Chapter 3

Convex functions

3.1 Basic properties and examples

3.1.1 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

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y). \tag{3.1}$$

Geometrically, this inequality means that the line segment between $(x, f(x))$ and $(y, f(y))$, which is the *chord* from x to y, lies above the graph of f (figure 3.1). A function f is *strictly convex* if strict inequality holds in (3.1) 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 (3.1), 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

Figure 3.1 Graph of a convex function. The chord (*i.e.*, line segment) between any two points on the graph lies above the graph.

all v, the function $g(t) = f(x + tv)$ 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.

3.1.2 Extended-value 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-value 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 need 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 (3.1). 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. (For $\theta = 0$ or $\theta = 1$ the inequality always holds.) 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 (3.1); 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-value 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 3.1 *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-value extension

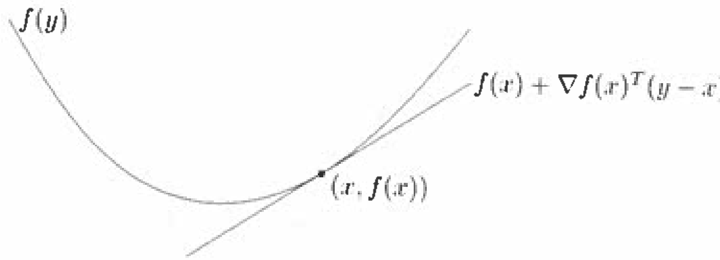

Figure 3.2 If f is convex and differentiable, then $f(x)+\nabla f(x)^T(y-x) \le f(y)$ for all x, $y \in \mathbf{dom}\, f$.

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 f (defined on all of $\mathbf{R}^n$, say) on the set C is the same as minimizing the function $f + \tilde{I}_C$ over all of $\mathbf{R}^n$. Indeed, the function $f + \tilde{I}_C$ is (by our convention) f restricted to the set C.

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

3.1.3 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

$$f(y) \ge f(x) + \nabla f(x)^T(y-x) \tag{3.2}$$

holds for all x, $y \in \mathbf{dom}\, f$. This inequality is illustrated in figure 3.2.

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 (3.2) states that for a convex function, the first-order Taylor approximation is in fact a *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 (3.2) shows that from *local information* about a convex function (*i.e.*, its value and 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. As one simple example, the inequality (3.2) shows that if $\nabla f(x) = 0$, then for all $y \in \mathbf{dom}\, f$, $f(y) \ge f(x)$, *i.e.*, x is a global minimizer of the function 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

$$f(y) > f(x) + \nabla f(x)^T(y - x). \tag{3.3}$$

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

Proof of first-order convexity condition

To prove (3.2), we first consider the case $n = 1$: We show that a differentiable function $f : \mathbf{R} \to \mathbf{R}$ is convex if and only if

$$f(y) \geq f(x) + f'(x)(y - x) \tag{3.4}$$

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

Assume first that f is convex and $x,\ y \in \mathbf{dom}\, f$. Since $\mathbf{dom}\, f$ is convex (*i.e.*, an interval), we conclude that for all $0 < t \leq 1$, $x + t(y - x) \in \mathbf{dom}\, f$, and by convexity of f,

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

If we divide both sides by t, we obtain

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

and taking the limit as $t \to 0$ yields (3.4).

To show sufficiency, assume the function satisfies (3.4) for all x and y in $\mathbf{dom}\, f$ (which is an interval). Choose any $x \neq y$, and $0 \leq \theta \leq 1$, and let $z = \theta x + (1 - \theta)y$. Applying (3.4) 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.

3.1.4 Second-order conditions

We now assume that f is twice differentiable, that is, 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: for all $x \in \mathbf{dom}\, f$,

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

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), which means that the derivative is nondecreasing. The condition $\nabla^2 f(x) \succeq 0$ can be interpreted geometrically as the requirement that the graph of the function have positive (upward) curvature at x. We leave the proof of the second-order condition as an exercise (exercise 3.8).

Similarly, f is concave if and only if $\mathbf{dom}\, f$ 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: for example, 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 3.2 *Quadratic functions.* Consider the quadratic function $f : \mathbf{R}^n \to \mathbf{R}$, with $\mathbf{dom}\, f = \mathbf{R}^n$, given by

$$f(x) = (1/2)x^T P x + q^T x + r,$$

with $P \in \mathbf{S}^n$, $q \in \mathbf{R}^n$, and $r \in \mathbf{R}$. Since $\nabla^2 f(x) = P$ 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 3.1 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.

