 $f(x) = a_k x + b_k$.

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 $\text{dom } f^* = [a_1, a_m]$, 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 $\mathbb{R}^{++}$, 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 $\mathbb{R}^+$. For $y > 0$ the function $y^x - x^y$ 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 $\text{dom } f^* = \mathbb{R}^{++}$ and $f^*(y) = \frac{-p}{q}(-y/p)^q$.

(e) **Geometric mean.** $f(x) = -(\prod x_i)^{1/n}$ on $\mathbb{R}^{++}$. (Recall that $\prod x_i$ is defined only if $x_i 
eq 0$ for all $i$.)

**Solution.** The conjugate function is

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

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

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

is unbounded above as a function of $t > 0$. Hence the condition $y \leq 0$ is indeed required.

Next assume that $y \leq 0$, but $(\prod_i(-y_i))^{1/n} < 1/n$. We choose $x_i = -t/y_i$, and obtain

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

as $t \to \infty$. This demonstrates that the second condition for the domain of $f^*$ is also needed.

Now assume that $y \geq 0$ and $(\prod_i(-y_i))^{1/n} \geq 1/n$, and $x \geq 0$. The arithmetic-geometric mean inequality states that

$$\frac{x^Ty}{n} \geq \left( \prod_i (-y_ix_i) \right)^{1/n} \geq \frac{1}{n} \left( \prod_i x_i \right)^{1/n},$$

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

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

**Solution.**

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

3.52 [MO79, §3.E.2] Log-convexity of moment functions. Suppose \( f : \mathbb{R} \to \mathbb{R} \) is nonnegative with \( \mathbb{R}_+ \subseteq \text{dom} \ f \). For \( x \geq 0 \) define
\[
\phi(x) = \int_0^\infty u^x f(u) \, du.
\]
Show that \( \phi \) is a log-convex function. (If \( x \) is a positive integer, and \( f \) is a probability density function, then \( \phi(x) \) is the \( x \)th moment of the distribution.)
Use this to show that the Gamma function,
\[
\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} \, du,
\]
is log-convex for \( x \geq 1 \).

**Solution.** \( g(x,u) = u^x f(u) \) is log-convex (as well as log-concave) in \( x \) for all \( u > 0 \). It follows directly from the property on page 106 that
\[
\phi(x) = \int_0^\infty g(x,u) \, du = \int_0^\infty u^x f(u) \, du
\]
is log-convex.

3.53 Suppose \( x \) and \( y \) are independent random vectors in \( \mathbb{R}^n \), with log-concave probability density functions \( f \) and \( g \), respectively. Show that the probability density function of the sum \( z = x + y \) is log-concave.

**Solution.** The probability density function of \( x + y \) is \( f \ast g \).

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

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

**Solution.** 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) Take exponentials and integrate. 
(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 
\[ \int_{-\infty}^{x} e^{-zt} dt = \frac{e^{-x^2}}{-x}. \]

3.55 Log-concavity of the cumulative distribution function of a log-concave probability density.
In this problem we extend the result of exercise 3.54. Let \( g(t) = \exp(-h(t)) \) be a differentiable log-concave probability density function, and let 
\[ f(x) = \int_{-\infty}^{x} g(t) dt = \int_{-\infty}^{x} e^{-h(t)} dt \]
be its cumulative distribution. We will show that \( f \) is log-concave, i.e., it satisfies \( f''(x)f(x) \leq (f'(x))^2 \) for all \( x \).

(a) Express the derivatives of \( f \) in terms of the function \( h \). Verify that \( f''(x)f(x) \leq (f'(x))^2 \) if \( h'(x) \geq 0 \).
(b) Assume that \( h'(x) < 0 \). Use the inequality 
\[ h(t) \geq h(x) + h'(x)(t-x) \]
(which follows from convexity of \( h \)), to show that 
\[ \int_{-\infty}^{x} e^{-h(t)} dt \leq \frac{e^{-h(x)}}{-h'(x)}. \]
Use this inequality to verify that \( f''(x)f(x) \leq (f'(x))^2 \) if \( h'(x) \geq 0 \).

Solution.
(a) \( f(x) = \int_{-\infty}^{x} e^{-h(t)} dt, f'(x) = e^{-h(x)}, f''(x) = -h'(x)e^{-h(x)}. \) Log-concavity means 
\[ -h'(x)e^{-h(x)} \int_{-\infty}^{x} e^{-h(t)} dt \leq e^{-2h(x)}, \]
which is obviously true if \( -h'(x) \leq 0 \).
(b) Take exponentials and integrate both sides of \( -h(t) \leq -h(x) - h'(x)(t-x) \): 
\[ \int_{-\infty}^{x} e^{-h(t)} dt \leq e^{xh'(x)-h(x)} \int_{-\infty}^{x} e^{-th'(x)} dt \]
\[ = e^{xh'(x)-h(x)} e^{-xh'(x)/(-h'(x))} \]
\[ = e^{-h(x)/-h'(x)} \]
\[ (h'(x)) \int_{-\infty}^{x} e^{-h(t)} dt \leq e^{-h(x)}. \]