Chapter 2

Convex functions

2.1 Basic properties and examples

2.1.1 Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *convex* if $\mathbf{dom}\, f$ is a convex set and if for all x, $y \in \mathbf{dom}\, f$, and θ with $0 \le \theta \le 1$, we have

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y). \qquad (2.1)$$

Geometrically, this inequality means that the line segment between $(x, f(x))$ and $(y, f(y))$ (*i.e.*, the *chord* from x to y) lies above the graph of f (figure 2.1). A function f is *strictly convex* if strict inequality holds in (2.1) whenever $x \ne y$ and $0 < \theta < 1$. We say f is *concave* if $-f$ is convex, and *strictly concave* if $-f$ is strictly convex.

For an affine function we always have equality in (2.1), so all affine (and therefore also linear) functions are both convex and concave. Conversely, any function that is convex and concave is affine.

A function is convex if and only if it is convex when restricted to any line that intersects its domain. In other words f is convex if and only if for all $x \in \mathbf{dom}\, f$ and all v, the function

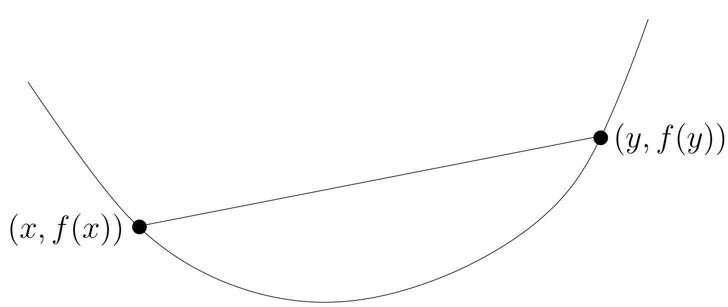


Figure 2.1: Graph of a convex function. The chord (*i.e.*, line segment) between any two points on the graph lies above the graph.

$h(t) = f(x + tv)$ is convex (on its domain, $\{t \mid x + tv \in \mathbf{dom}\, f\}$). This property is very useful, since it allows us to check whether a function is convex by restricting it to a line.

The *analysis* of convex functions is a well developed field, which we will not pursue in any depth. One simple result, for example, is that a convex function is continuous on the relative interior of its domain; it can have discontinuities only on its relative boundary.

2.1.2 Extended-valued extensions

It is often convenient to extend a convex function to all of $\mathbf{R}^n$ by defining its value to be ∞ outside its domain. If f is convex we define its extended-valued extension $\tilde{f} : \mathbf{R}^n \to \mathbf{R} \cup \{\infty\}$ by

$$\tilde{f}(x) = \begin{cases} f(x) & x \in \mathbf{dom}\, f, \\ +\infty & x \notin \mathbf{dom}\, f. \end{cases}$$

The extension $\tilde{f}$ is defined on all $\mathbf{R}^n$, and takes values in $\mathbf{R} \cup \{\infty\}$. We can recover the domain of the original function f from the extension $\tilde{f}$ as $\mathbf{dom}\, f = \{x \mid \tilde{f}(x) < \infty\}$.

The extension can simplify notation, since we do not have to explicitly describe the domain, or add the qualifier 'for all $x \in \mathbf{dom}\, f$' every time we refer to $f(x)$. Consider, for example, the basic defining inequality (2.1). In terms of the extension $\tilde{f}$, we can express it as: for $0 < \theta < 1$,

$$\tilde{f}(\theta x + (1-\theta)y) \leq \theta \tilde{f}(x) + (1-\theta)\tilde{f}(y)$$

for *any* x and y. Of course here we must interpret the inequality using extended arithmetic and ordering. For x and y both in $\mathbf{dom}\, f$, this inequality coincides with (2.1); if either is outside $\mathbf{dom}\, f$, then the righthand side is ∞, and the inequality therefore holds. As another example of this notational device, suppose f_1 and f_2 are two convex functions on $\mathbf{R}^n$. The pointwise sum $f = f_1 + f_2$ is the function with domain $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \mathbf{dom}\, f_2$, with $f(x) = f_1(x) + f_2(x)$ for any $x \in \mathbf{dom}\, f$. Using extended valued extensions we can simply say that for any x, $\tilde{f}(x) = \tilde{f}_1(x) + \tilde{f}_2(x)$. In this equation the domain of f has been automatically defined as $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \mathbf{dom}\, f_2$, since $\tilde{f}(x) = \infty$ whenever $x \notin \mathbf{dom}\, f_1$ or $x \notin \mathbf{dom}\, f_2$. In this example we are relying on extended arithmetic to automatically define the domain.

In this book we will use the same symbol to denote a convex function and its extension, whenever there is no harm from the ambiguity. This is the same as assuming that all convex functions are implicitly extended, *i.e.*, are defined as ∞ outside their domains.

Example 2.1 *Indicator function of a convex set.* Let $C \subseteq \mathbf{R}^n$ be a convex set, and consider the (convex) function I_C with domain C and $I_C(x) = 0$ for all $x \in C$. In other words, the function is identically zero on the set C. Its extended valued extension is given by

$$\tilde{I}_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C \end{cases}$$

The convex function $\tilde{I}_C$ is called the *indicator function* of the set C.

We can play several notational tricks with the indicator function $\tilde{I}_C$. For example the problem of minimizing a function g (defined on all of $\mathbf{R}^n$, say) on the set C is the same as minimizing the function $g + \tilde{I}_C$ over all of $\mathbf{R}^n$. Indeed, the function $g + \tilde{I}_C$ is (by our convention) g restricted to the set C.

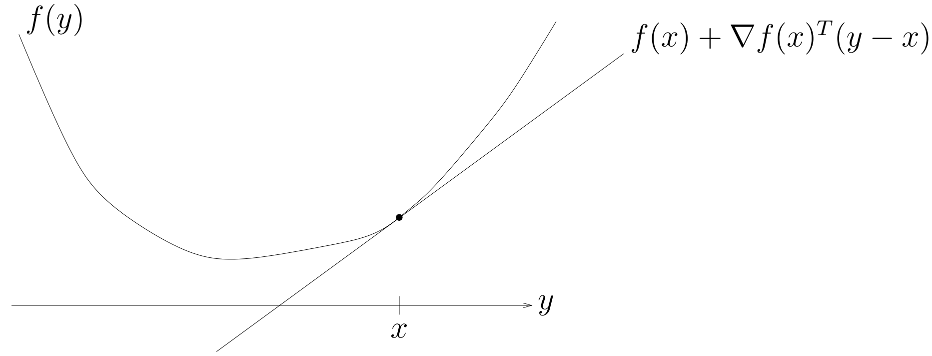


Figure 2.2: If f is convex and differentiable, then $f(x) + \nabla f(x)^T(y - x) \leq f(y)$ for all $x,\ y \in \mathbf{dom}\, f$.

In a similar way we can extend a concave function by defining it to be $-\infty$ outside its domain.

2.1.3 First order conditions

Suppose f is differentiable (*i.e.*, its gradient ∇f exists at each point in $\mathbf{dom}\, f$, which is open). Then f is convex if and only if $\mathbf{dom}\, f$ is convex and

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) \tag{2.2}$$

holds for all $x,\ y \in \mathbf{dom}\, f$. This inequality is illustrated in figure 2.2.

The affine function of y given by $f(x) + \nabla f(x)^T(y - x)$ is, of course, the first order Taylor approximation of f near x. The inequality (2.2) states that for a convex function, the first order Taylor approximation is in fact *global underestimator* of the function. Conversely, if the first order Taylor approximation of a function is always a global underestimator of the function, then the function is convex.

The inequality (2.2) shows that from *local information* about a convex function (*i.e.*, its derivative at a point) we can derive *global information* (*i.e.*, a global underestimator of it). This is perhaps the most important property of convex functions, and explains some of the remarkable properties of convex functions and convex optimization problems.

Proof. To prove (2.2), we first consider the case $n = 1$, *i.e.*, show that a differentiable function $f : \mathbf{R} \to \mathbf{R}$ is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x) \tag{2.3}$$

for all x and y.

Assume first that f is convex. Then for all $0 < t \leq 1$ we have

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

If we divide both sides by t, we obtain

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

and the limit as $t \to 0$ yields (2.3).

To show sufficiency, assume the function satisfies (2.3) for all x and y. Choose any $x \neq y$, and $0 \leq \theta \leq 1$, and let $z = \theta x + (1 - \theta)y$. Applying (2.3) twice yields

$$f(x) \geq f(z) + f'(z)(x - z), \qquad f(y) \geq f(z) + f'(z)(y - z).$$

Multiplying the first inequality by θ, the second by $1 - \theta$, and adding them yields

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

which proves that f is convex.

Now we can prove the general case, with $f : \mathbf{R}^n \to \mathbf{R}$. Let $x,\ y \in \mathbf{R}^n$ and consider f restricted to the line passing through them, *i.e.*, the function defined by $g(t) = f(ty + (1 - t)x)$, so $g'(t) = \nabla f(ty + (1 - t)x)^T(y - x)$.

First assume f is convex, which implies g is convex, so by the argument above we have $g(1) \geq g(0) + g'(0)$, which means

$$f(y) \geq f(x) + \nabla f(x)^T(y - x).$$

Now assume that this inequality holds for any x and y, so if $ty + (1 - t)x \in \mathbf{dom}\, f$ and $\tilde{t}y + (1 - \tilde{t})x \in \mathbf{dom}\, f$, we have

$$f(ty + (1 - t)x) \geq f(\tilde{t}y - (1 - \tilde{t})x) + \nabla f(\tilde{t}y - (1 - \tilde{t})x)^T(y - x)(t - \tilde{t}),$$

i.e., $g(t) \geq g(\tilde{t}) + g'(\tilde{t})(t - \tilde{t})$. We have seen that this implies that g is convex.

For concave functions we have the corresponding characterization: f is concave if and only if $\mathbf{dom}\, f$ is convex and

$$f(y) \leq f(x) + \nabla f(x)^T(y - x)$$

for all $x,\ y \in \mathbf{dom}\, f$.

Strict convexity can also be characterized by a first-order condition: f is strictly convex if and only if $\mathbf{dom}\, f$ is convex and for $x,\ y \in \mathbf{dom}\, f$, $x \neq y$, we have

$$f(y) > f(x) + \nabla f(x)^T(y - x). \tag{2.4}$$

2.1.4 Second order conditions

We now assume that f is twice differentiable, *i.e.*, its Hessian or second derivative $\nabla^2 f$ exists at each point in $\mathbf{dom}\, f$, which is open. Then f is convex if and only if $\mathbf{dom}\, f$ is convex and its Hessian is positive semidefinite:

$$\nabla^2 f(x) \succeq 0$$

for all $x \in \mathbf{dom}\, f$. For a function on $\mathbf{R}$, this reduces to the simple condition $f''(x) \geq 0$ (and $\mathbf{dom}\, f$ convex, *i.e.*, an interval). (The general second order condition is readily proved by reducing it to the case of $f : \mathbf{R} \to \mathbf{R}$). Similarly, f is concave if and only $\mathbf{dom}\, f$ is convex and $\nabla^2 f(x) \preceq 0$ for all $x \in \mathbf{dom}\, f$.

Strict convexity can be partially characterized by second order conditions. If $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$, then f is strictly convex. The converse, however, is not true: the function $f : \mathbf{R} \to \mathbf{R}$ given by $f(x) = x^4$ is strictly convex but has zero second derivative at $x = 0$.

Example 2.2 *Quadratic functions.* Consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, given by

$$f(x) = x^T P x + 2q^T x + r,$$

with $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Since $\nabla^2 f(x) = 2P$ for all x, f is convex if and only if $P \succeq 0$ (and concave if and only if $P \preceq 0$).

For quadratic functions, strict convexity is easily characterized: f is strictly convex if and only if $P \succ 0$ (and strictly concave if and only if $P \prec 0$).

Remark 2.1 The separate requirement that $\mathbf{dom}\, f$ be convex cannot be dropped from the first or second order characterizations of convexity and concavity. For example, the function $f(x) = 1/x^2$, with $\mathbf{dom}\, f = \{x \in \mathbf{R} \mid x \neq 0\}$, satisfies $f''(x) > 0$ for all $x \in \mathbf{dom}\, f$, but is not a convex function.

2.1.5 Examples

We have already mentioned that all linear and affine functions are convex (and concave), and have described the convex and concave quadratic functions. In this section we give a few more examples of convex and concave functions. We start with some functions on $\mathbf{R}$, with variable x.

- *Exponential.* e^{ax} is convex on $\mathbf{R}$.
- *Powers.* x^a is convex on $\mathbf{R}_{++}$ when $a \geq 1$ or $a \leq 0$, and concave for $0 \leq a \leq 1$.
- *Powers of absolute value.* $|x|^p$, for $p \geq 1$, is convex on $\mathbf{R}$.
- *Logarithm.* $\log x$ is concave on $\mathbf{R}_{++}$.
- *Negative entropy.* $x \log x$ (either on $\mathbf{R}_{++}$, or defined as 0 for $x = 0$) is convex.

Convexity or concavity of these examples can be shown by verifying the basic inequality (2.1), or by checking that the second derivative is nonnegative or nonpositive. For example, with $f(x) = x \log x$ we have

$$f'(x) = \log x + 1, \qquad f''(x) = 1/x,$$

so that $f''(x) > 0$ for $x > 0$. This shows that the negative entropy function is (strictly) convex.

We now give a few interesting examples of functions on $\mathbf{R}^n$.

- *Norms.* Every norm on $\mathbf{R}^n$ is convex.
- *Max function.* $f(x) = \max\{x_1, \ldots, x_n\}$ is convex on $\mathbf{R}^n$.
- *Quadratic-over-linear function.* The function $f(x, y) = x^2/y$, with $\mathbf{dom}\, f = \{(x, y) \in \mathbf{R}^2 | y > 0\}$, is convex.
- *Log-sum-exp.* The function $f(x) = \log(e^{x_1} + \cdots + e^{x_n})$ is convex on $\mathbf{R}^n$. This function can be interpreted as a smooth approximation of the max function, since

  $$\max\{x_1, \ldots, x_n\} \leq f(x) \leq \max\{x_1, \ldots, x_n\} + \log n$$

  for all x. (The second inequality is sharp when all components of x are equal.)
- *Geometric mean.* The geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{R}^n_{++}$.
- *Log determinant.* The function $f(X) = \log\det X^{-1}$ is convex on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

Convexity (or concavity) of these examples can be verified several ways, such as directly verifying the inequality (2.1), verifying that the Hessian is positive semidefinite, or restricting the function to an arbitrary line and verifying convexity of the resulting function of one variable.

Norms

If $f : \mathbf{R}^n \to \mathbf{R}$ is a norm, and $0 \leq \theta \leq 1$, then

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

since by definition a norm is homogeneous and satisfies the triangle inequality.

Max function

The function $f(x) = \max_i x_i$ satisfies, for $0 \leq \theta \leq 1$,

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

Quadratic-over-linear function

To show that the quadratic-over-linear function $f(x, y) = x^2/y$ is convex, we note that (for $y > 0$),

$$\nabla^2 f(x, y) = \frac{2}{y^3} \begin{bmatrix} y^2 & -xy \\ -xy & x^2 \end{bmatrix} = \frac{2}{y^3} \begin{bmatrix} y \\ -x \end{bmatrix} \begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0.$$

Log-sum-exp

The Hessian of the log-sum-exp function is

$$\nabla^2 f(x) = \frac{1}{(\mathbf{1}^T z)^2}\left((\mathbf{1}^T z)\,\mathbf{diag}(z) - zz^T\right)$$

where $z = (e^{x_1}, \ldots, e^{x_n})$. To verify that $\nabla^2 f(x) \succeq 0$ we must show that $v^T \nabla^2 f(x) v \geq 0$ for all v, *i.e.*,

$$v^T \nabla^2 f(x) v = \frac{1}{(\mathbf{1}^T z)^2}\left(\left(\sum_i z_i\right)\left(\sum_i v_i^2 z_i\right) - \left(\sum_i v_i z_i\right)^2\right) \geq 0.$$

But this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$ applied to the vectors with components $a_i = v_i\sqrt{z_i}$, $b_i = \sqrt{z_i}$.

Geometric mean

In a similar way we can show that the geometric mean $f(x) = \left(\prod_{i=1}^n x_i\right)^{1/n}$ is concave on $\mathbf{dom}\, f = \mathbf{R}^n_{++}$. Its Hessian $\nabla^2 f(x)$ is given by

$$\frac{\partial^2 f(x)}{\partial x_k^2} = -(n-1)\frac{\left(\prod_i x_i\right)^{1/n}}{n^2 x_k^2}, \qquad \frac{\partial^2 f(x)}{\partial x_k \partial x_l} = \frac{\left(\prod_i x_i\right)^{1/n}}{n^2 x_k x_l} \quad (k \neq l),$$

and can be expressed as

$$\nabla^2 f(x) = -\frac{\prod_i x_i^{1/n}}{n^2}\left(n\,\mathbf{diag}(1/x_1^2, \ldots, 1/x_n^2) - qq^T\right)$$

where $q_i = 1/x_i$. We must show that $\nabla^2 f(x) \preceq 0$, *i.e.*, that

$$v^T \nabla^2 f(x) v = -\frac{\prod_i x_i^{1/n}}{n^2}\left(n\sum_i v_i^2/x_i^2 - \left(\sum_i v_i/x_i\right)^2\right) \leq 0$$

for all v. Again this follows from the Cauchy-Schwarz inequality $(a^T a)(b^T b) \geq (a^T b)^2$, applied to the vectors $a = \mathbf{1}$ and $b_i = v_i/x_i$.