3.1.5 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 $\mathbf{R}$, for any $a \in \mathbf{R}$.
- *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 $\mathbf{R}_{++}$.

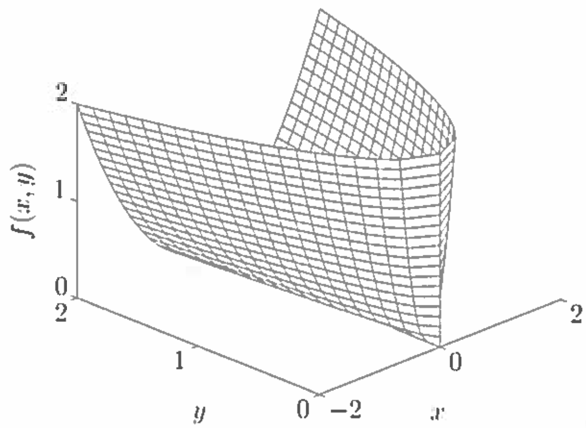

Figure 3.3 Graph of $f(x, y) = x^2/y$.

- *Negative entropy.* $x \log x$ (either on $\mathbf{R}_{++}$, or on $\mathbf{R}_+$, defined as 0 for $x = 0$) is convex.

Convexity or concavity of these examples can be shown by verifying the basic inequality (3.1), 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

$$\mathbf{dom}\, f = \mathbf{R} \times \mathbf{R}_{++} = \{(x, y) \in \mathbf{R}^2 \mid y > 0\},$$

  is convex (figure 3.3).

- *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 differentiable (in fact, analytic) approximation of the max function, since

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

  for all x. (The second inequality is tight when all components of x are equal.) Figure 3.4 shows f for $n = 2$.

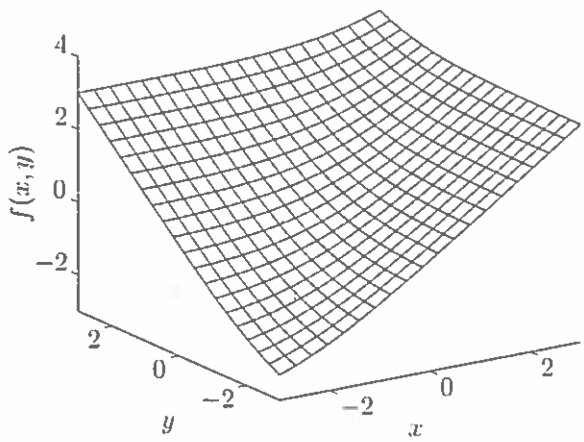

Figure 3.4 Graph of $f(x, y) = \log(e^x + e^y)$.

- *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 $f(X) = \log\det X$ is concave on $\mathbf{dom}\, f = \mathbf{S}^n_{++}$.

Convexity (or concavity) of these examples can be verified in several ways, such as directly verifying the inequality (3.1), 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

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

The inequality follows from the triangle inequality, and the equality follows from homogeneity of a norm.

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

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

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

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

and can be expressed as

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

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

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

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 $f(X) = \log\det X$, we can verify concavity 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

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

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

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

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

3.1.6 Sublevel sets

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

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

Sublevel sets of a convex function are convex, for any value of α. The proof is immediate from the definition of convexity: if x, $y \in C_\alpha$, then $f(x) \le \alpha$ and $f(y) \le \alpha$, and so $f(\theta x + (1-\theta)y) \le \alpha$ for $0 \le \theta \le 1$, 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 $\{x \in \mathbf{dom}\, f \mid f(x) \ge \alpha\}$, is a convex set. The sublevel set property is often a good way to establish convexity of a set, by expressing it as a sublevel set of a convex function, or as the superlevel set of a concave function.

Example 3.3 The geometric and arithmetic means of $x \in \mathbf{R}^n_+$ are, respectively,

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

(where we take $0^{1/n} = 0$ in our definition of G). The arithmetic-geometric mean inequality states that $G(x) \le A(x)$.

