2. Convexity

The concept of convexity has far-reaching consequences in variational analysis. In the study of maximization and minimization, the division between problems of convex or nonconvex type is as significant as the division in other areas of mathematics between problems of linear or nonlinear type. Furthermore, convexity can often be introduced or utilized in a local sense and in this way serves many theoretical purposes.

A. Convex Sets and Functions

For any two different points x_0 and x_1 in $I\!R^n$ and parameter value $\tau \in I\!R$ the point

$$x_\tau := x_0 + \tau(x_1 - x_0) = (1-\tau)x_0 + \tau x_1 \tag{2(1)}$$

lies on the line through x_0 and x_1. The entire line is traced as τ goes from $-\infty$ to ∞, with $\tau = 0$ giving x_0 and $\tau = 1$ giving x_1. The portion of the line corresponding to $0 \le \tau \le 1$ is called the *closed line segment* joining x_0 and x_1, denoted by $[x_0, x_1]$. This terminology is also used if x_0 and x_1 coincide, although the line segment reduces then to a single point.

2.1 Definition (convex sets and convex functions)**.**

(a) *A subset C of $I\!R^n$ is convex if it includes for every pair of points the line segment that joins them, or in other words, if for every choice of $x_0 \in C$ and $x_1 \in C$ one has $[x_0, x_1] \subset C$:*

$$(1-\tau)x_0 + \tau x_1 \ \in\ C \ \text{ for all } \ \tau \in (0,1). \tag{2(2)}$$

(b) *A function f on a convex set C is convex relative to C if for every choice of $x_0 \in C$ and $x_1 \in C$ one has*

$$f\big((1-\tau)x_0 + \tau x_1\big) \ \le\ (1-\tau)f(x_0) + \tau f(x_1) \ \text{ for all } \ \tau \in (0,1), \tag{2(3)}$$

and f is strictly convex relative to C if this inequality is strict for points $x_0 \neq x_1$ with $f(x_0)$ and $f(x_1)$ finite.

In plain English, the word 'convex' suggests a bulging appearance, but the convexity of a set in the mathematical sense of 2.1(a) indicates rather the absence of 'dents', 'holes', or 'disconnectedness'. A set like a cube or floating disk in $I\!R^3$, or even a general line or plane, is convex despite aspects of flatness. Note also that the definition doesn't require C to contain two different points, or even a point at all: the empty set is convex, and so is every singleton set $C = \{x\}$. At the other extreme, $I\!R^n$ is itself a convex set.

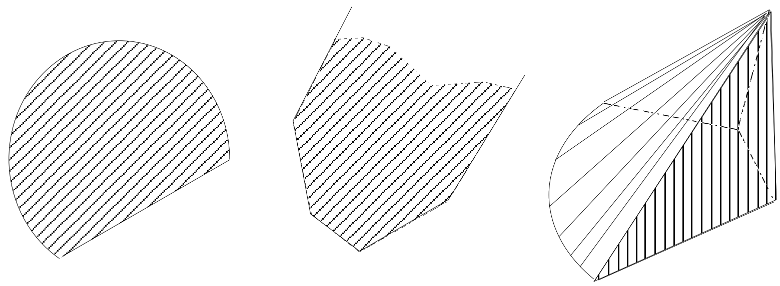

Fig. 2–1. Examples of closed, convex sets, the middle one unbounded.

Many connections between convex sets and convex functions will soon be apparent, and the two concepts are therefore best treated in tandem. In both 2.1(a) and 2.1(b) the τ interval (0,1) could be replaced by [0,1] without really changing anything. Extended arithmetic as explained in Chapter 1 is to be used in handling infinite values of f; in particular the inf-addition convention $\infty - \infty = \infty$ is to be invoked in 2(3) when needed. Alternatively, the convexity condition 2(3) can be recast as the condition that

$$
\begin{aligned}
&f(x_0) < \alpha_0 < \infty, \quad f(x_1) < \alpha_1 < \infty, \quad \tau \in (0,1) \\
&\qquad \Longrightarrow \quad f((1-\tau)x_0 + \tau x_1) \leq (1-\tau)\alpha_0 + \tau\alpha_1.
\end{aligned} \tag{2(4)}
$$

An extended-real-valued function f is said to be *concave* when $-f$ is convex; similarly, f is strictly concave if $-f$ is strictly convex. Concavity thus corresponds to the opposite direction of inequality in 2(3) (with the sup-addition convention $\infty - \infty = -\infty$).