Log-determinant

For the function $f(X) = \log\det X^{-1}$, we can verify convexity by considering an arbitrary line, given by $X = Z + tV$, where $Z, V \in \mathbf{S}^n$. We define $g(t) = f(Z + tV)$, and restrict g to the interval of values of t for which $Z + tV \succ 0$. Without loss of generality, we can assume that $t = 0$ is inside this interval, *i.e.*, $Z \succ 0$. We have

$$\begin{aligned} g(t) &= -\log\det(Z + tV) \\ &= -\log\det Z^{1/2}\left(I + tZ^{-1/2} V Z^{-1/2}\right) Z^{1/2} \\ &= -\sum_{i=1}^n \log(1 + t\lambda_i) - \log\det Z \end{aligned}$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $Z^{-1/2}VZ^{-1/2}$. Therefore we have

$$g'(t) = -\sum_{i=1}^{n} \frac{\lambda_i}{1+t\lambda_i}, \quad g''(t) = \sum_{i=1}^{n} \frac{\lambda_i^2}{(1+t\lambda_i)^2}.$$

Since $g''(t) \geq 0$, we conclude that f is convex.

2.1.6 Sublevel sets

The α-*sublevel set* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$C_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}.$$

Sublevel sets of a convex function are convex, for any value of α. The proof is immediate from the definition of convexity: if $x,\ y \in C_\alpha$, *i.e.*, $f(x) \leq \alpha$, and $f(y) \leq \alpha$, then $f(\theta x+(1-\theta)y) \leq \alpha$ for $0 \leq \theta \leq 1$, and hence $\theta x + (1-\theta)y \in C_\alpha$.

The converse is not true; a function can have all its sublevel sets convex, but not be a convex function. For example, $f(x) = -e^x$ is not convex on $\mathbf{R}$ (indeed, it is strictly concave) but all its sublevel sets are convex.

If f is concave, then its α-*superlevel set*, given by $\{x \mid f(x) \geq \alpha\}$, is convex.

Example 2.3 The geometric and arithmetic means of $x \in \mathbf{R}^n_{++}$ are, respectively,

$$G(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad A(x) = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

The geometric-arithmetic mean inequality states that $G(x) \leq A(x)$.

Suppose $0 \leq \alpha \leq 1$, and consider the set

$$\{x \in \mathbf{R}^n_+ \mid G(x) \geq \alpha A(x)\},$$

i.e., the set of vectors with geometric mean no more than a factor α smaller than arithmetic mean. This set is convex, since it is the 0-superlevel set of the function $G(x) - \alpha A(x)$, which is concave. In fact, the set is positively homogeneous, so it is a convex cone.

2.1.7 Epigraph

The graph of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$\{(x, f(x)) \mid x \in \mathbf{dom}\, f\},$$

which is a subset of $\mathbf{R}^{n+1}$. The *epigraph* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$\mathbf{epi}\, f = \{(x,t) \mid x \in \mathbf{dom}\, f,\ f(x) \leq t\ \}.$$

('Epi' means 'above' so epigraph means 'above the graph'.) The definition is illustrated in figure 2.3.

The link between convex sets and convex functions is via the epigraph: A function is convex if and only if its epigraph is a convex set. Similarly, a function is strictly convex if and only if its epigraph is a strictly convex set.

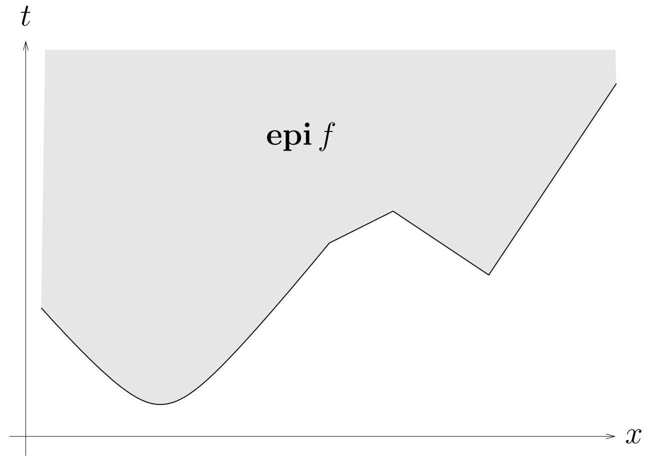


Figure 2.3: The epigraph of a function f. The boundary, shown darker, is the graph of f.

Example 2.4 *Matrix fractional function.* The function $f : \mathbf{R}^n \times \mathbf{S}^n \to \mathbf{R}$, defined as

$$f(x, Y) = x^T Y^{-1} x$$

is convex on $\mathbf{dom}\, f = \mathbf{R}^n \times \mathbf{S}^n_{++}$. (This generalizes the quadratic-over-linear function x^2/y, where $x,\ y \in \mathbf{R}$, $y > 0$.)

One easy way to establish convexity of f is via its epigraph:

$$\begin{aligned} \mathbf{epi}\, f &= \left\{(x, Y, t) \;\middle|\; Y \succ 0,\ x^T Y^{-1} x \leq t\right\} \\ &= \left\{(x, Y, t) \;\middle|\; \begin{bmatrix} Y & x \\ x^T & t \end{bmatrix} \succeq 0,\ Y \succ 0\right\} \end{aligned}$$

The last condition is a linear matrix inequality in (x, Y, t), and therefore is convex.

A function is concave if and only if its *hypograph*, defined as

$$\mathbf{hypo}\, f = \{(x, t) \mid t \leq f(x)\},$$

is a convex set.

Many results for convex functions can be proved (or interpreted) geometrically using epigraphs, and applying results for convex sets. For example, we can interpret the inequality (2.2) geometrically as follows:

$$\begin{aligned} (y, t) \in \mathbf{epi}\, f \quad &\Longrightarrow \quad t \geq f(x) + \nabla f(x)^T (y - x) \\ &\Longleftrightarrow \quad \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \begin{bmatrix} y - x \\ t - f(x) \end{bmatrix} \leq 0. \end{aligned}$$

This means that the hyperplane defined by $(\nabla f(x), -1)$ supports $\mathbf{epi}\, f$ at the boundary point $(x, f(x))$; see figure 2.4.

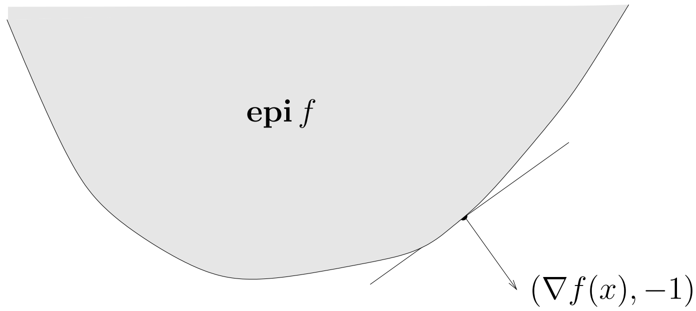


Figure 2.4: For a convex function f, the vector $(\nabla f(x), -1)$ defines a supporting hyperplane to the epigraph of f at x.

2.1.8 Jensen's inequality and extensions

The basic inequality (2.1), *i.e.*,

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

is sometimes called *Jensen's inequality*. It is easily extended to convex combinations of more than two points: If f is convex, $x_1, \ldots, x_k \in \mathbf{dom}\, f$ and $\theta_1, \ldots, \theta_k \ge 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

$$f(\theta_1 x_1 + \cdots + \theta_k x_k) \le \theta_1 f(x_1) + \cdots + \theta_k f(x_k).$$

As in the case of convex sets, the inequality extends to infinite sums, integrals, and expected values. For example, if $p(x) \ge 0$ on $S \subseteq \mathbf{dom}\, f$, $\int_S p(x)\, dx = 1$, then

$$f\left(\int_S p(x) x\, dx\right) \le \int_S f(x) p(x)\, dx$$

provided the integrals exist. In the most general case we can take any probability measure with support in $\mathbf{dom}\, f$. If x is a random variable such that $x \in \mathbf{dom}\, f$ with probability one, and f is convex, then we have

$$f(\mathbf{E}\, x) \le \mathbf{E}\, f(x), \tag{2.5}$$

provided the expectations exist. We can recover the basic inequality (2.1) from this general form, by taking the random variable x to have support $\{x_1, x_2\}$, with $\mathbf{Prob}(x = x_1) = \theta$, $\mathbf{Prob}(x = x_2) = 1 - \theta$. Thus the inequality (2.5) characterizes convexity: If f is not convex, there is a random variable x, with $x \in \mathbf{dom}\, f$ with probability one, such that $f(\mathbf{E}\, x) > \mathbf{E}\, f(x)$.

All of these inequalities are now called *Jensen's inequality*, even though the inequality studied by Jensen was the very simple one

$$f((x+y)/2) \le \frac{f(x) + f(y)}{2}.$$

Remark 2.2 We can interpret (2.5) as follows. Suppose $x \in \mathbf{dom}\, f \subseteq \mathbf{R}^n$ and z is any zero-mean random vector in $\mathbf{R}^n$. Then we have

$$\mathbf{E}\, f(x+z) \geq f(x).$$

Thus, randomization or dithering (*i.e.*, adding a zero mean random vector to the argument) cannot decrease the value of a convex function on average.

2.1.9 Inequalities

Many famous inequalities can be derived by applying Jensen's inequality to some appropriate convex function. (Indeed, convexity and Jensen's inequality can be made the foundation of a theory of inequalities.) As a simple example, consider the arithmetic-geometric mean inequality:

$$\sqrt{ab} \leq \frac{1}{2}(a+b) \tag{2.6}$$

for $a, b \geq 0$. The function $f(x) = -\log x$ is convex; Jensen's inequality with $\theta = 1/2$ yields

$$f((a+b)/2) = -\log(a+b)/2 \leq \frac{-\log a - \log b}{2} = \frac{f(a)+f(b)}{2}.$$

Taking the exponential of both sides yields (2.6).

As a less trivial example we prove Hölder's inequality: for $p > 1$, $1/p + 1/q = 1$,

$$\sum_{i=1}^n x_i y_i \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q}.$$

By convexity of $-\log x$, and Jensen's inequality with general θ, we obtain the more general arithmetic-geometric mean inequality

$$a^\theta b^{1-\theta} \leq \theta a + (1-\theta) b,$$

valid for $a, b \geq 0$ and $0 \leq \theta \leq 1$. Applying this with

$$a = \frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}, \qquad b = \frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}, \qquad \theta = 1/p,$$

yields

$$\left(\frac{|x_i|^p}{\sum_{j=1}^n |x_j|^p}\right)^{1/p} \left(\frac{|y_i|^q}{\sum_{j=1}^n |y_j|^q}\right)^{1/q} \leq \frac{|x_i|^p}{p\sum_{j=1}^n |x_j|^p} + \frac{|y_i|^q}{q\sum_{j=1}^n |y_j|^q}.$$

Summing over i then yields the Hölder inequality.

2.2 Operations that preserve convexity

In this section we describe some operations that preserve convexity or concavity of functions, or allow us to construct new convex and concave functions. We start with some simple operations such as addition, scaling, and pointwise supremum, and then describe some more sophisticated operations (some of which include the simple operations as special cases).

2.2.1 Nonnegative weighted sums

Evidently if f is a convex function and $\alpha \geq 0$, then the function αf is convex. If f_1 and f_2 are both convex functions, then so is their sum $f_1 + f_2$. Combining nonnegative scaling and addition, we see that the set of convex functions is itself a cone: a nonnegative weighted sum of convex functions,

$$f = w_1 f_1 + \cdots + w_m f_m,$$

is convex. Similarly, a nonnegative weighted sum of concave functions is concave. A nonnegative, nonzero weighted sum of strictly convex (concave) functions is strictly convex (concave).

These properties extend to infinite sums and integrals. For example if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, and $w(y) \geq 0$ for each $y \in \mathcal{A}$, then the function g defined as

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y)\, dy$$

is convex in x.

The fact that convexity is preserved under nonnegative scaling and addition is easily verified directly, or can be seen in terms of the associated epigraphs. For example, if $w \geq 0$ and f convex, we have

$$\mathbf{epi}(wf) = \begin{bmatrix} I & 0 \\ 0 & w \end{bmatrix} \mathbf{epi}\, f,$$

which is convex because the image of a convex set under a linear mapping is convex.

2.2.2 Composition with an affine mapping

Suppose $f : \mathbf{R}^n \to \mathbf{R}$, $A \in \mathbf{R}^{n \times m}$, and $b \in \mathbf{R}^n$. Define $g : \mathbf{R}^m \to \mathbf{R}$ by

$$g(x) = f(Ax + b)$$

with $\mathbf{dom}\, g = \{x \mid Ax + b \in \mathbf{dom}\, f\}$. Then if f is convex, so is g; if f is concave, so is g.

It is easy to directly establish the defining inequality (2.1). We can also establish the result using epigraphs. Suppose f is convex, so $\mathbf{epi}\, f$ is convex. The epigraph of g is

$$\mathbf{epi}\, g = \{(x, t) \mid (x, y) \in \mathcal{A},\ (y, t) \in \mathbf{epi}\, f \text{ for some } y\},$$

where $\mathcal{A} = \{(x, y) \mid y = Ax + b\}$. Hence $\mathbf{epi}\, g$ is the projection on the (x, t) subspace of the set

$$\{(x, t, y) \mid (x, y) \in \mathcal{A},\ (y, t) \in \mathbf{epi}\, f\},$$

which is the intersection of two convex sets. Hence, $\mathbf{epi}\, g$ is convex.

2.2.3 Pointwise maximum and supremum

If f_1 and f_2 are convex functions then their *pointwise maximum* f, defined by

$$f(x) = \max\{f_1(x), f_2(x)\},$$

with $\mathbf{dom}\, f = \mathbf{dom}\, f_1 \cap \mathbf{dom}\, f_2$, is also convex. This property is easily verified: if $0 \leq \theta \leq 1$ and $x,\ y \in \mathbf{dom}\, f$, then

$$\begin{array}{rcl} f(\theta x + (1-\theta)y) & = & \max\{f_1(\theta x + (1-\theta)y), f_2(\theta x + (1-\theta)y)\} \\ & \leq & \max\{\theta f_1(x) + (1-\theta)f_1(y), \theta f_2(x) + (1-\theta)f_2(y)\} \\ & \leq & \theta \max\{f_1(x), f_2(x)\} + (1-\theta)\max\{f_1(y), f_2(y)\} \\ & = & \theta f(x) + (1-\theta)f(y), \end{array}$$

which establishes convexity of f. It is easily shown that if $f_1, \ldots, f_m$ are convex, then their pointwise maximum

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

is also convex.

Example 2.5 *Piecewise linear functions.* The function

$$f(x) = \max_{i=1,\ldots,L} \{a_i^T x + b_i\}$$

defines a piecewise linear (or really, affine) function (with L or fewer regions). It is convex since it is the pointwise maximum of affine functions.

The converse can also be shown: any piecewise linear convex function with L or fewer regions can be expressed in this form. (See exercise 2.34.)

Example 2.6 *Sum of r largest components.* For $x \in \mathbf{R}^n$ we denote by $x_{[i]}$ the ith largest component of x, *i.e.*, $x_{[1]},\ x_{[2]}, \ldots, x_{[n]}$ are the components of x sorted in decreasing order. Then the function

$$f(x) = \sum_{i=1}^{r} x_{[i]},$$

i.e., the sum of the r largest elements of x, is a convex function. This can be seen by writing it 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 $n!/(r!(n-r)!)$ linear functions.

As an extension it can be shown that the function $\sum_{i=1}^{r} w_i x_{[i]}$ is convex, provided $w_1 \geq w_2 \geq \cdots \geq w_n \geq 0$. (See exercise 2.20.)

The pointwise maximum property extends to the pointwise supremum over an infinite set of convex functions. If for each $y \in \mathcal{A}$, $f(x,y)$ is convex in x, then the function g, defined as

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y) \tag{2.7}$$

is convex in x. Here the domain of g is (by our extended valued function convention)

$$\mathbf{dom}\, g = \left\{ x \;\middle|\; (x,y) \in \mathbf{dom}\, f \text{ for all } y \in \mathcal{A},\ \sup_{y \in \mathcal{A}} f(x,y) < \infty \right\}.$$

Similarly, the pointwise infimum of a set of concave functions is a concave function.

In terms of epigraphs, the pointwise supremum of functions corresponds to the intersection of epigraphs: with f, g, and $\mathcal{A}$ as defined in (2.7), we have

$$\mathbf{epi}\, g = \bigcap_{y \in \mathcal{A}} \mathbf{epi}\, f(\cdot, y).$$

Thus, the result follows from the fact that the intersection of a family of convex sets is convex.

Example 2.7 *Support function of a set.* Let $C \subseteq \mathbf{R}^n$, with $C \neq \emptyset$. The *support function S_C* associated with the set C is defined as

$$S_C(x) = \sup\{ \, x^T y \mid y \in C \, \}$$

(and, naturally, $\mathbf{dom}\, S_C = \{ \, x \mid \sup_{y \in C} x^T y < \infty \, \}$).

For each $y \in C$, $x^T y$ is a linear function of x, so S_C (which is the pointwise supremum of a family of linear functions) is convex.

Example 2.8 *Distance to farthest point of a set.* Let $C \subseteq \mathbf{R}^n$. The distance (in any norm) to the farthest point of C,

$$f(x) = \sup_{y \in C} \|x - y\|,$$

is convex. To see this, note that for any y, the function $\|x - y\|$ is convex in x. Since f is the pointwise supremum of a family of convex functions (indexed by $y \in C$), it is a convex function of x.

Example 2.9 *Least-squares cost as a function of weights.* Let $a_1, \ldots, a_n \in \mathbf{R}^m$. In a weighted least-squares problem we minimize the objective function $\sum_{i=1}^n w_i (a_i^T x - b_i)^2$ over $x \in \mathbf{R}^m$. We refer to w_i as *weights*, and allow negative w_i (which opens the possibility that the objective function is unbounded below).

We define the (optimal) *weighted least-squares cost* as

$$g(w) = \inf_x \sum_{i=1}^n w_i (a_i^T x - b_i)^2$$

(which is $-\infty$ if $\sum_i w_i (a_i^T x - b_i)^2$ is unbounded below).

Then g is a concave function of $w \in \mathbf{R}^n$, since it is the infimum of a family of linear functions of w (indexed by $x \in \mathbf{R}^m$).