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

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

i.e., the set of vectors with geometric mean at least as large as a factor α times the 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.

3.1.7 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

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

which is a subset of $\mathbf{R}^{n+1}$. ('Epi' means 'above' so epigraph means 'above the graph'.) The definition is illustrated in figure 3.5.

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

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

is a convex set.

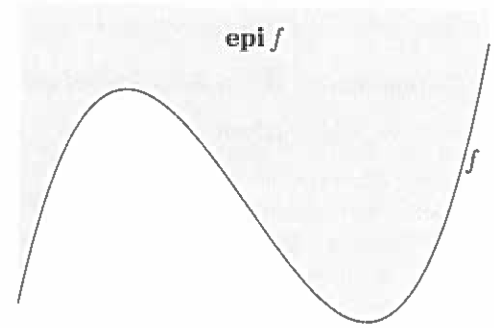

Figure 3.5 Epigraph of a function f, shown shaded. The lower boundary, shown darker, is the graph of f.

Example 3.4 *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 $f(x, y) = x^2/y$, with $\mathbf{dom}\, f = \mathbf{R} \times \mathbf{R}_{++}$.)

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

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

using the Schur complement condition for positive semidefiniteness of a block matrix (see §A.5.5). The last condition is a linear matrix inequality in (x, Y, t), and therefore $\mathbf{epi}\, f$ is convex.

For the special case $n = 1$, the matrix fractional function reduces to the quadratic-over-linear function x^2/y, and the associated LMI representation is

$$\begin{bmatrix} y & x \\ x & t \end{bmatrix} \succeq 0, \qquad y > 0$$

(the graph of which is shown in figure 3.3).

Many results for convex functions can be proved (or interpreted) geometrically using epigraphs, and applying results for convex sets. As an example, consider the first-order condition for convexity:

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

where f is convex and x, $y \in \mathbf{dom}\, f$. We can interpret this basic inequality geometrically in terms of $\mathbf{epi}\, f$. If $(y, t) \in \mathbf{epi}\, f$, then

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

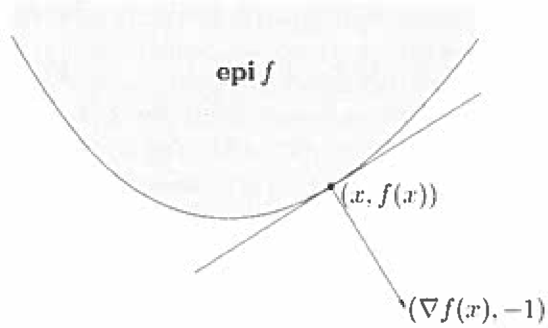

Figure 3.6 For a differentiable convex function f, the vector $(\nabla f(x), -1)$ defines a supporting hyperplane to the epigraph of f at x.

We can express this as:

$$(y,t) \in \mathbf{epi}\, f \implies \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

This means that the hyperplane defined by $(\nabla f(x), -1)$ supports $\mathbf{epi}\, f$ at the boundary point $(x, f(x))$; see figure 3.6.

3.1.8 Jensen's inequality and extensions

The basic inequality (3.1), *i.e.*,

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

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 $\theta_1, \ldots, \theta_k \geq 0$ with $\theta_1 + \cdots + \theta_k = 1$, then

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

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

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

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

$$f(\mathbf{E}\, x) \leq \mathbf{E}\, f(x), \tag{3.5}$$

provided the expectations exist. We can recover the basic inequality (3.1) 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 (3.5) 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

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

Remark 3.2 We can interpret (3.5) 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

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

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.

3.1.9 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:

$$\sqrt{ab} \le (a+b)/2 \tag{3.6}$$

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

$$-\log\left(\frac{a+b}{2}\right) \le \frac{-\log a - \log b}{2}.$$

Taking the exponential of both sides yields (3.6).

As a less trivial example we prove Hölder's inequality: for $p > 1$, $1/p + 1/q = 1$, and x, $y \in \mathbf{R}^n$,

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

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

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

valid for a, $b \ge 0$ and $0 \le \theta \le 1$. 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

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

Summing over i then yields Hölder's inequality.

3.2 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).

3.2.1 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 convex 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 (provided the integral exists).

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 is 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.

3.2.2 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.