The geometric picture of convexity, indispensable as it is to intuition, is only one motivation for this concept. Equally compelling is an association with 'mixtures', or 'weighted averages'. A *convex combination* of elements $x_0, x_1, \ldots, x_p$ of $I\!R^n$ is a linear combination $\sum_{i=0}^{p} \lambda_i x_i$ in which the coefficients λ_i are nonnegative and satisfy $\sum_{i=0}^{p} \lambda_i = 1$. In the case of just two elements a convex combination can equally well be expressed in the form $(1-\tau)x_0 + \tau x_1$ with $\tau \in [0,1]$ that we've seen so far. Besides having an interpretation as weights in many applications, the coefficients λ_i in a general convex combination can arise as probabilities. For a discrete random vector variable that can take on the value x_i with probability λ_i, the 'expected value' is the vector $\sum_{i=0}^{p} \lambda_i x_i$.

2.2 Theorem (convex combinations and Jensen's inequality)**.**

(a) *A set C is convex if and only if C contains all convex combinations of its elements.*

(b) *A function f is convex relative to a convex set C if and only if for every choice of points $x_0, x_1, \ldots, x_p$ in C one has*

$$f\Big(\sum_{i=0}^{p} \lambda_i x_i\Big) \;\le\; \sum_{i=0}^{p} \lambda_i f(x_i) \;\; \textit{when} \;\; \lambda_i \ge 0, \; \sum_{i=0}^{p} \lambda_i = 1. \qquad 2(5)$$

Proof. In (a), the convexity of C means by definition that C contains all convex combinations of two of its elements at a time. The 'if' assertion is therefore trivial, so we concentrate on 'only if'. Consider a convex combination $x = \lambda_0 x_0 + \cdots + \lambda_p x_p$ of elements x_i of a convex set C in the case where $p > 1$, we aim at showing that $x \in C$. Without loss of generality we can assume that $0 < \lambda_i < 1$ for all i, because otherwise the assertion is trivial or can be reduced notationally to this case by dropping elements with coefficient 0. Writing

$$x = (1-\lambda_p)\sum_{i=0}^{p-1} \lambda_i' x_i + \lambda_p x_p \;\; \text{with} \;\; \lambda_i' = \frac{\lambda_i}{1-\lambda_p},$$

where $0 < \lambda_i' < 1$ and $\sum_{i=0}^{p-1} \lambda_i' = 1$, we see that $x \in C$ if the convex combination $x' = \sum_{i=0}^{p-1} \lambda_i' x_i$ is in C. The same representation can now be applied to x' if necessary to show that it lies on the line segment joining x_{p-1} with some convex combination of still fewer elements of C, and so forth. Eventually one gets down to combinations of only two elements at a time, which do belong to C. A closely parallel argument establishes (b). □

Any convex function f on a convex set $C \subset I\!R^n$ can be identified with a convex function on all of $I\!R^n$ by defining $f(x) = \infty$ for all $x \notin C$. Convexity is thereby preserved, because the inequality 2(3) holds trivially when $f(x_0)$ or $f(x_1)$ is ∞. For most purposes, the study of convex functions can thereby be reduced to the framework of Chapter 1 in which functions are everywhere defined but extended-real-valued.

2.3 Exercise (effective domains of convex functions)**.** *For any convex function $f : I\!R^n \to \overline{I\!R}$, $\operatorname{dom} f$ is a convex set with respect to which f is convex. The proper convex functions on $I\!R^n$ are thus the functions obtained by taking a finite, convex function on a nonempty, convex set $C \subset I\!R^n$ and giving it the value ∞ everywhere outside of C.*

The indicator δ_C of a set $C \subset I\!R^n$ is convex if and only if C is convex. In this sense, convex sets in $I\!R^n$ correspond one-to-one with special convex functions on $I\!R^n$. On the other hand, convex functions on $I\!R^n$ correspond one-to-one with special convex sets in $I\!R^{n+1}$, their epigraphs.