In fact we can derive an explicit expression for g:

$$g(w) = \begin{cases} \sum_{i=1}^n w_i b_i^2 \left(1 - w_i a_i^T F^\dagger a_i\right), & F = \sum_{i=1}^n w_i a_i a_i^T \succeq 0, \\ -\infty & \text{otherwise,} \end{cases}$$

where $X^\dagger$ is the pseudo-inverse of a matrix X. Concavity of g from this expression is not obvious.

Example 2.10 *Maximum eigenvalue of a symmetric matrix.* The function $f(X) = \lambda_{\max}(X)$, with $\mathbf{dom}\, f = \mathbf{S}^m$, is convex. To see this, we express f as

$$f(X) = \sup \left\{ y^T X y \;\middle|\; \|y\|_2 = 1 \right\},$$

i.e., as the pointwise supremum of (an infinite number of) linear functions of X (*i.e.*, $y^T X y$) indexed by $y \in \mathbf{R}^m$.

Example 2.11 *Norm of a matrix.* Consider $f(X) = \|X\|_2$ with $\mathbf{dom}\, f = \mathbf{R}^{p\times q}$, where $\|\cdot\|_2$ denotes the spectral norm or maximum singular value. Convexity of f follows from

$$f(X) = \sup\{u^T X v \mid \|u\|_2 \leq 1,\ \|v\|_2 \leq 1\},$$

which shows it is the pointwise supremum of a family of affine functions of X.

As a generalization suppose $\|\cdot\|_a$ and $\|\cdot\|_b$ are norms on $\mathbf{R}^p$ and $\mathbf{R}^q$, respectively. Then the induced norm

$$\|X\|_{a,b} = \sup_{v\neq 0} \frac{\|Xv\|_a}{\|v\|_b}$$

is a convex function of $X \in \mathbf{R}^{p\times q}$, since it can be expressed as

$$f(X) = \sup\{u^T X v \mid \|u\|_{a*} \leq 1,\ \|v\|_b \leq 1\},$$

where $\|\cdot\|_{a*}$ is the dual norm of $\|\cdot\|_a$.

These examples illustrate a good method for establishing convexity of a function, *i.e.*, by expressing it as the pointwise supremum of a family of affine functions. Except for a technical condition, a sort of converse holds, *i.e.*, (almost) every convex function can be expressed as the pointwise supremum of a family of affine functions. As an example of a simple result, if $f : \mathbf{R}^n \to \mathbf{R}$ is convex, with $\mathbf{dom}\, f = \mathbf{R}^n$, then

$$f(x) = \sup\{\ g(x) \mid g \text{ affine},\ g(z) \leq f(z) \text{ for all } z\ \}.$$

In other words, f is the pointwise supremum of the set of all affine global underestimators of it. We give the proof of this result below, and leave the case where $\mathbf{dom}\, f \neq \mathbf{R}^n$ to the exercises.

Proof. Suppose f is convex with $\mathbf{dom}\, f = \mathbf{R}^n$. The inequality

$$f(x) \geq \sup\{\ g(x) \mid g \text{ affine},\ g(z) \leq f(z) \text{ for all } z\ \}$$

is clear, since if g is any affine underestimator of f, we have $g(x) \leq f(x)$. To establish equality, we will show that for each $x \in \mathbf{R}^n$, there is an affine function g, which is global underestimator of f, and satisfies $g(x) = f(x)$.

The epigraph of f is, of course, a convex set. Hence we can find a supporting hyperplane to it at $(x, f(x))$, *i.e.*, $a \in \mathbf{R}^n$ and $b \in \mathbf{R}$ with $(a, b) \neq 0$ and

$$\begin{bmatrix} a \\ b \end{bmatrix}^T \begin{bmatrix} x - z \\ f(x) - t \end{bmatrix} \leq 0$$

for all $(z, t) \in \mathbf{epi}\, f$. This means that

$$a^T(x - z) + b(f(x) - f(z) - s) \leq 0 \tag{2.8}$$

for all $z \in \mathbf{dom}\, f = \mathbf{R}^n$ and all $s \geq 0$ (since $(z, t) \in \mathbf{epi}\, f$ means $t = f(z) + s$ for some $s \geq 0$). For the inequality (2.8) to hold for all $s \geq 0$, we must have $b \geq 0$. If $b = 0$, then the inequality (2.8) reduces to $a^T(x - z) \leq 0$ for all $z \in \mathbf{R}^n$, which implies $a = 0$ and contradicts $(a, b) \neq 0$. We conclude that $b > 0$, *i.e.*, that the separating hyperplane is not vertical.

Using the fact that $b > 0$ we rewrite (2.8) as

$$g(z) = f(x) + (a/b)^T(x - z) \leq f(z)$$

for all z. The function g is an affine underestimator of f, and satisfies $g(x) = f(x)$.

2.2.4 Composition

In this section we examine conditions on $h : \mathbf{R}^k \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}^k$ that guarantee convexity or concavity of their composition $f = h \circ g : \mathbf{R}^n \to \mathbf{R}$, defined by

$$f(x) = h(g(x)), \qquad \mathbf{dom}\, f = \{\ x \in \mathbf{dom}\, g \mid g(x) \in \mathbf{dom}\, h\ \}.$$

Scalar composition

We first consider the case $k = 1$, so $h : \mathbf{R} \to \mathbf{R}$ and $g : \mathbf{R}^n \to \mathbf{R}$. We can restrict ourselves to the case $n = 1$ (since convexity is determined by the behavior of a function on arbitrary lines that intersect its domain).

To discover the composition rules, we start by assuming that h and g are twice differentiable, with $\mathbf{dom}\, g = \mathbf{dom}\, h = \mathbf{R}$. In this case, convexity of f reduces to $f'' \geq 0$ (meaning, $f''(x) \geq 0$ for all $x \in \mathbf{R}$).

The second derivative of the composition function $f = h \circ g$ is given by

$$f''(x) = (g'(x))^2 h''(g(x)) + g''(x)h'(g(x)). \tag{2.9}$$

Now suppose, for example, that g is convex (so $g'' \geq 0$) and h is convex and nondecreasing (so $h'' \geq 0$ and $h' \geq 0$). It follows from (2.9) that $f'' \geq 0$, *i.e.*, f is convex. In a similar way, the expression (2.9) gives the results

$$\begin{array}{l} f \text{ is convex if } g \text{ is convex and } h \text{ is convex and nondecreasing} \\ f \text{ is convex if } g \text{ is concave and } h \text{ is convex and nonincreasing} \\ f \text{ is concave if } g \text{ is concave and } h \text{ is concave and nondecreasing} \\ f \text{ is concave if } g \text{ is convex and } h \text{ is concave and nonincreasing} \end{array} \tag{2.10}$$

which are valid when the functions g and h are twice differentiable and have domains that are all of $\mathbf{R}$.

There are very similar composition rules in the general case $n > 1$, without assuming differentiability of h and g, or that $\mathbf{dom}\, g = \mathbf{R}^n$ and $\mathbf{dom}\, h = \mathbf{R}$:

$$\begin{array}{l} f \text{ is convex if } g \text{ is convex, } h \text{ is convex, and } \tilde{h} \text{ is nondecreasing} \\ f \text{ is convex if } g \text{ is concave, } h \text{ is convex, and } \tilde{h} \text{ is nonincreasing} \\ f \text{ is concave if } g \text{ is concave, } h \text{ is concave, and } \tilde{h} \text{ is nondecreasing} \\ f \text{ is concave if } g \text{ is convex, } h \text{ is concave, and } \tilde{h} \text{ is nonincreasing} \end{array} \tag{2.11}$$

where $\tilde{h}$ denotes that extended value extension of h. The only difference between these results, and the results (2.10), is that we require that the *extended value extension* function h be nonincreasing or nondecreasing.

To understand what this means, suppose h is convex, so $\tilde{h}$ takes on the value $+\infty$ outside $\mathbf{dom}\, h$. To say that $\tilde{h}$ is nondecreasing means that for *any* $x,\ y \in \mathbf{R}$, with $x < y$, we have $\tilde{h}(x) \leq \tilde{h}(y)$. In particular, this means that if $y \in \mathbf{dom}\, f$, then $x \in \mathbf{dom}\, f$. In other words, the domain of h extends infinitely in the negative direction; it is either $\mathbf{R}$, or interval of the form $(-\infty, a)$ or $(-\infty, a]$. In a similar way, to say that $\tilde{h}$ is nonincreasing means that h is nonincreasing and $\mathbf{dom}\, h$ extends infinitely in the positive direction. When the function h is concave, $\tilde{h}$ takes on the value $-\infty$ outside $\mathbf{dom}\, h$, and $\tilde{h}$ is nonincreasing means that h is nonincreasing and $\mathbf{dom}\, h$ extends infinitely in the negative direction.

Example 2.12 • The function $h(x) = \log x$, with $\mathbf{dom}\, h = \{x \in \mathbf{R} \mid x > 0\}$, is concave and satisfies $\tilde{h}$ nondecreasing.

- The function $h(x) = x^{1/2}$, with $\mathbf{dom}\, h = \mathbf{R}_+$, is concave and satisfies the condition $\tilde{h}$ nondecreasing.
- The function $h(x) = x^{3/2}$, with $\mathbf{dom}\, h = \mathbf{R}_+$, is convex but *does not* satisfy the condition $\tilde{h}$ nondecreasing. For example, we have $\tilde{h}(-1) = \infty$, but $\tilde{h}(1) = 1$.
- The function $h(x) = x^{3/2}$ for $x \geq 0$, and $h(x) = 0$ for $x < 0$, with $\mathbf{dom}\, h = \mathbf{R}$, is convex and *does* satisfy the condition $\tilde{h}$ nondecreasing.

The composition results (2.11) can be proved directly, without assuming differentiability, or using the formula (2.9). For example, suppose g is convex and $\tilde{h}$ is convex and nondecreasing. Assume that $x,\ y \in \mathbf{dom}\, f$, and $0 \leq \theta \leq 1$. Since $x,\ y \in \mathbf{dom}\, f$, we have that $x,\ y \in \mathbf{dom}\, g$ and $g(x),\ g(y) \in \mathbf{dom}\, h$. Since $\mathbf{dom}\, g$ is convex, we conclude that $\theta x + (1-\theta)y \in \mathbf{dom}\, g$, and from convexity of g, we have

$$g(\theta x + (1-\theta)y) \leq \theta g(x) + (1-\theta)g(y). \tag{2.12}$$

Since $g(x),\ g(y) \in \mathbf{dom}\, h$, we conclude that $\theta g(x) + (1-\theta)g(y) \in \mathbf{dom}\, h$, *i.e.*, the righthand side of (2.12) is in $\mathbf{dom}\, h$. Now we use the assumption that $\tilde{h}$ is nondecreasing, which means that its domain extends infinitely in the negative direction. Since the righthand side of (2.12) is in $\mathbf{dom}\, h$, we conclude that the lefthand side, *i.e.*, $g(\theta x + (1-\theta)y \in \mathbf{dom}\, h$. This means that $\theta x + (1-\theta)y \in \mathbf{dom}\, f$. At this point, we have shown that $\mathbf{dom}\, f$ is convex.

Now using the fact that $\tilde{h}$ is nondecreasing and the inequality (2.12), we get

$$h(g(\theta x + (1-\theta)y)) \leq h(\theta g(x) + (1-\theta)g(y)). \tag{2.13}$$

From convexity of h, we have

$$h(\theta g(x) + (1-\theta) g(y)) \leq \theta h(g(x)) + (1-\theta) h(g(y)). \tag{2.14}$$

Putting (2.13) and (2.14) together, we have

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

which proves the composition theorem.

Example 2.13

- If g is convex then $f(x) = \exp g(x)$ is convex.
- If g is concave and positive, then $\log g(x)$ is concave.
- If g is concave and positive, then $1/g(x)$ is convex.
- If g is convex and nonnegative and $p \geq 1$, then $g(x)^p$ is convex.
- If f is convex then $-\log(-f(x))$ is convex on $\{x \mid f(x) < 0\}$.

Remark 2.3 The requirement that the monotonicity hold for the extended valued extension $\tilde{h}$, and not just the function h, cannot be removed. For example, consider the function $g(x) = x^2$, with $\mathbf{dom}\, g = \mathbf{R}$, and $h(x) = 0$, with $\mathbf{dom}\, h = [1, 2]$. Here g is convex, and h is convex and nondecreasing. But the function $f = h \circ g$, given by

$$f(x) = 0, \quad \mathbf{dom}\, f = [-\sqrt{2}, 1] \;\cup\; [1, \sqrt{2}]$$

is not convex, since its domain is not convex. Here, of course, the function $\tilde{h}$ is *not* nondecreasing.

Vector composition

We now turn to the more complicated case when $k \geq 1$. Suppose

$$f(x) = h(g_1(x), \ldots, g_k(x))$$

with $h : \mathbf{R}^k \to \mathbf{R}$, $g_i : \mathbf{R}^n \to \mathbf{R}$. Again without loss of generality we can assume $n = 1$.

As in the case $k = 1$, we start by assuming the functions are twice differentiable, with $\mathbf{dom}\, g = \mathbf{R}$ and $\mathbf{dom}\, h = \mathbf{R}^k$, in order to discover the composition rules. We have

$$f''(x) = \nabla h(g(x))^T \begin{bmatrix} g_1(x)'' \\ \vdots \\ g_k(x)'' \end{bmatrix} + \begin{bmatrix} g_1(x)' \\ \vdots \\ g_k(x)' \end{bmatrix}^T \nabla^2 h \begin{bmatrix} g_1(x)' \\ \vdots \\ g_k(x)' \end{bmatrix}. \tag{2.15}$$

Again the issue is to determine conditions under which $f(x)'' \geq 0$ for all x (or $f(x)'' \leq 0$ for all x for concavity). From (2.15) we can derive many rules, for example:

- If h is convex and nondecreasing in each argument, and g_i are convex, then f is convex.
- If h is convex and nonincreasing in each argument, and g_i are concave, then f is convex.

- If h is concave and nondecreasing in each argument, and g_i are concave, then f is concave.

As in the scalar case, similar composition results hold in general, with $n > 1$, no assumption of differentiability of h or g, and general domains. For the general results, the monotonicity condition on h must hold for the extended valued extension $\tilde{h}$.

To understand the meaning of the condition that the extended valued extension $\tilde{h}$ be monotonic, we consider the case where $h : \mathbf{R}^k \to \mathbf{R}$ is convex, and $\tilde{h}$ nondecreasing, *i.e.*, whenever $u \preceq v$, we have $\tilde{h}(u) \leq \tilde{h}(v)$. This implies that if $v \in \mathbf{dom}\, h$, then so is u; *i.e.*, $\mathbf{dom}\, h - \mathbf{R}^k_+ = \mathbf{dom}\, h$.

Example 2.14 • Let $h(z) = z_{[1]} + \cdots + z_{[r]}$, *i.e.*, h is the sum of the r largest components of z. Then h is convex and nondecreasing in each argument. Suppose $g_1, \ldots, g_k$ are convex functions on $\mathbf{R}^n$. Then the composition function $f = h(g)$, *i.e.*, the pointwise sum of the r largest g_i's, is convex.

- $h(z) = \log \sum_{i=1,\ldots,k} \exp z_i$ is convex and nondecreasing in each argument, so $\log \sum_{i=1,\ldots,m} \exp g_i$ is convex whenever g_i are.
- For $0 < p \leq 1$, the function $h(z) = (\sum z_i^p)^{1/p}$ on $\mathbf{R}^n_+$ is concave, and its extension (which has the value $-\infty$ for $x \notin \mathbf{R}^n_+$) nondecreasing in each component. So if g_i are convex and positive, we conclude that $f(x) = (\sum g_i(x)^p)^{1/p}$ is concave.
- For $p \geq 1$, the function h defined as

$$h(z) = \begin{cases} (\sum z_i^p)^{1/p} & \text{for } x \in \mathbf{R}^n_+, \\ 0 & \text{for } x \notin \mathbf{R}^n_+, \end{cases}$$

  is convex and nondecreasing. Therefore $h(x) = (\sum g_i(x)^p)^{1/p}$ is convex if g_i are convex and positive.

  (Note that to apply the composition rule in this case, we needed to extend h as 0 outside $\mathbf{R}^n_+$.)

2.2.5 Minimization

We have seen that the maximum or supremum of an arbitrary family of convex functions is convex. It turns out that some special forms of minimization also yield convex functions. If f is convex in (x, y), and C is a convex nonempty set, then the function

$$g(x) = \inf_{y \in C} f(x, y) \tag{2.16}$$

is convex in x. Here, of course, we take

$$\mathbf{dom}\, g = \left\{ x \;\middle|\; \inf_{y \in C} f(x, y) > -\infty \right\}$$

(which can be shown to be convex).

We prove this by verifying Jensen's inequality for $x_1,\ x_2 \in \mathbf{dom}\, g$. Let $\epsilon > 0$. Then there are $y_1,\ y_2 \in C$ such that $f(x_i, y_i) \le g(x_i) + \epsilon$ for $i = 1,\ 2$. Now let $\theta \in [0,1]$. Then we have

$$\begin{aligned} g(\theta x_1 + (1-\theta)x_2)) &= \inf_{y \in C} f(\theta x_1 + (1-\theta)x_2, y) \\ &\le f(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta)y_2) \\ &\le \theta f(x_1, y_1) + (1-\theta) f(x_2, y_2) \\ &\le \theta g(x_1) + (1-\theta) g(x_2) + \epsilon. \end{aligned}$$

Since this holds for any $\epsilon > 0$, we have

$$g(\theta x_1 + (1-\theta)x_2) \le \theta g(x_1) + (1-\theta) g(x_2).$$

The result can also be seen in terms of epigraphs. With f, g, and C defined as in (2.16), we have

$$\mathbf{epi}\, g = \{(x,t) \mid \exists y,\ (x,y,t) \in \mathbf{epi}\, f,\ y \in C\}.$$

Thus $\mathbf{epi}\, g$ is convex, since it is the projection of a convex set on a subset of its components.