2.4 Proposition (convexity of epigraphs)**.** *A function $f : I\!R^n \to \overline{I\!R}$ is convex if and only if its epigraph set $\operatorname{epi} f$ is convex in $I\!R^n \times I\!R$, or equivalently, its strict epigraph set $\big\{(x,\alpha) \,\big|\, f(x) < \alpha < \infty\big\}$ is convex.*

Proof. The convexity of $\operatorname{epi} f$ means that whenever $(x_0,\alpha_0) \in \operatorname{epi} f$ and $(x_1,\alpha_1) \in \operatorname{epi} f$ and $\tau \in (0,1)$, the point $(x_\tau,\alpha_\tau) := (1-\tau)(x_0,\alpha_0) + \tau(x_1,\alpha_1)$ belongs to $\operatorname{epi} f$. This is the same as saying that whenever $f(x_0) \le \alpha_0 \in I\!R$ and $f(x_1) \le \alpha_1 \in I\!R$, one has $f(x_\tau) \le \alpha_\tau$. The latter is equivalent to the convexity inequality 2(3) or its variant 2(4). The 'strict' version follows similarly. □

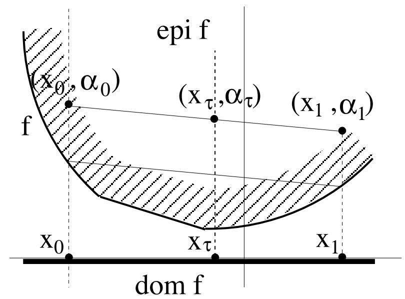

Fig. 2–2. Convexity of epigraphs and effective domains.

For concave functions on $\mathbb{R}^n$ it is the hypograph rather than the epigraph that is a convex subset of $\mathbb{R}^n \times \mathbb{R}$.

2.5 Exercise (improper convex functions)**.** *An improper convex function f must have $f(x) = -\infty$ for all $x \in \operatorname{int}(\operatorname{dom} f)$. If such a function is lsc, it can only have infinite values: there must be a closed, convex set D such that $f(x) = -\infty$ for $x \in D$ but $f(x) = \infty$ for $x \notin D$.*

Guide. Argue first from the definition of convexity that if $f(x_0) = -\infty$ and $f(x_1) < \infty$, then $f(x_\tau) = -\infty$ at all intermediate points x_τ as in 2(1). □

Improper convex functions are of interest mainly as possible by-products of various constructions. An example of an improper convex function having finite as well as infinite values (it isn't lsc) is

$$f(x) = \begin{cases} -\infty & \text{for } x \in (0,\infty) \\ 0 & \text{for } x = 0 \\ \infty & \text{for } x \in (-\infty,0). \end{cases} \qquad 2(6)$$

The chief significance of convexity and strict convexity in optimization derives from the following facts, which use the terminology in 1.4 and its sequel.

2.6 Theorem (characteristics of convex optimization)**.** *In a problem of minimizing a convex function f over $\mathbb{R}^n$ (where f may be extended-real-valued), every locally optimal solution is globally optimal, and the set of all such optimal solutions (if any), namely* $\operatorname{argmin} f$, *is convex.*

Furthermore, if f is strictly convex and proper, there can be at most one optimal solution: the set $\operatorname{argmin} f$, *if nonempty, must be a singleton.*

Proof. If x_0 and x_1 belong to $\operatorname{argmin} f$, or in other words, $f(x_0) = \inf f$ and $f(x_1) = \inf f$ with $\inf f < \infty$, we have for $\tau \in (0,1)$ through the convexity inequality 2(3) that the point x_τ in 2(1) satisfies

$$f(x_\tau) \leq (1-\tau)\inf f + \tau \inf f = \inf f,$$

where strict inequality is impossible. Hence $x_\tau \in \operatorname{argmin} f$, and $\operatorname{argmin} f$ is convex. When f is strictly convex and proper, this shows that x_0 and x_1 can't be different; then $\operatorname{argmin} f$ can't contain more than one point.

In a larger picture, if x_0 and x_1 are any points of $\operatorname{dom} f$ with $f(x_0) > f(x_1)$, it's impossible for x_0 to furnish a local minimum of f because every neighborhood of x_0 contains points x_τ with $\tau \in (0,1)$, and such points satisfy $f(x_\tau) \leq (1-\tau)f(x_0) + \tau f(x_1) < f(x_0)$. Thus, there can't be any locally optimal solutions outside of $\operatorname{argmin} f$, where global optimality reigns. □

The uniqueness criterion provided by Theorem 2.6 for optimal solutions is *virtually the only one that can be checked in advance*, without somehow going through a process that would individually determine all the optimal solutions to a problem, if any.

B. Level Sets and Intersections

For any optimization problem of convex type in the sense of Theorem 2.6 the set of feasible solutions is convex, as seen from 2.3. This set may arise from inequality constraints, and here the convexity of constraint functions is vital.