Example 2.15 *Schur complement.* Suppose the quadratic function

$$f(x,y) = x^T A x + 2x^T B y + y^T C y,$$

(where A and C are symmetric) is convex, *i.e.*,

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0.$$

We can express $g(x) = \inf_y f(x,y)$ as

$$g(x) = x^T (A - BC^\dagger B^T) x,$$

where $C^\dagger$ is the pseudo-inverse of C. By the minimization rule, g is convex, so we conclude that $A - BC^\dagger B^T \succeq 0$.

(If C is invertible, *i.e.*, $C \succ 0$, then the matrix $A - BC^{-1}B^T$ is called the *Schur complement* of C in the matrix $\begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$.)

Example 2.16 *Distance to a set.* The distance of a point x to a set $S \subseteq \mathbf{R}^n$ is defined as

$$\mathrm{dist}(x,S) = \inf_{y \in S} \|x - y\|.$$

The function $\|x - y\|$ is convex in (x,y), so if the set S is convex, the distance function $\mathrm{dist}(x,S)$ is a convex function of x.

Example 2.17 Suppose h is convex. Then the function g defined as

$$g(x) = \inf\{h(y) \mid Ay = x\}$$

is convex. To see this we define f by

$$f(x,y) = \begin{cases} h(y) & \text{if } Ay = x \\ +\infty & \text{otherwise,} \end{cases}$$

which is convex in (x,y). Then g is the minimum of f over y, and hence is convex. (It is not hard to show directly that g is convex.)

2.2.6 Perspective of a function

If $f : \mathbf{R}^n \to \mathbf{R}$, then the *perspective* of f is the function $g : \mathbf{R}^{n+1} \to \mathbf{R}$ defined by

$$g(x,t) = tf(x/t)$$

for $t > 0$. Its domain is, naturally,

$$\mathbf{dom}\, g = \{(x,t) \mid x/t \in \mathbf{dom}\, f,\ t > 0\}$$

(which is easily shown to be convex). If f is convex, then the perspective function g is convex.

This can be proved several ways, for example, direct verification of the defining inequality (2.1) (see exercises). We give a short proof here using epigraphs and the perspective mapping on $\mathbf{R}^{n+1}$ described in §1.3.3 (which will also explain the name 'perspective'). For $t > 0$ we have

$$\begin{aligned} (x,t,s) \in \mathbf{epi}\, g &\iff tf(x/t) \le s \\ &\iff f(x/t) \le s/t \\ &\iff (x/t, s/t) \in \mathbf{epi}\, f. \end{aligned}$$

Therefore $\mathbf{epi}\, g$ is the inverse image of $\mathbf{epi}\, f$ under the perspective mapping that takes (u,v,w) to $(u,v)/w$. It follows (see §1.3.3) that $\mathbf{epi}\, g$ is convex, so the function g is convex.

Example 2.18 The perspective of the convex function $f(x) = x^T x$ on $\mathbf{R}^n$ is

$$g(x,t) = t(x/t)^T(x/t) = \frac{x^T x}{t},$$

which is convex in (x,t) for $t > 0$.

Example 2.19 $f(x) = -\log x$ on $\mathbf{R}_{++}$. Then

$$g(x,t) = -t\log(x/t) = t\log t - t\log x$$

is convex (in (x,t)). For $x = 1$, this reduces to the negative entropy function.

From convexity of g we can establish convexity of several interesting related functions. It follows that the *Kullback-Leibler* distance (or divergence) between $x > 0$ and $t > 0$, given by

$$D_{\mathrm{kl}}(x,t) = x\log(x/t) - x + t,$$

is convex. (D_{kl} also satisfies $D_{\mathrm{kl}} \ge 0$ and $D_{\mathrm{kl}} = 0$ if and only if $x = t$; see exercises.)

It also follows that the function

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

is convex in $(x, t_1, \ldots, t_n)$ (for $x,\ t_i > 0$). Taking $x = \mathbf{1}^T t$, we conclude that the *normalized negative entropy* function,

$$\sum_{i=1}^n t_i \log t_i - (\mathbf{1}^T t)\log(\mathbf{1}^T t) = \sum_{i=1}^n \frac{t_i}{\mathbf{1}^T t} \log \frac{t_i}{\mathbf{1}^T t}$$

is convex. (Note that $t_i/\mathbf{1}^T t$ normalizes t to be a probability distribution; the normalized entropy is the entropy of this distribution.)

Example 2.20 Suppose $f : \mathbf{R}^m \to \mathbf{R}$ is convex, and $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$, and $d \in \mathbf{R}$. We define

$$g(x) = (c^T x + d) f\left((Ax+b)/(c^T x + d)\right)$$

with

$$\mathbf{dom}\, g = \{x \mid c^T x + d > 0,\ (Ax+b)/(c^T x + d) \in \mathbf{dom}\, f\}.$$

Then g is convex.

2.3 The conjugate function

In this section we introduce an operation that will play an important role in later chapters.

2.3.1 Definition and examples

Let $f : \mathbf{R}^n \to \mathbf{R}$. The function $f^\star : \mathbf{R}^n \to \mathbf{R}$, defined as

$$f^\star(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x)\right), \tag{2.17}$$

is called the *conjugate* of the function f. The domain of the conjugate function consists of $y \in \mathbf{R}^n$ for which the supremum is finite, *i.e.*, for which the difference $x^T y - f(x)$ is bounded above on $\mathbf{dom}\, f$.

We see immediately that $f^\star$ is a convex function, since it is the pointwise supremum of a family of convex (indeed, affine) functions of y. This is true whether or not f is convex. (Note that when f is convex, the subscript $x \in \mathbf{dom}\, f$ is not necessary, since by convention, $y^T x - f(x) = -\infty$ for $x \notin \mathbf{dom}\, f$.)

We start with some simple examples, and then describe some rules for conjugating functions. This allows us to derive an analytical expression for the conjugate of many common convex functions.

Example 2.21 We derive the conjugate of some convex functions on $\mathbf{R}$.

- *Affine function.* $f(x) = ax + b$. As a function of x, $yx - ax - b$ is bounded if and only if $y = a$, in which case it is constant. Therefore the domain of the conjugate function $f^\star$ is the singleton $\{a\}$, and $f^\star(a) = -b$.
- *Logarithm.* $f(x) = -\log x$, with $\mathbf{dom}\, f = \mathbf{R}_{++}$. The function $xy + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $\mathbf{dom}\, f^\star = \{y \mid y < 0\}$ and $f^\star(y) = -\log(-y) - 1$ for $y < 0$.
- *Exponential.* $f(x) = e^x$. $xy - e^x$ is unbounded if $y \leq 0$. For $y > 0$, $xy - e^x$ reaches its maximum at $x = \log y$, so we have $f^\star(y) = y \log y - y$. For $y = 0$, $f^\star(y) = \sup_x -e^x = 0$. In summary, $\mathbf{dom}\, f^\star = \mathbf{R}_+$ and $f^\star(y) = y\log y - y$ (with the interpretation $0 \log 0 = 0$).
- *Negative entropy.* $f(x) = x \log x$, with $\mathbf{dom}\, f = \mathbf{R}_+$ (and $f(0) = 0$). The function $xy - x\log x$ is bounded above on $\mathbf{R}_+$ for all y, hence $\mathbf{dom}\, f^\star = \mathbf{R}$. It attains its maximum at $x = e^{y-1}$, and substituting we find $f^\star(y) = e^{y-1}$.

- *Inverse.* $f(x) = 1/x$ on $\mathbf{R}_{++}$. For $y > 0$, $yx - 1/x$ is unbounded above. For $y = 0$ this function has supremum 0; for $y < 0$ the supremum is attained at $x = (-y)^{-1/2}$. Therefore we have $f^\star(y) = -2(-y)^{1/2}$, with $\mathbf{dom}\, f^\star = \mathbf{R}_-$.

Example 2.22 *Strictly convex quadratic function.* Consider $f(x) = \frac{1}{2}x^TQx$, with $Q \in \mathbf{S}^n_+$. The function $x^Ty - \frac{1}{2}x^TQx$ is bounded above as a function of x for all y. It attains its maximum at $x = Q^{-1}y$, so

$$f^\star(y) = \frac{1}{2}y^TQ^{-1}y.$$

Example 2.23 *Log determinant.* $f(X) = \log\det X^{-1}$ on $\mathbf{S}^n_{++}$. The conjugate function is defined as

$$f^\star(Y) = \sup_{X \succ 0} \mathbf{Tr}\, XY + \log\det X.$$

The argument of the supremum is unbounded above unless $Y \prec 0$; when $Y \prec 0$ we can find the maximizing X by setting the gradient with respect to X equal to zero:

$$\nabla_X \left(\mathbf{Tr}\, XY + \log\det X\right) = Y + X^{-1} = 0$$

(see §B.3.2), which yields $X = -Y^{-1}$. Therefore we have

$$f^\star(Y) = \log\det(-Y)^{-1} - n$$

on $\mathbf{dom}\, f^\star = -\mathbf{S}^n_{++}$.

Example 2.24 *Indicator function.* Let I_S be the indicator function of a (not necessarily convex) set $S \subseteq \mathbf{R}^n$, *i.e.*, $I_S(x) = 0$ on $\mathbf{dom}\, I_S = S$. Its conjugate is

$$I^\star_S(y) = \sup_{x \in S} y^Tx,$$

which is the support function of the set S.

Example 2.25 *Log-sum-exp function.* To derive the conjugate of $f(x) = \log\sum_{i=1}^n e^{x_i}$, we first determine the values of y for which the maximum over x of $x^Ty - f(x)$ is attained. By setting the gradient with respect to x equal to zero, we obtain the condition

$$y_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}, \quad i = 1, \ldots, n.$$

These equations are solvable for x if and only if $y \succ 0$ and $\mathbf{1}^Ty = 1$. By substituting the expression for y_i into in $x^Ty - f(x)$ we obtain $f^\star(y) = \sum_i y_i \log y_i$. This expression for $f^\star$ is still correct if some components of y are zero, as long as $y \succeq 0$ and $\mathbf{1}^Ty = 1$, and we interpret $0\log 0$ as 0.

In fact the domain of $f^\star$ is exactly given by $\mathbf{1}^Ty = 1$, $y \succeq 0$. Suppose that a component of y is negative, say, $y_k < 0$. Then we can show that $x^Ty - f(x)$ is unbounded above by choosing $x_k = -t$, and $x_i = 0$, $i \neq k$, and letting t go to infinity.

If $y \succeq 0$ but $\mathbf{1}^Ty \neq 1$, one chooses $x = t\mathbf{1}$, so that

$$x^Ty - f(x) = t\mathbf{1}^Ty - t - \log n,$$

which grows unboundedly as $t \to \infty$.

In summary,

$$f^\star(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i & \text{if } y \succeq 0 \text{ and } \mathbf{1}^T y = 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Example 2.26 *Norm and norm squared.* Let $\|\cdot\|$ be a norm on $\mathbf{R}^n$, with dual norm $\|\cdot\|_*$. We will show that the conjugate of $f(x) = \|x\|$ is

$$f^\star(y) = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & \text{otherwise} \end{cases}$$

in other words, the conjugate of a norm is the indicator function of the dual norm unit ball. If $\|y\|_* > 1$, then by definition of the dual norm, there is a $z \in \mathbf{R}^n$ with $\|z\| \le 1$ and $y^T z > 1$. Taking $x = tz$ and letting $t \to \infty$, we have

$$y^T x - \|x\| = t(y^T z - \|z\|) \to \infty,$$

which shows that $f^\star(y) = \infty$. Conversely, if $\|y\|_* \le 1$, then we have $x^T y \le \|x\|\|y\|_*$ for all x, which implies for all x, $y^T x - \|x\| \le 0$. Therefore $x = 0$ is the value that maximizes $y^T x - \|x\|$, with maximum value 0.

Now consider the function $f(x) = (1/2)\|x\|^2$. We will show that its conjugate is $f^\star(y) = (1/2)\|y\|_*^2$. From $y^T x \le \|y\|_* \|x\|$, we conclude

$$y^T x - (1/2)\|x\|^2 \le \|y\|_* \|x\| - (1/2)\|x\|^2$$

for all x. The righthand side is a quadratic function of $\|x\|$, which has maximum value $(1/2)\|y\|_*^2$. Therefore for all x, we have

$$y^T x - (1/2)\|x\|^2 \le (1/2)\|y\|_*^2,$$

which shows that $f^\star(y) \ge (1/2)\|y\|_*^2$.

To show the other inequality, let x be any vector with $y^T x = \|y\|_* \|x\|$, scaled so that $\|x\| = \|y\|_*$. Then we have, for this x,

$$y^T x - (1/2)\|x\|^2 = \|y\|_* \|x\| - (1/2)\|x\|_* = (1/2)\|y\|_*^2.$$

This shows that $f^\star(y) \le (1/2)\|y\|_*^2$.

Example 2.27 *Revenue and profit functions.* We consider a business or enterprise that consumes n resources and produces a product that can be sold. We let $r = (r_1, \ldots, r_n)$ denote the vector of resource quantities consumed, and $S(r)$ denote the sales revenue derived from the product produced (as a function of the resources consumed). Now let p_i denote the price (per unit) of resource i, so the total amount paid for resources by the enterprise is $p^T r$. The profit derived by the firm is then $S(r) - p^T r$. Let us fix the prices of the resources, and ask what is the maximum profit that can be made, by wisely choosing the quantities of resources consumed. This maximum profit is given by

$$M(p) = \sup_r \left(S(r) - p^T r \right).$$

The function $M(p)$ gives the maximum profit attainable, as a function of the resource prices. In terms of conjugate functions, we can express M as

$$M(p) = (-S)^\star(-p).$$

Thus the maximum profit (as a function of resource prices) is closely related to the conjugate of gross sales (as a function of resources consumed).

2.3.2 Basic properties

Fenchel's inequality

From the definition of conjugate function, we immediately obtain the inequality

$$f(x) + f^\star(y) \geq x^T y$$

for all x, y. This is called *Fenchel's inequality* (or *Young's inequality* if f is differentiable).

For example with $f(x) = (1/2)x^T Q x$, where $Q \in \mathbf{S}^n_{++}$, we obtain the inequality

$$x^T y \leq (1/2)x^T Q x + (1/2)y^T Q^{-1} y.$$

Conjugate of the conjugate

The examples above, and the name 'conjugate', suggest that the conjugate of the conjugate of a convex function is the original function. This is the case provided a technical condition holds: if f is convex, and $\mathbf{epi}\, f$ is a closed set, then $f^{\star\star} = f$. For example, if $\mathbf{dom}\, f = \mathbf{R}^n$, then we have $f^{\star\star} = f$, *i.e.*, the conjugate of the conjugate of f is f again (see exercises).

This means that if f is convex and $\mathbf{epi}\, f$ is closed, then for each x there exists a y such that Fenchel's inequality is tight.

Differentiable functions

The conjugate of a differentiable function f is also called the *Legendre transform* of f. (To distinguish the general definition from the differentiable case, the term *Fenchel conjugate* is sometimes used instead of conjugate.)

Suppose f is convex and differentiable, with $\mathbf{dom}\, f = \mathbf{R}^n$. Any maximizer $x^\star$ of $x^T y - f(x)$ satisfies $y = \nabla f(x^\star)$, and conversely, if $x^\star$ satisfies $y = \nabla f(x^\star)$, then $x^\star$ maximizes $x^T y - f(x)$. Therefore, if $y = \nabla f(x^\star)$, we have

$$f^\star(y) = x^{\star T} \nabla f(x^\star) - f(x^\star).$$

This allows us to determine $f^\star(y)$ for any y for which we can solve the gradient equation $y = \nabla f(z)$ for z.

We can express this another way. Let $z \in \mathbf{R}^n$ be arbitrary and define $y = \nabla f(z)$. Then we have

$$f^\star(y) = z^T \nabla f(z) - f(z).$$

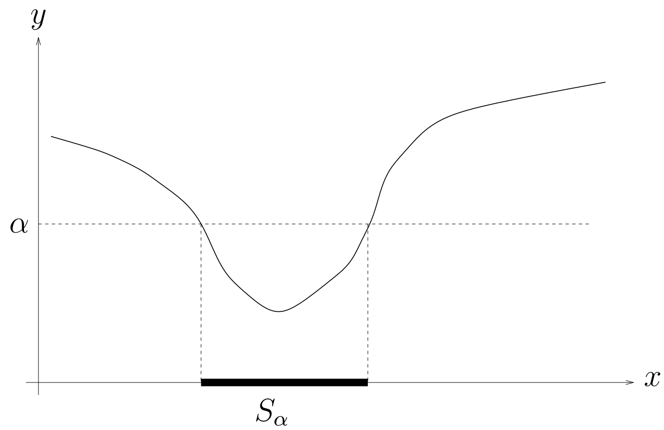


Figure 2.5: A quasiconvex function on $\mathbf{R}$. For each α, the α-sublevel set S_α is convex, *i.e.*, an interval.

Scaling and composition with affine transformation

For $a > 0$ and $b \in \mathbf{R}$, the conjugate of $g(x) = af(x) + b$ is $g^\star(y) = af^\star(y/a) - b$.

Suppose $A \in \mathbf{R}^{n\times n}$ is nonsingular and $b \in \mathbf{R}^n$. Then the conjugate of $g(x) = f(Ax + b)$ is

$$g^\star(y) = f^\star(A^{-T}y) - b^T A^{-T}y,$$

with $\mathbf{dom}\, g^\star = A^T \mathbf{dom}\, f^\star$.

Sums of independent functions

If $f(u, v) = f_1(u) + f_2(v)$, where f_1 and f_2 are convex functions with conjugates $f_1^\star$ and $f_2^\star$, respectively, then

$$f^\star(w, z) = f_1^\star(w) + f_2^\star(z).$$

In other words, the conjugate of the sum of *independent* convex functions is the sum of the conjugates. ('Independent' means they are functions of different variables.)

2.4 Quasiconvex functions

2.4.1 Definition and examples

A function $f : \mathbf{R}^n \to \mathbf{R}$ is called *quasiconvex* (or *unimodal*) if its domain and all its sublevel sets

$$S_\alpha = \{x \in \mathbf{dom}\, f \mid f(x) \leq \alpha\}$$

are convex. This is illustrated in figure 2.5. A function is *quasiconcave* if $-f$ is quasiconvex, *i.e.*, every superlevel set $\{x \mid f(x) \geq \alpha\}$ is convex. A function that is both quasiconvex and quasiconcave is called *quasilinear*.