2.7 Proposition (convexity of level sets)**.** *For a convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ all the level sets of type $\operatorname{lev}_{\leq\alpha} f$ and $\operatorname{lev}_{<\alpha} f$ are convex.*

Proof. This follows right from the convexity inequality 2(3). □

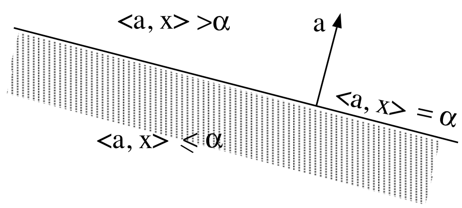

Fig. 2–3. A hyperplane and its associated half-spaces.

Level sets of the type $\operatorname{lev}_{\geq\alpha} f$ and $\operatorname{lev}_{>\alpha} f$ are convex when, instead, f is concave. Those of the type $\operatorname{lev}_{=\alpha} f$ are convex when f is both convex and concave at the same time, and in this respect the following class of functions is important. Here we denote by $\langle x, y\rangle$ the canonical inner product in $\mathbb{R}^n$:

$$\langle x, y\rangle = x_1 y_1 + \cdots + x_n y_n \text{ for } x = (x_1, \ldots, x_n),\ y = (y_1, \ldots, y_n).$$

2.8 Example (affine functions, half-spaces and hyperplanes)**.** *A function f on $\mathbb{R}^n$ is said to be affine if it differs from a linear function by only a constant:*

$$f(x) = \langle a, x\rangle + \beta \text{ for some } a \in \mathbb{R}^n \text{ and } \beta \in \mathbb{R}.$$

Any affine function is both convex and concave. As level sets of affine functions, all sets of the form $\{x \mid \langle a, x\rangle \leq \alpha\}$ and $\{x \mid \langle a, x\rangle \geq \alpha\}$, as well as all those of the form $\{x \mid \langle a, x\rangle < \alpha\}$ and $\{x \mid \langle a, x\rangle > \alpha\}$, are convex in $\mathbb{R}^n$, and so too are all those of the form $\{x \mid \langle a, x\rangle = \alpha\}$. For $a \neq 0$ and α finite, the sets in *the last category are the hyperplanes in $I\!R^n$, while the others are the closed half-spaces and open half-spaces associated with such hyperplanes.*

Affine functions are the only *finite* functions on $I\!R^n$ that are both convex and concave, but other functions with infinite values can have this property, not just the constant functions ∞ and $-\infty$ but examples such as 2(6).

A set defined by several equations or inequalities is the intersection of the sets defined by the individual equations or inequalities, so it's useful to know that convexity of sets is preserved under taking intersections.

2.9 Proposition (intersection, pointwise supremum and pointwise limits).

(a) $\bigcap_{i\in I} C_i$ *is convex if each set* C_i *is convex.*

(b) $\sup_{i\in I} f_i$ *is convex if each function* f_i *is convex.*

(c) $\sup_{i\in I} f_i$ *is strictly convex if each* f_i *is strictly convex and* I *is finite.*

(d) f *is convex if* $f(x) = \limsup_\nu f^\nu(x)$ *for all* x *and each* f^ν *is convex.*

Proof. These assertions follow at once from Definition 2.1. Note that (b) is the epigraphical counterpart to (a): taking the pointwise supremum of a collection of functions corresponds to taking the intersection of their epigraphs. □

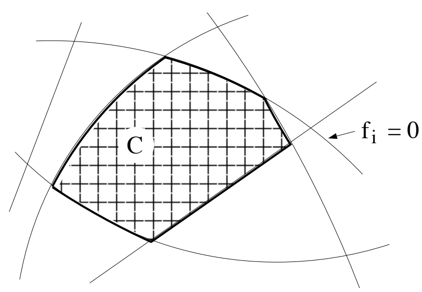

Fig. 2–4. A feasible set defined by convex inequalities.

As an illustration of the intersection principle in 2.9(a), any set $C \subset I\!R^n$ consisting as in Example 1.1 of the points satisfying a constraint system

$$x \in X \;\text{ and }\; \begin{cases} f_i(x) \le 0 & \text{for } i \in I_1, \\ f_i(x) = 0 & \text{for } i \in I_2, \end{cases}$$

is convex if the set $X \subset I\!R^n$ is convex and the functions f_i are convex for $i \in I_1$ but affine for $i \in I_2$. Such sets are common in convex optimization.