Convex functions have convex sublevel sets, and hence are quasiconvex. But simple examples, such as the one shown in figure 2.5, show that the converse is not true.

Example 2.28 Some examples on $\mathbf{R}$:

- *Logarithm.* $\log x$ on $\mathbf{R}_{++}$ is quasiconvex (and quasiconcave, hence quasilinear).
- *Ceiling function.* $\mathrm{ceil}(x) = \min\{z \in \mathbf{Z} \mid z \geq x\}$ is quasiconvex (and quasiconcave).

These examples show that quasiconvex functions can be concave, or discontinuous (indeed, integer valued). We now give some examples on $\mathbf{R}^n$.

Example 2.29 *Length of a vector.* We define the *length* of $x \in \mathbf{R}^n$ as the largest index of a nonzero component, *i.e.*,

$$f(x) = \max\{\ k \leq n \mid x_i = 0 \text{ for } i = k+1, \ldots, n\ \}.$$

This function is quasiconvex on $\mathbf{R}^n$; its sublevel sets are subspaces.

Example 2.30 Consider $f : \mathbf{R}^2 \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^2_+$ and $f(x_1, x_2) = x_1 x_2$. The function is neither convex or concave since its Hessian

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

is indefinite; it has one positive and one negative eigenvalue. The function f is quasiconcave, however, since the superlevel sets

$$\left\{ x \in \mathbf{R}^2_+ \;\middle|\; x_1 x_2 \geq \alpha \right\}$$

are convex sets for all α.

Example 2.31 *Linear-fractional function.* The function

$$f(x) = \frac{a^T x + b}{c^T x + d},$$

with $\mathbf{dom}\, f = \{x \mid c^T x + d > 0\}$, is quasiconvex (and quasiconcave).

Its α-sublevel set is

$$\begin{aligned} S_\alpha &= \{x \mid c^T x + d > 0,\ (a^T x + b)/(c^T x + d) \leq \alpha\} \\ &= \{x \mid c^T x + d > 0,\ (a - \alpha c)^T x + b - \alpha d \leq 0\}, \end{aligned}$$

which is convex.

Example 2.32 *Distance ratio function.* Suppose $a, b \in \mathbf{R}^n$, and define

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2},$$

i.e., the ratio of the Euclidean distance to a to the distance to b. Then f is quasiconvex on the halfspace $\{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$. To see this, we consider the α-sublevel set of f, with $\alpha \leq 1$ since $f(x) \leq 1$ on the halfspace $\{x \mid \|x - a\|_2 \leq \|x - b\|_2\}$. This sublevel set is the set of points satisfying

$$\|x - a\|_2 \leq \alpha \|x - b\|_2.$$

Squaring both sides, and rearranging terms, we see that this equivalent to

$$(1 - \alpha^2) x^T x - 2(a - \alpha^2 b)^T x + a^T a - \alpha^2 b^T b \leq 0.$$

This describes a convex set (in fact a Euclidean ball) if $\alpha \leq 1$.

Example 2.33 *Internal rate of return.* Let $x = (x_0, x_1, \ldots, x_n)$ denote a cash flow sequence over n periods, where $x_i > 0$ means a payment to us in period i, and $x_i < 0$ means a payment by us in period i. We define the *present value* of a cash flow, with interest rate $r \geq 0$, to be

$$\mathrm{PV}(x, r) = \sum_{i=0}^{n} (1+r)^{-i} x_i.$$

(The factor $(1+r)^{-i}$ is a discount factor for a payment by or to us in period i.)

Now we consider cash flows for which $x_0 < 0$ and $x_0 + x_1 + \cdots + x_n > 0$. This means that we start with an investment of $|x_0|$ in period 0, and that the total of the remaining cash flow, $x_1 + \cdots + x_n$, (not taking any discount factors into account) exceeds our initial investment.

For such a cash flow, $\mathrm{PV}(x, 0) > 0$ and $\mathrm{PV}(x, r) \to x_0 < 0$ as $r \to \infty$, so it follows that for at least one $r \geq 0$, we have $\mathrm{PV}(x, r) = 0$. We define the *internal rate of return* of the cash flow as the smallest interest rate $r \geq 0$ for which the present value is zero:

$$\mathrm{IRR}(x) = \inf\{r \geq 0 \mid \mathrm{PV}(x, r) = 0\}.$$

Internal rate of return is a quasiconcave function. To see this, we note that

$$\mathrm{IRR}(x) \geq R \iff \mathrm{PV}(x, r) \geq 0 \text{ for } 0 \leq r \leq R.$$

The lefthand side defines the R-superlevel set of IRR. The righthand side is the intersection of the sets $\{x \mid \mathrm{PV}(x, r) \geq 0\}$ indexed by r, over the range $0 \leq r \leq R$. For each r, $\mathrm{PV}(x, r) \geq 0$ defines a halfspace, so the righthand side defines a convex set.

2.4.2 Basic properties

The examples above show that quasiconvexity is a considerable generalization of convexity. Still, many of the properties of convex functions hold, or have analogs, for quasiconvex functions. For example, there is a variation on Jensen's inequality that characterizes quasiconvexity. A function f is quasiconvex if and only if for any $x,\ y \in \mathbf{dom}\, f$ and $0 \leq \theta \leq 1$,

$$f(\theta x + (1-\theta) y) \leq \max\{f(x), f(y)\}, \tag{2.18}$$

i.e., the value of the function on a segment does not exceed the maximum of its values at the endpoints. The inequality (2.18) is sometimes called Jensen's inequality for quasiconvex functions. This is illustrated in figure 2.6.

Example 2.34 *Rank of positive semidefinite matrix.* The function $\mathbf{Rank}(X)$ is quasiconcave on $\mathbf{S}^n_+$. This follows from the modified Jensen inequality (2.18),

$$\mathbf{Rank}(X+Y) \geq \max\{\mathbf{Rank}(X), \mathbf{Rank}(Y)\}$$

which holds for positive semidefinite $X,\ Y \in \mathbf{S}^n$.

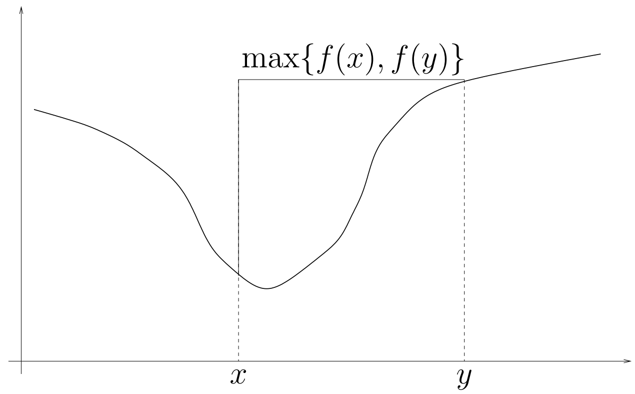


Figure 2.6: A quasiconvex function on $\mathbf{R}$. The value of f between x and y is no more than $\max\{f(x), f(y)\}$.

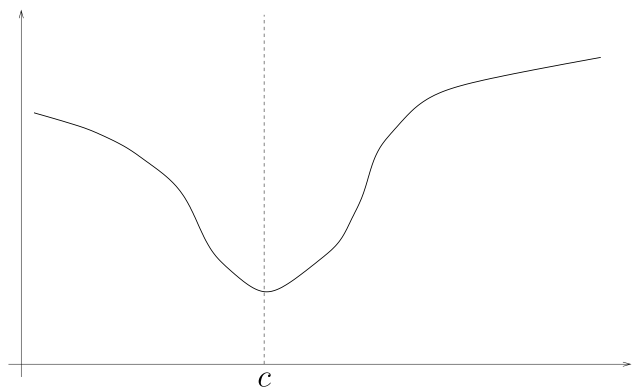


Figure 2.7: A quasiconvex function on $\mathbf{R}$. The function is nonincreasing for $t \leq c$ and nondecreasing for $t \geq c$.

Like convexity, quasiconvexity is characterized by the behavior of a function f on lines: f is quasiconvex if and only if its restriction to any line intersecting its domain is quasiconvex. We can give a simple characterization of quasiconvex functions on $\mathbf{R}$. We consider continuous functions, since stating the conditions in the general case is cumbersome. A continuous function $f : \mathbf{R} \to \mathbf{R}$ is quasiconvex if and only if at least one of following conditions holds:

- f is nondecreasing
- f is nonincreasing
- there is a point $c \in \mathbf{dom}\, f$ such that for $t \leq c$ (and $t \in \mathbf{dom}\, f$), f is nonincreasing, and for $t \geq c$ (and $t \in \mathbf{dom}\, f$), f is nondecreasing.

The point c can be chosen as any point which is a global minimizer of f. Figure (2.7) illustrates this.

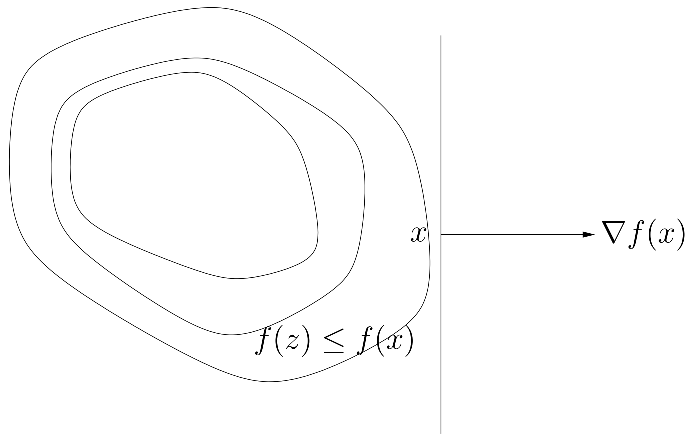


Figure 2.8: Three level curves of a quasiconvex function f are shown. The vector $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{z \mid f(z) \leq f(x)\}$ at x.

2.4.3 Differentiable quasiconvex functions

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable. Then f is quasiconvex if and only if $\mathbf{dom}\, f$ is convex and for all x, $y \in \mathbf{dom}\, f$

$$f(y) \leq f(x) \Longrightarrow \nabla f(x)^T (y - x) \leq 0. \tag{2.19}$$

This is the analog of inequality (2.2), for quasiconvex functions.

The condition (2.19) has a simple geometric interpretation when $\nabla f(x) \neq 0$. It states that $\nabla f(x)$ defines a supporting hyperplane to the sublevel set $\{y \mid f(y) \leq f(x)\}$, as illustrated in figure 2.8.

Now suppose f is twice differentiable. If f is quasiconvex, then for all $x \in \mathbf{dom}\, f$, and all $y \in \mathbf{R}^n$, we have

$$y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y \geq 0. \tag{2.20}$$

For a quasiconvex function on $\mathbf{R}$, this reduces to the simple condition

$$f'(x) = 0 \Longrightarrow f''(x) \geq 0,$$

i.e., at any point with zero slope, the second derivative is nonnegative. For a quasiconvex function on $\mathbf{R}^n$, the interpretation of the condition (2.20) is a bit more complicated. As in the case $n = 1$, we conclude that whenever $\nabla f(x) = 0$, we must have $\nabla^2 f(x) \succeq 0$. When $\nabla f(x) \neq 0$, the condition (2.20) means that $\nabla^2 f(x)$ is positive semidefinite on the $(n-1)$-dimensional subspace $\nabla f(x)^\perp$. This implies that $\nabla^2 f(x)$ can have at most one negative eigenvalue.

As a (partial) converse, if f satisfies

$$y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y > 0 \tag{2.21}$$

for all $x \in \mathbf{dom}\, f$ and all $y \in \mathbf{R}^n$, $y \neq 0$, then f is quasiconvex. This condition is the same as requiring $\nabla^2 f(x)$ to be positive definite for any point with $\nabla f(x) = 0$, and for all other points, requiring $\nabla^2 f(x)$ to be positive definite on the $(n-1)$-dimensional subspace $\nabla f(x)^\perp$.

Proof. By restricting the function to an arbitrary line, it suffices to consider the one-dimensional case.

We first show that if $f : \mathbf{R} \to \mathbf{R}$ is quasiconvex on an interval (a, b), then it must satisfy (2.20), *i.e.*, if $f'(c) = 0$ with $c \in (a, b)$, then we must have $f''(c) \geq 0$. If $f'(c) = 0$ with $c \in (a, b)$, $f''(c) < 0$, then for small positive ϵ we have $f(c - \epsilon) < f(c)$ and $f(c+\epsilon) < f(c)$. It follows that the sublevel set $\{x \mid f(x) \leq f(c) - \epsilon\}$ is disconnected for small positive ϵ, and therefore not convex, which contradicts our assumption that f is quasiconvex.

Now we show that if the condition (2.21) holds, then f is quasiconvex. Assume that (2.21) holds, *i.e.*, for each $c \in (a, b)$ with $f'(c) = 0$, we have $f''(c) > 0$. This means that whenever the function f' crosses the value 0, it is strictly increasing. Therefore it can cross the value 0 at most once. If f' does not cross the value 0 at all, then f is either nonincreasing on nondecreasing on (a, b), and therefore quasiconvex. Otherwise it must cross it exactly once, say at $c \in (a, b)$. Since $f''(c) > 0$, it follows that $f'(t) \leq 0$ for $a < t \leq c$, and $f'(t) \geq 0$ for $c \leq t < b$. This shows that f is quasiconvex.

2.4.4 Operations that preserve quasiconvexity

Nonnegative weighted maximum

A nonnegative weighted maximum of quasiconvex functions, *i.e.*,

$$f = \max\{w_1 f_1, \ldots, w_m f_m\},$$

with $w_i \geq 0$ and f_i quasiconvex, is quasiconvex. The property extends to the general pointwise supremum

$$f(x) = \sup_{y \in C} w(y) g(x, y)$$

where $w(y) \geq 0$ and $g(x, y)$ is quasiconvex in x for each y. This fact can be easily verified: $f(x) \leq \alpha$ if and only if

$$w(y) g(x, y) \leq \alpha, \text{ for all } y \in C,$$

i.e., the α-sublevel set of f is the intersection of the α-sublevel sets of the functions $w(y)g(x, y)$ in the variable x.

Example 2.35 *Generalized eigenvalue.* The *largest generalized eigenvalue* of a pair of symmetric matrices (X, Y), with $Y \succ 0$, is defined as

$$\lambda_{\max}(X, Y) = \sup_{u \neq 0} \frac{u^T X u}{u^T Y u} = \sup\{\lambda \mid \det(\lambda Y - X) = 0\}.$$

This function is quasiconvex on $\mathbf{dom}\, f = \mathbf{S}^n \times \mathbf{S}^n_{++}$.

To see this we note that

$$\lambda_{\max}(X, Y) = \sup_{u \neq 0} \frac{u^T X u}{u^T Y u}.$$

For each $u \neq 0$, the function $u^T X u / u^T Y u$ is linear fractional, hence quasiconvex, so $\lambda_{\max}$ is the supremum of a family of quasiconvex functions.

Composition

If $h : \mathbf{R}^n \to \mathbf{R}$ is quasiconvex and $g : \mathbf{R} \to \mathbf{R}$ is nondecreasing, then $f(x) = g(h(x))$ is quasiconvex.

The composition of a quasiconvex function with an affine or linear-fractional transformation yields a quasiconvex function. If g is quasiconvex, then $f(x) = g(Ax+b)$ is quasiconvex, and $\tilde{f}(x) = g((Ax+b)/(c^T x + d))$ is quasiconvex on the set

$$\left\{x \;\middle|\; c^T x + d > 0,\ (Ax+b)/(c^T x + d) \in \mathbf{dom}\, g\right\}.$$

Minimization

If $g(x, y)$ is quasiconvex jointly in x and y and C is a convex set, then the function

$$f(x) = \inf_{y \in C} g(x, y)$$

is quasiconvex.

Proof. From the definition of f, $f(x) \le \alpha$ if and only if for any $\epsilon > 0$ there exists a $y \in C$ with $g(x, y) \le \alpha + \epsilon$. Let x_1 and x_2 be two points in the α-sublevel set of f. Then for any $\epsilon > 0$, there exists y_1, $y_2 \in C$ with

$$g(x_1, y_1) \le \alpha + \epsilon, \quad g(x_2, y_2) \le \alpha + \epsilon,$$

and since g is quasiconvex in x and y, we also have

$$g(\theta x_1 + (1-\theta)x_2, \theta y_1 + (1-\theta) y_2) \le \alpha + \epsilon,$$

for $0 \le \theta \le 1$. Hence $f(\theta x_1 + (1-\theta)x_2) \le \alpha$.

2.4.5 Representation via family of convex functions

In the sequel, it will be convenient to represent the sublevel sets of a quasiconvex function f (which are convex) via inequalities of convex functions. We seek a family of convex functions $\phi_t : \mathbf{R}^n \to \mathbf{R}$, indexed by $t \in \mathbf{R}$, with

$$f(x) \le t \Leftrightarrow \phi_t(x) \le 0, \tag{2.22}$$

i.e., the t-sublevel set of the quasiconvex function f is the 0-sublevel set of the convex function ϕ_t. Evidently ϕ_t must satisfy the property that for all $x \in \mathbf{R}^n$, $\phi_t(x) \le 0 \implies \phi_s(x) \le 0$ for $s \ge t$. This is satisfied if for each x, $\phi_t(x)$ is a nonincreasing function of t, *i.e.*, $\phi_s(x) \le \phi_t(x)$ whenever $s \ge t$.

To see that such a representation always exists, we can take

$$\phi_t(x) = \begin{cases} 0 & f(x) \le t, \\ \infty & \text{otherwise.} \end{cases}$$

i.e., ϕ_t is the indicator function of the t-sublevel of f. Obviously this representation is not unique; for example if the sublevel sets of f are closed, we can take

$$\phi_t(x) = \mathbf{dist}\,(x, \{z \mid f(z) \le t\}).$$

We are usually interested in a family ϕ_t with nice properties, such as differentiability.