2.10 Example (polyhedral sets and affine sets). *A set $C \subset I\!R^n$ is said to be a polyhedral set if it can be expressed as the intersection of a finite family of closed half-spaces or hyperplanes, or equivalently, can be specified by finitely many linear constraints, i.e., constraints $f_i(x) \le 0$ or $f_i(x) = 0$ where f_i is affine. It is called an affine set if it can be expressed as the intersection of hyperplanes alone, i.e., in terms only of constraints $f_i(x) = 0$ with f_i affine.*

Affine sets are in particular polyhedral, while polyhedral sets are in particular closed, convex sets. The empty set and the whole space are affine.

Detail. The empty set is the intersection of two parallel hyperplanes, whereas $\mathbb{R}^n$ is the intersection of the 'empty collection' of hyperplanes in $\mathbb{R}^n$. Thus, the empty set and the whole space are affine sets, hence polyhedral. Note that since every hyperplane is the intersection of two opposing closed half-spaces, hyperplanes are superfluous in the definition of a polyhedral set.

The alternative descriptions of polyhedral and affine sets in terms of linear constraints are based on 2.8. For an affine function f_i that happens to be a constant function, a constraint $f_i(x) \leq 0$ gives either the empty set or the whole space, and similarly for a constraint $f_i(x) = 0$. Such possible degeneracy in a system of linear constraints therefore doesn't affect the geometric description of the set of points satisfying the system as being polyhedral or affine. $\square$

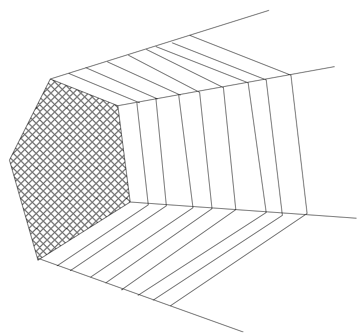

Fig. 2–5. A polyhedral set.

2.11 Exercise (characterization of affine sets)**.** *For a nonempty set $C \subset \mathbb{R}^n$ the following properties are equivalent:*

(a) *C is an affine set;*

(b) *C is a translate $M + p$ of a linear subspace M of $\mathbb{R}^n$ by a vector p;*

(c) *C has the form $\{x \mid Ax = b\}$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$.*

(d) *C contains for each pair of distinct points the entire line through them: if $x_0 \in C$ and $x_1 \in C$ then $(1 - \tau)x_0 + \tau x_1 \in C$ for all $\tau \in (-\infty, \infty)$.*

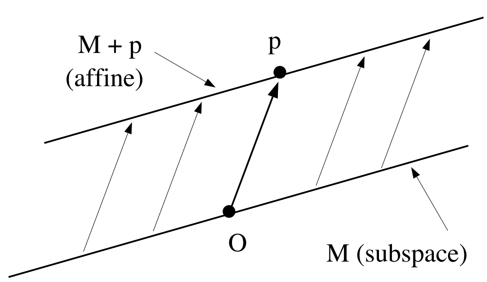

Fig. 2–6. Affine sets as translates of subspaces.

Guide. Argue that an affine set containing 0 must be a linear subspace. A subspace M of dimension $n-m$ can be represented as the set of vectors orthogonal to certain vectors $a_1,\ldots,a_m$. Likewise reduce the analysis of (d) to the case where $0 \in C$. □

Taking the pointwise supremum of a family of functions is just one of many convexity-preserving operations. Others will be described shortly, but we first expand the criteria for verifying convexity directly. For differentiable functions, conditions on first and second derivatives serve this purpose.

C. Derivative Tests

We begin the study of such conditions with functions of a single real variable. This approach is expedient because convexity is essentially a one-dimensional property; behavior with respect to line segments is all that counts. For instance, a set C is convex if and only if its intersection with every line is convex. By the same token, *a function f is convex if and only if it's convex relative to every line*. Many of the properties of convex functions on $I\!R^n$ can thus be derived from an analysis of the simpler case where $n=1$.

2.12 Lemma (slope inequality)**.** *A real-valued function f on an interval $C \subset I\!R$ is convex on C if and only if for arbitrary points $x_0 < y < x_1$ in C one has*

$$\frac{f(y)-f(x_0)}{y-x_0} \le \frac{f(x_1)-f(x_0)}{x_1-x_0} \le \frac{f(x_1)-f(y)}{x_1-y}. \qquad 2(7)$$

Then for any $x \in C$ the difference quotient $\Delta_x(y) := \big[f(y)-f(x)\big]/(y-x)$ is a nondecreasing function of $y \in C \setminus \{x\}$, i.e., one has $\Delta_x(y_0) \le \Delta_x(y_1)$ for all choices of y_0 and y_1 not equal to x with $y_0 < y_1$.

Similarly, strict convexity is characterized by strict inequalities between the difference quotients, and then $\Delta_x(y)$ is an increasing function of $y \in C \setminus \{x\}$.

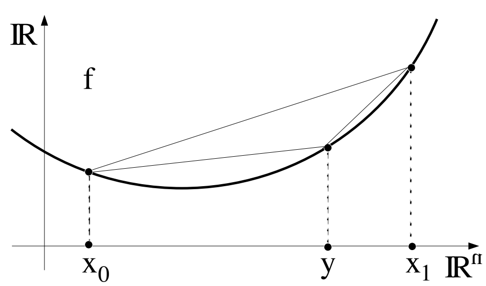

Fig. 2–7. Slope inequality.

Proof. The convexity of f is equivalent to the condition that

$$f(y) \;\le\; \frac{x_1-y}{x_1-x_0}f(x_0) + \frac{y-x_0}{x_1-x_0}f(x_1) \;\text{ when }\; x_0 < y < x_1 \;\text{ in }\; C, \qquad 2(8)$$

since this is 2(3) when y is x_τ for $\tau = (y - x_0)/(x_1 - x_0)$. The first inequality is what one gets by subtracting $f(x_0)$ from both sides in 2(8), whereas the second corresponds to subtracting $f(x_1)$. The case of strict convexity is parallel. □

2.13 Theorem (one-dimensional derivative tests)**.** *For a differentiable function f on an open interval $O \subset I\!R$, each of the following conditions is both necessary and sufficient for f to be convex on O:*

(a) *f' is nondecreasing on O, i.e., $f'(x_0) \le f'(x_1)$ when $x_0 < x_1$ in O;*
(b) *$f(y) \ge f(x) + f'(x)(y - x)$ for all x and y in O;*
(c) *$f''(x) \ge 0$ for all x in O (assuming twice differentiability).*

Similarly each of the following conditions is both necessary and sufficient for f to be strictly convex on O:

(a$'$) *f' is increasing on O: $f'(x_0) < f'(x_1)$ when $x_0 < x_1$ in O.*
(b$'$) *$f(y) > f(x) + f'(x)(y - x)$ for all x and y in O with $y \ne x$.*

A sufficient (but not necessary) condition for strict convexity is:

(c$'$) *$f''(x) > 0$ for all x in O (assuming twice differentiability).*

Proof. The equivalence between (a) and (c) when f is twice differentiable is well known from elementary calculus, and the same is true of the implication from (c$'$) to (a$'$). We'll show now that [convexity] $\Rightarrow$ (a) $\Rightarrow$ (b) $\Rightarrow$ [convexity]. If f is convex, we have

$$f'(x_0) \le \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{f(x_0) - f(x_1)}{x_0 - x_1} \le f'(x_1) \ \text{ when } \ x_0 < x_1 \ \text{ in } \ O$$

from the monotonicity of difference quotients in Lemma 2.12, and this gives (a). On the other hand, if (a) holds we have for any $y \in O$ that the function $g_y(x) := f(x) - f(y) - f'(y)(x - y)$ has $g_y'(x) \ge 0$ for all $x \in (y, \infty) \cap O$ but $g_y'(x) \le 0$ for all $x \in (-\infty, y) \cap O$. Then g_y is nondecreasing to the right of y but nonincreasing to the left, and it therefore attains its global minimum over O at y. This means that (b) holds. Starting now from (b), consider the family of affine functions $l_y(x) = f(y) + f'(y)(x - y)$ indexed by $y \in O$. We have $f(x) = \max_{y \in O} l_y(x)$ for all $x \in O$, so f is convex on O by 2.9(b).