Example 2.36 *Convex over concave function.* Suppose p is a convex function, q is a concave function, with $p(x) \geq 0$ and $q(x) > 0$ on a convex set C. Then the function f defined by $f(x) = p(x)/q(x)$, on C, is quasiconvex.

Here we have

$$f(x) \leq t \iff p(x) - tq(x) \leq 0,$$

so we can take $\phi_t(x) = p(x) - tq(x)$ for $t \geq 0$. For each t, ϕ_t is convex and for each x, $\phi_t(x)$ is decreasing in t.

2.5 Log-concave and log-convex functions

2.5.1 Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is *logarithmically concave* or *log-concave* if $f(x) > 0$ for all $x \in \mathbf{dom}\, f$ and $\log f$ is concave. It is said to be *logarithmically convex* or *log-convex* if $\log f$ is convex. Thus f is log-convex if and only if $1/f$ is log-concave. It is convenient to allow f to take on the value zero, in which case we take $\log f(x) = -\infty$. In this case we say f is log-concave if the extended-valued function $\log f$ is concave.

From the composition rules we know that e^h is convex if h is convex, so a log-convex function is convex. Similarly, a concave function is log-concave. It is also clear that a log-convex function is quasiconvex and a log-concave function is quasiconcave.

Example 2.37 Some simple examples are:

- *Affine function.* The function $f(x) = a^T x + b$ is log-concave on $\{x | a^T x + b > 0\}$.
- *Powers.* $f(x) = x^a$, on $\mathbf{R}_{++}$ is log-convex for $a \leq 0$ and log-concave for $a \geq 0$.
- *Exponentials.* $f(x) = e^{ax}$ is log-convex and log-concave.
- The cumulative distribution function for a Gaussian density, $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} e^{-u^2/2}\, du$, is log-concave (see exercise 2.54).
- *Gamma function.* $\Gamma(x) = \displaystyle\int_0^\infty u^{x-1} e^{-u}\, du$ is log-convex for $x \geq 1$ (see exercise 2.53).
- *Determinant.* $\det X$ is log concave on $\mathbf{S}^n_{++}$.

Example 2.38 *Log-concave density functions.* Many common probability density functions are log-concave. Two examples are the multivariate normal distribution,

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} e^{-\frac{1}{2}(x-\bar{x})^T \Sigma^{-1}(x-\bar{x})}$$

(where $\bar{x} \in \mathbf{R}^n$ and $\Sigma \in \mathbf{S}^n_{++}$), and the exponential distribution on $\mathbf{R}^n_+$,

$$f(x) = \left(\prod_{i=1}^{n} \lambda_i \right) e^{-\lambda^T x}$$

(where $\lambda \succeq 0$). Another example is the uniform distribution over a convex set C,

$$f(x) = \begin{cases} 1/\alpha & x \in C \\ 0 & x \notin C \end{cases}$$

where α is the Lebesgue measure of C. In this case $\log f$ takes on the value $-\infty$ outside C, and $-\log \alpha$ on C, hence is concave.

As a more exotic example consider the Wishart distribution, defined as follows. Let $x_1, \ldots, x_p \in \mathbf{R}^n$ be independent Gaussian random vectors with zero mean and covariance $\Sigma \in \mathbf{S}^n$, with $p > n$. The random matrix $X = \sum_{i=1}^p x_i x_i^T$ has the Wishart density

$$f(X) = a\,(\det X)^{(p-n-1)/2}\, e^{-\frac{1}{2}\,\mathbf{Tr}\,\Sigma^{-1}X},$$

with $\mathbf{dom}\, f = \mathbf{S}^n_{++}$, and a is a positive constant. The Wishart density is log-concave, since

$$\log f(X) = \log a + \frac{p-n-1}{2} \log\det X - \frac{1}{2}\,\mathbf{Tr}\,\Sigma^{-1}X$$

is a concave function of X.

2.5.2 Properties

Twice differentiable log-convex/concave functions

Suppose f is twice differentiable, and $\mathbf{dom}\, f$ is convex. Then f is log-convex if and only if for all $x \in \mathbf{dom}\, f$,

$$f(x)\nabla^2 f(x) \succeq \nabla f(x) \nabla f(x)^T,$$

and log-concave if and only if for all $x \in \mathbf{dom}\, f$,

$$f(x)\nabla^2 f(x) \preceq \nabla f(x) \nabla f(x)^T.$$

Multiplication, addition, and integration

Log-convexity and log-concavity are evidently closed under multiplication and positive scaling.

Simple examples show that the sum of log-concave functions is not, in general, log-concave. Log-convexity, however, is preserved under sums. Let f and g be log-convex functions, *i.e.*, $F = \log f$ and $G = \log g$ are convex. From the composition rules for convex functions, it follows that

$$\log(\exp F + \exp G) = \log(f + g)$$

is convex. Therefore the sum of two log-convex functions is log-convex.

More generally, if $f(x, y)$ is log-convex in x for each $y \in C$ then

$$\int_C f(x, y)\, dy$$

is log-convex.

Example 2.39 *Laplace transform of nonnegative function and the moment and cumulant generating functions.* Suppose $p : \mathbf{R}^n \to \mathbf{R}$ satisfies $p(x) \geq 0$ for all x. The Laplace transform of p,

$$P(z) = \int p(x) e^{-z^T x} \, dx,$$

is log-convex on $\mathbf{R}^n$. (Here $\mathbf{dom}\, P$ is, naturally, $\{z \mid P(z) < \infty\}$.)

Now suppose p is a density, *i.e.*, satisfies $\int p(x)\, dx = 1$. The function $M(z) = P(-z)$ is called the *moment generating function* of the density. It gets its name from the fact that the moments of the density can be found from the derivatives of the moment generating function, evaluated $z = 0$, *e.g.*,

$$\nabla M(0) = \mathbf{E}\, v, \quad \nabla^2 M(0) = \mathbf{E}\, vv^T,$$

where v is a random variable with density p.

The function $\log M(z)$, which is convex, is called the *cumulant generating function* for p, since its derivatives give the cumulants of the density. For example, the first and second derivatives of the cumulant generating function, evaluated at zero, are the mean and covariance of the associated random variable:

$$\nabla \log M(0) = \mathbf{E}\, v, \quad \nabla^2 \log M(0) = \mathbf{E}(v - \mathbf{E}\, v)(v - \mathbf{E}\, v)^T.$$

Integration of log-concave functions

In some special cases log-concavity is preserved by integration. If $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is log-concave, then

$$g(x) = \int f(x, y) \, dy \tag{2.23}$$

is a log-concave function of x (on $\mathbf{R}^n$). (The integration here is over $\mathbf{R}^m$.) A proof of this result is not simple; see the notes and references.

This result has many important consequences, some of which we describe in the rest of this section. It implies, for example, that marginal distributions of log-concave probability densities are log-concave. It also implies that log-concavity is closed under convolution, *i.e.*, if f and g are log-concave on $\mathbf{R}^n$, then the convolution

$$(f * g)(x) = \int f(x - y) g(y) \, dy$$

(To see this, note that $g(y)$ and $f(x-y)$ are log-concave in (x, y), hence the product $f(x - y)g(y)$ is; then the result (2.23) applies.)

Suppose $C \subseteq \mathbf{R}^n$ is a convex set and w is a random vector in $\mathbf{R}^n$ with log-concave probability density p. Then the function

$$f(x) = \mathbf{Prob}(x + w \in C)$$

is log-concave in x. To see this, express f as

$$f(x) = \int g(x + z) p(z) \, dz$$

where g is defined as

$$g(u) = \begin{cases} 1 & u \in C \\ 0 & u \notin C \end{cases}$$

(which is log-concave) and apply (2.23).

Example 2.40 The *cumulative distribution function* (CDF) of a probability density function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$F(x) = \mathbf{Prob}(w \preceq x)$$

where w is a random variable with density f. If f is log-concave, then F is log-concave. We have already encountered a special case: the CDF a Gaussian random variable,

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

is log-concave. (See example 2.37 and exercise 2.54.)

Example 2.41 *Yield function.* Let $x \in \mathbf{R}^n$ denote the nominal or target value of a set of parameters of a product that is manufactured. Variation in the manufacturing process causes the parameters of the product, when manufactured, to have the value $x + w$, where $w \in \mathbf{R}^n$ is a random vector that represents manufacturing variation, and is usually assumed to have zero mean. The *yield* of the manufacturing process, as a function of the nominal parameter values, is given by

$$Y(x) = \mathbf{Prob}(x + w \in S)$$

where $S \subseteq \mathbf{R}^n$ denotes the set of acceptable parameter values for the product, *i.e.*, the product *specifications*.

If the density of the manufacturing error w is log-concave (for example, Gaussian) and the set S of product specifications is convex, then the yield function Y is log-concave. This implies that the α-*yield region*, defined as the set of nominal parameters for which the yield exceeds α, is convex. For example, the 95% yield region

$$\{x \mid Y(x) \geq 0.95\} = \{x \mid \log Y(x) \geq \log 0.95\}$$

is convex, since it is a superlevel set of the concave function $\log Y$.

Example 2.42 *Volume of polyhedron.* Let $A \in \mathbf{R}^{m \times n}$. Define

$$\mathcal{P}_u = \{\ x \ \in \mathbf{R}^n \mid Ax \preceq u\ \}.$$

Then $\mathbf{vol}\, P_u$ is a log-concave function of u.

To prove this, note that the function

$$\Psi(x, u) = \begin{cases} 1 & Ax \preceq u \\ 0 & \text{otherwise} \end{cases}$$

is log-concave. By (2.23), we conclude that

$$\int \Psi(x, u)\, dx = \mathbf{vol}\, P_u$$

is log-concave.

2.6 Convexity with respect to generalized inequalities

We now consider generalizations of the notions of monotonicity and convexity, using generalized inequalities instead of the usual ordering on $\mathbf{R}$.

2.6.1 Monotonicity with respect to a generalized inequality

Suppose $K \subseteq \mathbf{R}^n$ is a proper cone with associated generalized inequality $\preceq_K$. A function $f : \mathbf{R}^n \to \mathbf{R}$ is called *K-nondecreasing* if

$$x \preceq_K y \Longrightarrow f(x) \leq f(y),$$

and *K-increasing* if

$$x \preceq_K y, \; x \neq y \Longrightarrow f(x) < f(y).$$

We define *K-nonincreasing* and *K-decreasing* functions in a similar way.

Example 2.43 A function $f : \mathbf{R}^n \to \mathbf{R}$ is nondecreasing with respect to $\mathbf{R}^n_+$ if and only if

$$x_1 \leq y_1, \ldots, x_n \leq y_n \;\; \Rightarrow \;\; f(x) \leq f(y)$$

for all $x, \; y$. This is the same as saying that f, when restricted to any component x_i (*i.e.*, x_i is considered the variable while x_j for $j \neq i$ are fixed), is nondecreasing.

Example 2.44 *Matrix monotone functions.* A function $f : \mathbf{S}^n \to \mathbf{R}$ is called *matrix monotone* (increasing, decreasing) if it is monotone with respect to the positive semidefinite cone.

Some examples of matrix monotone functions of the variable $X \in \mathbf{S}^n$:

- $\mathbf{Tr}\, WX$, where $W \in \mathbf{S}^n$, is matrix nondecreasing if $W \succeq 0$, and matrix increasing if $W \succ 0$ (it is matrix nonincreasing if $W \preceq 0$, and matrix decreasing if $W \prec 0$).
- $\mathbf{Tr}\, X^{-1}$ is matrix decreasing on the set of positive definite matrices.
- $\det X$ is matrix increasing on the set of positive semidefinite matrices.

Gradient conditions for monotonicity

Recall that a differentiable function $f : \mathbf{R} \to \mathbf{R}$ is nondecreasing if and only if $f'(x) \geq 0$ for all $x \in \mathbf{dom}\, f$, and increasing if $f'(x) > 0$ for all $x \in \mathbf{dom}\, f$ (but the converse is not true). These conditions are readily extended to the case of monotonicity with respect to a generalized inequality. A differentiable function f is K-nondecreasing if and only if

$$\nabla f(x) \succeq_{K^*} 0 \tag{2.24}$$

for all $x \in \mathbf{dom}\, f$. Note the difference with the simple scalar case: the gradient must be nonnegative in the *dual* inequality. For the strict case, we have the following: If

$$\nabla f(x) \succ_{K^*} 0 \tag{2.25}$$

for all $x \in \mathbf{dom}\, f$ then f is K-increasing. As in the scalar case, the converse is not true.

Proof. First, assume that f satisfies (2.24) for all x, but is not K-nondecreasing, *i.e.*, there exist x, y with $x \preceq_K y$ and $f(y) < f(x)$. By differentiability of f there exists a $t \in [0,1]$ with

$$\frac{d}{dt} f(x + t(y-x)) = \nabla f(x + t(y-x))^T (y - x) < 0.$$

Since $y - x \in K$ this means

$$-\nabla f(x + t(y-x)) \notin K^*,$$

which contradicts our assumption that (2.24) is satisfied everywhere. In a similar way it can be shown that (2.25) implies f is K-increasing.

It is also straightforward to see that it is necessary that (2.24) hold everywhere. Assume (2.24) does not hold for $x = z$. By the definition of dual cone this means there exists a $v \in K$ with

$$\nabla f(z)^T v < 0.$$

Now consider $h(t) = f(z + tv)$ as a function of t. We have $h'(0) = \nabla f(z)^T v < 0$, and therefore there exists $t > 0$ with $h(t) = f(z + tv) < h(0) = f(z)$, which means f is not K-nondecreasing.

2.6.2 Convexity with respect to a generalized inequality

Definition and examples

Suppose $K \subseteq \mathbf{R}^m$ is a proper cone with associated generalized inequality $\preceq_K$. We say $f : \mathbf{R}^n \to \mathbf{R}^m$ is *K-convex* if for all x, y, and $0 \le \theta \le 1$,

$$f(\theta x + (1-\theta)y) \;\preceq_K\; \theta f(x) + (1-\theta) f(y).$$

The function is *strictly K-convex* if

$$f(\theta x + (1-\theta)y) \;\prec_K\; \theta f(x) + (1-\theta) f(y).$$

for all $x \ne y$ and $0 < \theta < 1$. These definitions reduce to ordinary convexity and strict convexity when $m = 1$ (and $K = \mathbf{R}_+$).

Example 2.45 *Convexity with respect to componentwise inequality.* A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is convex with respect to componentwise inequality (*i.e.*, the generalized inequality induced by $\mathbf{R}^m_+$) if and only if for all x, y and $0 \le \theta \le 1$,

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta) f(y),$$

i.e., each component f_i is a convex function. The function f is strictly convex with respect to componentwise inequality if and only if each component f_i is strictly convex.

Example 2.46 *Matrix convexity.* Suppose f is a symmetric matrix valued function, *i.e.*, $f : \mathbf{R}^n \to \mathbf{S}^m$. f is convex with respect to matrix inequality (*i.e.*, the positive semidefinite cone) if

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta) f(y)$$

for any x and y, and for $\theta \in [0,1]$. This is sometimes called *matrix convexity*. An equivalent definition is that the scalar function $z^T f(x) z$ is convex for all vectors z. (This is often a good way to prove matrix convexity). A matrix function is strictly matrix convex if

$$f(\theta x + (1-\theta)y) \prec \theta f(x) + (1-\theta) f(y)$$

when $x \neq y$ and $0 < \theta < 1$ or, equivalently, if $z^T f z$ is strictly convex for every $z \neq 0$.

Some examples:

- $f(X) = XX^T$ where $X \in \mathbf{R}^{n\times m}$ is convex, since for fixed z the function $z^T XX^T z = \|X^T z\|_2^2$ is a convex quadratic function of the components of X. For the same reason, $f(X) = X^2$ is convex on $\mathbf{S}^n$.
- The functions X^p is matrix convex on the set of positive definite matrices for $1 \le p \le 2$ or $-1 \le p \le 0$, and matrix concave for $0 \le p \le 1$.
- The function $f(X) = e^X$ is not convex on the space of symmetric matrices.

Many of the results for convex functions have extensions to K-convex functions. As a simple example, a function is K-convex if and only if its restriction to any line in its domain is K-convex. In the rest of this section list a few results for K-convexity that we will use later; more results are explored in the exercises.

Dual characterization of K-convexity

A function f is K-convex if and only if for every $w \succeq_{K^*} 0$, the (real-valued) function $w^T f$ is convex (in the ordinary sense); f is strictly K-convex if and only if for every nonzero $w \succeq_{K^*} 0$ the function $w^T f$ is strictly convex. (These follow directly from the definitions and properties of dual inequality.)

Differentiable K-convex functions

A differentiable function f is K-convex if and only if for all $x,\ y \in \mathbf{dom}\, f$,

$$f(y) \succeq_K f(x) + Df(x)(y-x).$$

(Here $Df(x) \in \mathbf{R}^{m\times n}$ is the derivative or Jacobian matrix of f at x; see §B.3.2.)

The function f is strictly K-convex if for all $x,\ y \in \mathbf{dom}\, f$ with $x \neq y$,

$$f(y) \succ_K f(x) + Df(x)(y-x).$$

(As in the scalar case, a function can be strictly K-convex without satisfying this inequality everywhere.)

Composition theorem

If $g : \mathbf{R}^n \to \mathbf{R}^p$ is K-convex, $h : \mathbf{R}^p \to \mathbf{R}$ is convex, and $\tilde{h}$ (the extended-value extension of h) is K-nondecreasing, then $h(g(x))$ is convex. This generalizes the fact that a nondecreasing convex function of a convex function is convex. The condition that $\tilde{h}$ be K-nondecreasing implies that $\mathbf{dom}\, h - K = \mathbf{dom}\, h$.

Proof. The function g is K-convex. Therefore, $g(\theta x_1 + (1-\theta)x_2) \preceq_K \theta g(x_1) + (1-\theta)g(x_2)$. The function h is K-increasing, and convex. Therefore

$$\begin{aligned} h(g(\theta x_1 + (1-\theta)x_2)) &\leq h(\theta g(x_1) + (1-\theta)g(x_2)) \\ &\leq \theta h(g(x_1)) + (1-\theta)h(g(x_2)). \end{aligned}$$

Example 2.47 The quadratic matrix function $g : \mathbf{R}^{m\times n} \to \mathbf{S}^n$ defined by

$$g(X) = X^T AX + B^T X + X^T B + C,$$

where $A \in \mathbf{S}^m$, $B \in \mathbf{R}^{m\times n}$, and $C \in \mathbf{S}^n$, is convex when $A \succeq 0$.

The function $h : \mathbf{S}^n \to \mathbf{R}$ defined by $h(Y) = -\log\det(-Y)$ is convex and increasing on $\mathbf{dom}\, h = \{Y \in \mathbf{S}^n \mid Y \prec 0\}$.

By the composition theorem, we conclude that

$$f(X) = -\log\det(-(X^T AX + B^T X + X^T B + C))$$

is convex on

$$\mathbf{dom}\, f = \left\{ X \in \mathbf{R}^{m\times n} \;\middle|\; X^T AX + B^T X + X^T B + C \prec 0 \right\}.$$

(This generalizes the fact that $-\log(-(ax^2+bx+c))$ is convex on $\{x \in \mathbf{R} \mid ax^2+bx+c < 0\}$, provided $a \geq 0$.)

Exercises

Definition of convexity

Exercise 2.1. *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?

Exercise 2.2. Suppose f is convex, $\mathbf{dom}\, f = \mathbf{R}^n$, and bounded above on $\mathbf{R}^n$. Show that it is constant.

Exercise 2.3. Let $f : \mathbf{R} \to \mathbf{R}$ be a convex function with $\mathbf{dom}\, f = \mathbf{R}$. Suppose $a < x < b$.

(a) Show that

$$f(x) \leq \frac{b-x}{b-a} f(a) + \frac{x-a}{b-a} f(b).$$

(b) Show that

$$\frac{f(x)-f(a)}{x-a} \leq \frac{f(b)-f(a)}{b-a} \leq \frac{f(b)-f(x)}{b-x}.$$

Draw a sketch that illustrates this inequality.

(c) Suppose now f is differentiable. Show that

$$f'(a) \leq \frac{f(b)-f(a)}{b-a}.$$

Exercise 2.4. *Expressing a convex function as the pointwise supremum of a family of affine functions.* Let $f : \mathbf{R}^n \to \mathbf{R}$ be a convex function. Define $\tilde{f} : \mathbf{R}^n \to \mathbf{R}$ as

$$\tilde{f}(x) = \sup\{g(z) \mid g \text{ affine},\ g(z) \leq f(z) \text{ for all } z\}.$$

The function $\tilde{f}$ is the pointwise supremum of all affine functions that are global underestimators of f. Show that $f(x) = \tilde{f}(x)$ for $x \in \mathbf{int}\,\mathbf{dom}\, f$.

Exercise 2.5. *Second order conditions for a function restricted to an affine set.* When is a twice differentiable function $f : \mathbf{R}^n \to \mathbf{R}$ convex when restricted to the affine set $\mathcal{A} = \{Fz + g \mid z \in \mathbf{R}^m\}$? What if the set is described as $\mathcal{A} = \{x \mid Ax = b\}$?

Exercise 2.6. *An extension of Jensen's inequality.* One interpretation of Jensen's inequality is: 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: the larger the variance of v, the larger should be $\mathbf{E}\, f(x_0 + v)$.

(a) Show this is false in general. In other words find zero mean random variables v and w, with $\mathbf{var}(v) > \mathbf{var}(w)$, 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 $\mathbf{E}\, f(x_0 + av)$ is monotonic increasing in $a \geq 0$, when f is convex and v is zero mean.

Exercise 2.7. 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 on the endpoints of the segment: for every $x,\ y \in \mathbf{R}^n$,

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

Exercise 2.8. *Convexity of inverse.* Suppose $f : \mathbf{R} \to \mathbf{R}$ is increasing and convex on its domain (a, b). Let g denote its inverse, *i.e.*, the function with domain $(f(a), f(b))$ with $g(f(x)) = x$ for $a < x < b$. What can you say about convexity or concavity of g?

Exercise 2.9. Suppose $f : \mathbf{R} \to \mathbf{R}$ is convex. Show that its "running average", *i.e.*,

$$F(x) = \frac{1}{x} \int_0^x f(t)\, dt$$

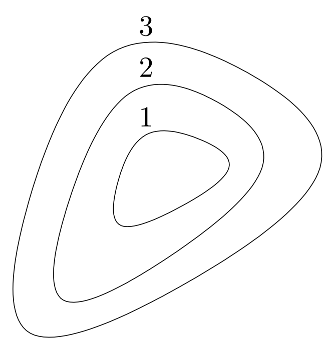
(with $F(0) = 0$) is also convex.

Exercise 2.10. *Interpolation with a convex function.* Suppose you are given k points $x_1, \ldots, x_k \in \mathbf{R}^n$, and k numbers $y_1, \ldots, y_k \in \mathbf{R}$. When does there exist a convex function f such that $f(x_i) = y_i$, $i = 1, \ldots, k$? Describe the condition geometrically.

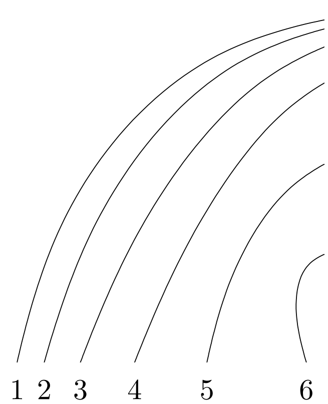
Also consider the following extension: you are given not only $x_i \in \mathbf{R}^n$ and $y_i \in \mathbf{R}$ but also $g_i \in \mathbf{R}^n$. Under what conditions does there exist a convex function f such that $f(x_i) = y_i$, $\nabla f(x_i) = g_i$ $i = 1, \ldots, k$? Describe the conditions geometrically.

In both cases, you do not have to give all the details of how to construct such a function when the conditions hold.

Exercise 2.11. *Level sets of convex, concave, quasiconvex, and quasiconcave functions.* Some level sets of a function f are shown below. The curve labeled 1 shows $\{x \mid f(x) = 1\}$, etc.



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



Exercise 2.12. *Monotone mappings.* A function $\psi : \mathbf{R}^n \to \mathbf{R}^n$ is called *monotone* if for all $x, y \in \mathbf{R}^n$,

$$(\psi(x) - \psi(y))^T (x - y) \geq 0.$$

(Note that 'monotone' as defined here is not the same as the definition given in §2.6.1. Unfortunately both meanings are widely used, so you have to figure out from context which is meant.)

Suppose that f is a differentiable convex function on $\mathbf{R}^n$. Show that the function $\psi(x) = \nabla f(x)$ is monotone.

Is the converse true, *i.e.*, is every monotone mapping the gradient of a convex function?

Exercise 2.13. Suppose $f, g : \mathbf{R}^n \to \mathbf{R}$, f is convex, g is concave, $\mathbf{dom}\, f = \mathbf{dom}\, g = \mathbf{R}^n$, and for all x, $g(x) \leq f(x)$. In other words, the concave function g is an underestimator of the convex function f.

Show that there exists an affine function h such that for all x, $g(x) \leq h(x) \leq f(x)$. In other words: we can fit an affine function between the concave and the convex function.

Exercise 2.14. *Kullback-Leibler divergence and information inequality.* Let $p \in \mathbf{R}^n$ and $q \in \mathbf{R}^n$ be two discrete probability distributions (*i.e.*, two vectors satisfying $p \succeq 0$, $q \succeq 0$, $\mathbf{1}^T p = 1$, and $\mathbf{1}^T q = 1$). The *relative entropy* or *Kullback-Leibler divergence* between p and q is defined as

$$D_{\mathrm{kl}}(p, q) = \sum_{i=1}^n p_i \log(p_i / q_i).$$

(a) Show that $D_{\mathrm{kl}}(p, q)$ is jointly convex in p and q.

(b) Prove the information inequality: $D_{\mathrm{kl}}(p, q) \geq 0$ for all probability distributions p and q. *Hint.* The Kullback-Leibler divergence can be expressed as

$$D_{\mathrm{kl}}(p, q) = f(p) - f(q) + \nabla f(p)^T (q - p)$$

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

Exercise 2.15. Suppose that $f : \mathbf{R}^n \to \mathbf{R}$ is convex and symmetric, *i.e.*, $f(Px) = f(x)$ for every permutation P. Show that f always has a minimizer of the form $\alpha\mathbf{1}$.

Examples

Exercise 2.16. *A family of concave utility functions.* For $0 < \alpha \leq 1$ let

$$u_\alpha(x) = \frac{x^\alpha - 1}{\alpha},$$

with $\mathbf{dom}\, u_\alpha = \mathbf{R}_+$. We also define $u_0(x) = \log x$ (with $\mathbf{dom}\, u_0 = \mathbf{R}_{++}$).

(a) Show that for $x > 0$, $u_0(x) = \lim_{\alpha \to 0} u_\alpha(x)$.

(b) Show that u_α 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_α 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.*

Exercise 2.17. For each of the following functions determine whether it is convex, concave, quasiconvex, quasiconcave, log-convex, or log-concave.

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

(b) $f(x_1, x_2) = x_1 x_2$ on $x_1, x_2 > 0$.

(c) $f(x_1, x_2) = 1/x_1 x_2$ on $x_1, x_2 > 0$.

(d) $f(x_1, x_2) = x_1/x_2$ on $x_1, x_2 > 0$.

(e) $f(x_1, x_2) = x_1^2/x_2$ on $x_2 > 0$.

(f) $f(x_1, x_2) = x_1^\alpha x_2^{1-\alpha}$, where $0 \le \alpha \le 1$, on $x_1, x_2 > 0$.

Exercise 2.18. Show that the following functions are concave on $\mathbf{R}_+^n$.

(a) $f(x) = \left(\sum_{i=1}^n x_i^{1/2}\right)^2$. Generalize to $f(x) = \left(\sum_i c_i x_i^{1/2}\right)^2$ and $f(x) = (\sum_i c_i x_i^p)^{1/p}$ where $c_i > 0$ and $0 < p < 1$.

(b) *Harmonic mean.* $f(x) = (\sum_{i=1}^n 1/x_i)^{-1}$.

Exercise 2.19. Show that the function $f(X) = \mathbf{Tr}\, X^{-1}$ is convex on $\mathbf{dom}\, f = \mathbf{S}_{++}^n$.
Hint. Adapt the proof of convexity of $\log\det X^{-1}$ in §2.1.5.

Exercise 2.20. *Nonnegative weighted sums and integrals.*

(a) Show that $f(x) = \sum_{i=1}^r \alpha_i x_{[i]}$ is a convex function of x, where $\alpha_1 \ge \alpha_2 \ge \cdots \ge \alpha_r \ge 0$, and $x_{[i]}$ denotes the ith largest component of x.
(You can use the fact that $f(x) = \sum_{i=1}^r x_{[i]}$ is convex on $\mathbf{R}^n$.)

(b) Let $T(x, \omega)$ denote the trigonometric polynomial

$$T(x, \omega) = x_1 + x_2 \cos\omega + x_3 \cos 2\omega + \cdots + x_n \cos(n-1)\omega.$$

Show that the function

$$f(x) = -\int_0^{2\pi} \log T(x, \omega)\, d\omega$$

is convex on $\{ x \in \mathbf{R}^n \mid T(x, \omega) > 0,\ 0 \le \omega \le 2\pi \}$.

Exercise 2.21. *Composition with an affine function.* Show that the following functions are convex.

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

(b) $f(x) = -(\det(A_0 + x_1 A_1 + \cdots + x_n A_n))^{1/m}$ on $\{ x \mid A_0 + x_1 A_1 + \cdots + x_n A_n \succ 0 \}$ where $A_i \in \mathbf{S}^m$.

(c) $f(X) = \mathbf{Tr}\,(A_0 + x_1 A_1 + \cdots + x_n A_n)^{-1}$ on $\{ x \mid A_0 + x_1 A_1 + \cdots + x_n A_n \succ 0 \}$ where $A_i \in \mathbf{S}^m$.
(Use the fact that $\mathbf{Tr}\, X^{-1}$ is convex on $\mathbf{S}_{++}^m$; see exercise 2.19.)

Exercise 2.22. *Pointwise maximum and supremum.* Show that the following functions are convex.

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

(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 ith 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 decreasing order.)

Exercise 2.23. *Composition rules.* Show that the following functions are convex.

(a) $f(x) = -\log(-\log(\sum_{i=1}^m e^{a_i^T x+b_i}))$.

(You can use the fact that $\log \sum_i e^{y_i}$ is convex.)

(b) $f(x,u,v) = -\sqrt{uv - x^T x}$ on $\mathbf{dom}\, f = \{ (x,u,v) \mid uv \geq x^T x, u, v \geq 0 \}$.

(Use the fact that $x^T x/u$ is convex in (x,u) for $u > 0$, and that $-\sqrt{xy}$ is convex on $\{ (x,y) \in \mathbf{R}^2 \mid x, y \geq 0 \}$.

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

(d) $f(x,t) = -(t^p - \|x\|_p^p)^{1/p}$ where $p > 1$ and $\mathbf{dom}\, f = \{(x,t) \mid t \geq \|x\|_p\}$.

(You can use the fact that $\|x\|_p^p/u^{p-1}$ $(p > 1)$ is convex in (x,u) for $u > 0$, and that $-x^{1/p}y^{1-1/p}$ is convex on $\{ (x,y) \in \mathbf{R}^2 \mid x, y \geq 0 \}$.)

(e) $f(x,t) = -\log(t^p - \|x\|_p^p)$ where $p > 1$ and $\mathbf{dom}\, f = \{(x,t) \mid t > \|x\|_p\}$.

Exercise 2.24. *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 $\{(x,t) \mid t > \|x\|_p\}$.

(b) Show that

$$f(x) = \frac{\|Ax + b\|_2^2}{c^T x + d}$$

is convex on $\{ x \mid c^T x + d > 0 \}$, where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, $c \in \mathbf{R}^n$ and $d \in \mathbf{R}$.

Exercise 2.25. *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 $\mathbf{R}$. Prove the following.

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

(b) if f, g are concave, positive, with one increasing and the other decreasing then fg concave.

(c) if f convex, increasing, positive, g concave, decreasing, positive, then f/g convex.

Exercise 2.26. *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 $\mathbf{Prob}(x = a_i) = p_i$, $i = 1, \ldots, n$. For each of the following functions of p on the probability simplex $\{p \in \mathbf{R}_+^n \mid \mathbf{1}^T p = 1\}$, is it convex, concave, quasiconvex, quasiconcave? If the function is positive, is it log-convex or log-concave?

(a) $\mathbf{E}\, x$

(b) $\mathbf{Prob}(x \geq \alpha)$

(c) $\mathbf{Prob}(\alpha \leq x \leq \beta)$

(d) $\sum_i p_i \log p_i$ (the negative entropy of the distribution)

(e) $\mathbf{var}\, x$

(f) Quartile$(x) = \inf\{\alpha \mid \mathbf{Prob}(x \leq \alpha) \geq 0.25\}$.

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

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

$$\inf\left\{\beta - \alpha \mid \mathbf{Prob}(x \in [\alpha, \beta]) \geq 0.9\right\}.$$

Exercise 2.27. *More functions of eigenvalues.* Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_n$ denote the eigenvalues of a matrix $X \in \mathbf{S}^n$. We have already seen several functions of the eigenvalues that are convex or concave functions of X, for example:

- λ_1 is convex; λ_n is concave
- $\lambda_1 + \cdots + \lambda_n = \mathbf{Tr}\, X$ is linear
- for $X \succ 0$, $\left(\prod_{i=1}^n \lambda_i\right)^{1/n} = (\det X)^{1/n}$ and $\sum_{i=1}^n \log \lambda_i = \log\det X$ are concave

In this problem we explore some more functions of eigenvalues, by exploiting variational characterizations.

(a) *Sum of k largest eigenvalues.* Use the variational characterization

$$\sum_{i=1}^k \lambda_i = \sup\{\mathbf{Tr}\, V^T X V \mid V \in \mathbf{R}^{n\times k},\ V^T V = I\ \},$$

to show that $\sum_{i=1}^k \lambda_i$ is convex.

(b) *Geometric mean of k smallest eigenvalues.* For $X \succ 0$, we have

$$\left(\prod_{i=n-k+1}^n \lambda_i\right)^{1/k} = \inf\{\ \left(\mathbf{Tr}\, V^T X V\right)/k \mid V \in \mathbf{R}^{n\times k},\ \det V^T V = 1\ \}.$$

Use this to show that $\left(\prod_{i=n-k+1}^n \lambda_i\right)^{1/k}$ is a concave functions of X.

(c) *Log of product of k smallest eigenvalues.* For $X \succ 0$, we have

$$\prod_{i=n-k+1}^n \lambda_i = \inf\{\ \det \mathbf{diag}\, V^T X V \mid V \in \mathbf{R}^{n\times k},\ V^T V = I\ \}.$$

Use this to show that $\sum_{i=n-k+1}^n \log \lambda_i(X)$ is a concave function of X.

Exercise 2.28. *Another composition rule.* Prove the following, which generalizes the composition rules given in §2.2.4.

Suppose

$$f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

where $h : \mathbf{R}^k \to \mathbf{R}$, $g_i : \mathbf{R}^n \to \mathbf{R}$. If h is convex and for each i, either h is nondecreasing in the ith argument and g_i is convex, or, h_i is nonincreasing in the ith argument and g_i is concave, and g_i are convex, then f is convex.

Exercise 2.29. *A composition rule based on copositivity.* A matrix $A \in \mathbf{S}^n$ is *copositive* if $x^T Ax \geq 0$ for all $x \succeq 0$. In general it is difficult computational task to check whether a matrix is copositive, but a simple sufficient condition is that A is the sum of a positive semidefinite matrix and an elementwise nonnegative matrix.

Establish the following composition theorem: Consider

$$f(x) = h(g_1(x), g_2(x), \ldots, g_k(x))$$

where $h : \mathbf{R}^k \to \mathbf{R}$, $g_i : \mathbf{R}^n \to \mathbf{R}$. You can assume that $\mathbf{dom}\, h = \mathbf{R}^k$ and $\mathbf{dom}\, g_i = \mathbf{R}^n$, so you don't have to worry about the issue of $\mathbf{dom}\, f$.

Suppose h is monotone nondecreasing in each component, $\nabla^2 h(z)$ is copositive for all $z \in \mathbf{dom}\, h$, each g_i is convex and nondecreasing. Then f is convex.

Exercise 2.30. *Largest homogenous lower bound.* Let f be a convex function. Define the function g as

$$g(x) = \inf_{\alpha > 0} \frac{f(\alpha x)}{\alpha}.$$

(a) Show that g is the largest homogeneous lower bound on f: if h is homogeneous (*i.e.*, satisfies $h(\alpha x) = \alpha h(x)$ for $\alpha \geq 0$) and $h(x) \leq f(x)$ for all x, then we have $h(x) \leq g(x)$ for all x.

(b) Show that g is convex.

Exercise 2.31. *Convex hull or envelope of a function.* The convex hull or convex envelope of a function $f : \mathbf{R}^n \to \mathbf{R}$ is defined as

$$g(x) = \inf\{t \mid (x,t) \in \mathbf{Co}\,\mathbf{epi}\, f\}.$$

Geometrically, the epigraph of g is the convex hull of the epigraph of f.

Show that g is the largest convex function that is less than or equal to f. In other words, show that if h is convex and satisfies $h(x) \leq f(x)$ for all x, then $h(x) \leq g(x)$ for all x.

Exercise 2.32. *The Minkowski function.* Let C be a closed convex set. We define the *Minkowski function* for C as

$$M_C(x) = \inf\{t > 0 \mid t^{-1}x \in C\}.$$

(a) Draw a picture giving a geometric interpretation of how to find $M_C(x)$.

(b) Show that M_C is homogeneous, *i.e.*, $M_C(\alpha x) = \alpha M_C(x)$ for $\alpha > 0$.

(c) What is $\mathbf{dom}\, M_C$? (Recall that $\inf \emptyset = +\infty$.)

(d) When do we have $M_C(x) = 0$?

(e) Show that M_C is a convex function.

Exercise 2.33. *Support function calculus.* Recall that the support function of a set $C \subseteq \mathbf{R}^n$ is defined as $S_C(y) = \sup\{y^T x \mid x \in C\}$. On page 56 we showed that S_C is a convex function.

(a) Let A denote the closure of the convex hull of B. Show that $S_A = S_B$.

(b) Show that $S_{A+B} = S_A + S_B$.

(c) Show that $S_{A \cup B} = S_A + S_B$.

(d) Let A and B be convex and closed. Show that $A \subseteq B$ if and only if $S_A(x) \leq S_B(x)$ for all x.

Exercise 2.34. *Representation of piecewise linear convex functions.* In this problem you will show that any piecewise linear convex function can be expressed as the maximum of a finite number of affine functions. (This is the converse of the result in example 2.5.) Suppose $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, is piecewise linear. This means the following: we have a partition

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

where $\mathbf{int}\, X_i \neq \emptyset$ and $\mathbf{int}\, X_i \cap \mathbf{int}\, X_j$ for $i \neq j$, and f is given by $f(x) = a_i^T x + b_i$ for $x \in X_i$. We assume that $a_i^T x + b_i = a_j^T x + b_j$ whenever $x \in X_i \cap X_j$.

Assume now that f is convex. Show that f can be expressed as the maximum of a finite set of affine functions.

Hint: define $\tilde{f}$ as $\tilde{f}(x) = \max_{i=1,\ldots,L}\{a_i^T x + b_i\}$, and then show that $f = \tilde{f}$.

Conjugate functions

Exercise 2.35. Derive the conjugates of the following functions.

(a) *Max function.* $f(x) = \max_i x_i$ 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* $\mathbf{R}$. $f(x) = \max_{i=1,\ldots,m} a_i x + b_i$ 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 where $f(x) = a_k x + b_k$.)

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

(e) *Geometric mean.* $f(x) = -(\prod x_i)^{1/n}$ on $\mathbf{R}_+^n$.

Exercise 2.36. Show that the conjugate of $f(X) = \mathbf{Tr}\, X^{-1}$ with $\mathbf{dom}\, f = \mathbf{S}_{++}^n$ is given by

$$f^\star(Y) = -2\,\mathbf{Tr}(-Y)^{1/2}$$

with $\mathbf{dom}\, f^\star = -\mathbf{S}_{++}^n$ (*i.e.*, $\mathbf{dom}\, f^\star = \{Y \in \mathbf{S}^n \mid Y \prec 0\}$).

Exercise 2.37. *Conjugate of conjugate.* Show that if f is convex with closed epigraph, then $f^{\star\star} = f$, *i.e.*, the conjugate of the conjugate of f is f itself.

Exercise 2.38. *Conjugate of convex plus affine function.* Define $g(x) = f(x) + c^T x + d$, where f is convex. Express $g^\star$ in terms of $f^\star$ (and c, d).

Exercise 2.39. *Conjugate of perspective.* Let $f : \mathbf{R}^n \to \mathbf{R}$ be convex, and let $F : \mathbf{R}^{n+1} \to \mathbf{R}$ be its perspective function, defined by $F(x,t) = tf(x/t)$ for $t > 0$. Express the conjugate of F in terms of the conjugate of f.

Exercise 2.40. *Conjugate and minimization.*

(a) Let $f(x,z)$ be convex in (x,z) and define $g(x) = \inf_z f(x,z)$. Express the conjugate $g^\star$ in terms of $f^\star$.

(b) Let f be convex, and define $h(x) = \inf_u\{f(u) \mid Au + b = x\}$. Express $h^\star$ in terms of $f^\star$ (and A, b).

Exercise 2.41. *Gradient and Hessian of conjugate function.* Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is convex and twice differentiable. Suppose $\tilde{y}$ and $\tilde{x}$ are related by $\tilde{y} = \nabla f(\tilde{x})$, and that the gradient mapping $\nabla f : \mathbf{R}^n \to \mathbf{R}^n$ is invertible near $\tilde{x}$. In other words, for each y near $\tilde{y}$, there is a unique x near $\tilde{x}$ such that $y = \nabla f(x)$.

(a) Show that $\nabla f^\star(\tilde{y}) = \tilde{x}$

(b) Show that $\nabla^2 f^\star(\tilde{y}) \nabla^2 f(\tilde{x}) = I$.

Exercise 2.42. *Young's inequality.* Let f be an increasing function on $\mathbf{R}$, with $f(0) = 0$, and let g be its inverse. Define F and G as

$$F(x) = \int_0^x f(a)\, da, \quad G(y) = \int_0^y g(a)\, da.$$

Show that F and G are conjugates.

Give a simple graphical interpretation of Young's inequality

$$xy \leq F(x) + G(y).$$

Quasiconvex functions

Exercise 2.43. *Approximation width.* Let $f_0, \ldots, f_n : \mathbf{R}_+ \to \mathbf{R}$ be given functions. We consider the problem of approximating f_0 as a linear combination of $f_1, \ldots, f_n$. For $x \in \mathbf{R}^n$, we say that $f = x_1 f_1 + \cdots + x_n f_n$ approximates f_0 with tolerance $\epsilon > 0$ over the interval $[0, T]$ if $|f(t) - f_0(t)| \leq \epsilon$ for $0 \leq t \leq T$. Now we choose a fixed tolerance $\epsilon > 0$ and define the *approximation width* as the largest T such that f approximates f_0 over the interval $[0, T]$:

$$W(x) = \sup\{\, T \mid |x_1 f_1(t) + \cdots + x_n f_n(t) - f_0(t)| \leq \epsilon \text{ for } 0 \leq t \leq T \,\}.$$

Show that W is quasiconcave function of x.

Exercise 2.44. *First-order condition for quasiconvexity.* Prove the first-order condition for quasiconvexity given in §2.4.3: A differentiable function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f$ convex, is quasiconvex if and only if for all $x, y \in \mathbf{dom}\, f$,

$$f(y) \leq f(x) \Longrightarrow \nabla f(x)^T (y - x) \leq 0.$$

Hint. It suffices to prove the result for a function on $\mathbf{R}$; the general result follows by restriction to an arbitrary line.

Exercise 2.45. *Second-order conditions for quasiconvexity.* In this problem we derive alternate representations of the second-order conditions for quasiconvexity given in §2.4.3.

(a) Consider the necessary condition: for all $x \in \mathbf{dom}\, f$

$$y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y \geq 0.$$

Show that this condition can be expressed in the following two equivalent ways.

- for all $x \in \mathbf{dom}\, f$, there exists a $\lambda(x) \geq 0$ such that

$$\nabla^2 f(x) + \lambda(x) \nabla f(x) \nabla f(x)^T \succeq 0.$$

- for all $x \in \mathbf{dom}\, f$, the matrix

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

has at most one negative eigenvalue.

(b) Consider the sufficient condition: for all $x \in \mathbf{dom}\, f$

$$y^T \nabla f(x) = 0 \Longrightarrow y^T \nabla^2 f(x) y > 0.$$

Show that this condition can be expressed in the following two equivalent ways.

- for all $x \in \mathbf{dom}\, f$, there exists a $\lambda(x) \geq 0$ such that

$$\nabla^2 f(x) + \lambda(x) \nabla f(x) \nabla f(x)^T \succ 0$$

- for all $x \in \mathbf{dom}\, f$, the matrix

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

has one nonpositive and n positive eigenvalues.

Exercise 2.46. Use the first and second order conditions for quasiconvexity given in §2.4.3 to verify quasiconvexity of the function $f(x) = -x_1 x_2$, with $\mathbf{dom}\, f = \mathbf{R}^2_{++}$.

Exercise 2.47. *Quasilinear functions with domain* $\mathbf{R}^n$. A function on $\mathbf{R}$ that is quasilinear (*i.e.*, quasiconvex and quasiconcave) is monotone, *i.e.*, either nondecreasing or nonincreasing. In this problem we consider a generalization of this result to functions on $\mathbf{R}^n$.

Suppose the function $f : \mathbf{R}^n \to \mathbf{R}$ is quasilinear and $\mathbf{dom}\, f = \mathbf{R}^n$. You can assume that f is continuous. Show that it can be expressed as $f(x) = g(a^T x)$, for $g : \mathbf{R} \to \mathbf{R}$ is monotone and $a \in \mathbf{R}^n$. In other words: a quasilinear function wih domain $\mathbf{R}^n$ must be a monotone function of a linear function. (The converse is also true.)

Log-concave and log-convex functions

Exercise 2.48. *Logistic function.* Show that the function $e^x/(1+e^x)$, which is sometimes called the *logistic function*, is log-concave.

Exercise 2.49. *Harmonic mean.* The harmonic mean of $x_1, \ldots, x_n > 0$ is defined as

$$H(x) = \frac{1}{1/x_1 + \cdots + 1/x_n}.$$

Show that the harmonic mean is log-concave.

Exercise 2.50. *Subtracting a constant from a log-concave function.* Show that if $f : \mathbf{R}^n \to \mathbf{R}$ is log-concave and $a \geq 0$, then the function $g = f - a$ is log-concave, where $\mathbf{dom}\, g = \{x \in \mathbf{dom}\, f \mid f(x) > a\}$.

Exercise 2.51. Let p be a polynomial of $x \in \mathbf{R}$, with all its roots real. Show that it is log-concave on any interval on which it is positive.

Exercise 2.52. Let $y \in \mathbf{R}^n$ be a random variable with a log-concave probability density. Assume $g_i(x, y)$, $i = 1, \ldots, r$, are concave functions on $\mathbf{R}^m \times \mathbf{R}^n$. Then

$$h(x) = \mathbf{Prob}(g_1(x, y) \geq 0, \ldots, g_r(x, y) \geq 0)$$

is log-concave in x. A special case is

$$h(x) = \mathbf{Prob}(g_1(x) \geq y_1, \ldots, g_r(x) \geq y_r)$$

where $g_i(x)$ is concave, and y_i have log-concave denisities. with a log-concave probability distribution.

Exercise 2.53. *Log-convexity of moment functions.* Suppose $f : \mathbf{R}_+ \to \mathbf{R}$ is nonnegative. For $x \geq 0$ define

$$M(x) = \int_0^\infty u^x f(u)\, du.$$

When x is a positive integer, and f is a probability distribution (*i.e.*, with integral one), $M(x)$ is the xth moment of the distribution. Show that M is a log-convex function.

Hint: for each $u \geq 0$, u^x is log-convex on $\mathbf{R}_{++}$.

Use this to show that the Gamma function,

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du,$$

is log-convex for $x \geq 1$.

Exercise 2.54. *Log-concavity of Gaussian CDF.* 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 is log-concave for $x \geq 0$. That leaves us the hard part, which is to show that f is log-concave 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

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

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

Exercise 2.55. *Log-concavity of the CDF of a log-concave probability density.* In this problem we extend the result of exercise 2.54. 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 $f''(x)f(x) \leq (f'(x))^2$.

(a) Express the derivatives of f in terms of the function h. Verify that f is log-concave 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 is log-concave.

Exercise 2.56. *Log-concave measures.* A probability measure π on $\mathbf{R}^n$ is log-concave if

$$\pi((1-\theta)C_1 + \theta C_2) \geq \pi(C_1)^{1-\theta}\pi(C_2)^{\theta}$$

for all convex subsets C_1 and C_2 of $\mathbf{R}^n$ and all $\theta \in [0, 1]$.

Show that when the measure π is given by a density p, *i.e.*, $\pi(A) = \int_A p(x)dx$, then it is log-concave if and only if the density p is a log-concave function.

Exercise 2.57. *More log-concave densities.* Show that the following densities are log-concave.

(a) The gamma density, defined by

$$f(x) = \frac{\alpha^{\lambda}}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}$$

on $\mathbf{R}_+$. The parameters λ and α satisfy $\lambda \geq 1$, $\alpha > 0$.

(b) The multivariate hyperbolic density:

$$f(x) = c e^{-\alpha\left(\delta + (x-\bar{x})^T \Sigma^{-1} (x-\bar{x})\right)^{1/2} + \beta^T (x-\bar{x})}.$$

Here $\Sigma \in \mathbf{S}^n_{++}$, $\beta \in \mathbf{R}^n$, and α and c are positive constants.

(c) The Dirichlet density:

$$f(x) = \frac{\Gamma(\lambda)}{\Gamma(\lambda_1)\cdots\Gamma(\lambda_{n+1})} x_1^{\lambda_1-1} \cdots x_n^{\lambda_n-1} \left(1 - \sum_{i=1}^{n} x_i\right)^{\lambda_{n+1}-1}$$

with domain

$$\mathbf{dom}\, f = \{x \in \mathbf{R}^n_+ \mid \mathbf{1}^T x \leq 1\}.$$

The parameter λ satisfies $\lambda \succeq \mathbf{1}$.

Exercise 2.58. Show that the function

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

is log-concave on $\mathbf{R}^n_{++}$.

Convexity with respect to a generalized inequality

Exercise 2.59. Show that the function $f(X) = X^{-1}$ is matrix convex on the positive definite cone.

Exercise 2.60. *Second order conditions for K-convexity.* Let $K \subseteq \mathbf{R}^m$ be a convex cone that induces a generalized inequality. Show that a twice differentiable function $f : \mathbf{R}^n \to \mathbf{R}^m$ is K-convex if for all $x \in \mathbf{dom}\, f$ and all $y \in \mathbf{R}^n$,

$$\sum_{i,j=1,\ldots,n} \frac{\partial^2 f}{\partial x_i \partial x_j} y_i y_j \succeq_K 0,$$

i.e., the second derivative is a K-nonnegative bilinear form. (Here $\partial^2 f/\partial x_i \partial x_j \in \mathbf{R}^m$, with components $\partial^2 f_k/\partial x_i \partial x_j$, for $k = 1, \ldots, m$; see §B.3.2.)

Exercise 2.61. *Sublevel sets and epigraph of K-convex functions.* Let $K \subseteq \mathbf{R}^m$ indice a generalized inequality, and let $f : \mathbf{R}^n \to \mathbf{R}^m$. For $\alpha \in \mathbf{R}^m$, the α-sublevel set of f (with respect to the generalized inequality $\preceq_K$) is defined as

$$C_\alpha = \{ \ x \in \mathbf{R}^n \mid f(x) \preceq_K \alpha \ \}.$$

The epigraph of f, with respect to $\preceq_K$, is defined as the set

$$\mathbf{epi}_K f = \{ \ (x, t) \in \mathbf{R}^{n+m} \mid f(x) \preceq_K t \ \}.$$

Show the following:

(a) If f is K-convex, then its sublevel sets are convex.

(b) f is K-convex if and only if $\mathbf{epi}_K f$ is a convex set.

Exercise 2.62. *Pointwise maximum of K-convex functions.* The pointwise maximum of two (or more) K-convex functions is K-convex, but the situation is far trickier here than in the scalar case. Recall that a, $b \in \mathbf{R}^m$ need not, in general, have a maximum with respect to K; in other words, there need not exist a $c \in \mathbf{R}^m$ such that $a \preceq_K c$, $b \preceq_K c$, and

$$a \preceq_K d, \quad b \preceq_K d \implies c \preceq_K d.$$

Therefore the pointwise maximum of f_1 and f_2, given by

$$f(x) = \max_K \{f_1(x), f_2(x)\},$$

is defined only if for each x the points $f_1(x)$ and $f_2(x)$ have a maximum.

Show that when the pointwise maximum of two K-convex functions f_1, f_2 does exist, it is K-convex.

Exercise 2.63. *Minimization and K-convexity.* Suppose $g(x, y)$ is K-convex, jointly in x and y, and C is a convex set. Suppose for each x, the set

$$\{g(x, y) \mid y \in C\}$$

has a minimum element (with respect to K), which we denote as $f(y)$. Show that f is K-convex.