In parallel fashion we see that [strict convexity] $\Rightarrow$ (a$'$) $\Rightarrow$ (b$'$). To establish that (b$'$) implies not just convexity but strict convexity, consider $x_0 < x_1$ in O and an intermediate point x_τ as in 2(1). For the affine function $l(x) = f(x_\tau) + f'(x_\tau)(x - x_\tau)$ we have $f(x_0) > l(x_0)$ and $f(x_1) > l(x_1)$, but $f(x_\tau) = l(x_\tau) = (1 - \tau)l(x_0) + \tau l(x_1)$. Therefore, $f(x_\tau) < (1 - \tau)f(x_0) + \tau f(x_1)$. □

Here are some examples of functions of a single variable whose convexity or strict convexity can be established by the criteria in Theorem 2.13, in particular by the second-derivative tests:

- $f(x) = ax^2 + bx + c$ on $(-\infty, \infty)$ when $a \ge 0$; strictly convex when $a > 0$.
- $f(x) = e^{ax}$ on $(-\infty, \infty)$; strictly convex when $a \ne 0$.
- $f(x) = x^r$ on $(0, \infty)$ when $r \ge 1$; strictly convex when $r > 1$.
- $f(x) = -x^r$ on $(0, \infty)$ when $0 \le r \le 1$; strictly convex when $0 < r < 1$.

- $f(x) = x^{-r}$ on $(0, \infty)$ when $r > 0$; strictly convex in fact on this interval.
- $f(x) = -\log x$ on $(0, \infty)$; strictly convex in fact on this interval.

The case of $f(x) = x^4$ on $(-\infty, \infty)$ furnishes a counterexample to the common misconception that positivity of the second derivative in 2.13(c′) is not only sufficient but necessary for strict convexity: this function is strictly convex relative to the entire real line despite having $f''(0) = 0$. The strict convexity can be verified by applying the condition to f on the intervals $(-\infty, 0)$ and $(0, \infty)$ separately and then invoking the continuity of f at 0. In general, it can be seen that 2.13(c′) remains a sufficient condition for the strict convexity of a twice differentiable function when relaxed to allow $f''(x)$ to vanish at finitely many points x (or even on a subset of O having Lebesgue measure zero).

To extend the criteria in 2.13 to functions of $x = (x_1, \ldots, x_n)$, we must call for appropriate conditions on gradient vectors and Hessian matrices. For a differentiable function f, the *gradient* vector and *Hessian* matrix at x are

$$\nabla f(x) := \left[\frac{\partial f}{\partial x_j}(x)\right]_{j=1}^{n}, \qquad \nabla^2 f(x) := \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right]_{i,j=1}^{n,n}.$$

Recall that a matrix $A \in I\!R^{n \times n}$ is called *positive-semidefinite* if $\langle z, Az\rangle \geq 0$ for all z, and *positive-definite* if $\langle z, Az\rangle > 0$ for all $z \neq 0$. This terminology applies even if A is not symmetric, but of course $\langle z, Az\rangle$ depends only on the *symmetric part of* A, i.e., the matrix $\frac{1}{2}(A + A^*)$, where A^* denotes the transpose of A. In terms of the components a_{ij} of A and z_j of z, one has

$$\langle z, Az\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} z_i z_j.$$

2.14 Theorem (higher-dimensional derivative tests)**.** *For a differentiable function f on an open convex set $O \subset I\!R^n$, each of the following conditions is both necessary and sufficient for f to be convex on O:*

(a) $\langle x_1 - x_0, \nabla f(x_1) - \nabla f(x_0)\rangle \geq 0$ *for all x_0 and x_1 in O;*
(b) $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle$ *for all x and y in O;*
(c) $\nabla^2 f(x)$ *is positive-semidefinite for all x in O (f twice differentiable).*

For strict convexity, a necessary and sufficient condition is (a) holding with strict inequality when $x_0 \neq x_1$, or (b) holding with strict inequality when $x \neq y$. A condition that is sufficient for strict convexity (but not necessary) is the positive definiteness of the Hessian matrix in (c) for all x in O.

Proof. As already noted, f is convex on O if and only if it is convex on every line segment in O. This is equivalent to the property that for every choice of $y \in O$ and $z \in I\!R^n$ the function $g(t) = f(y + tz)$ is convex on any open interval of t values for which $y + tz \in O$. Here $g'(t) = \big\langle z, \nabla f(y + tz)\big\rangle$ and $g''(t) = \big\langle z, \nabla^2 f(y + tz)z\big\rangle$. The asserted conditions for convexity and strict convexity are equivalent to requiring in each case that the corresponding condition in 2.13 hold for all such functions g. □

Twice differentiable concave or strictly concave functions can similarly be characterized in terms of Hessian matrices that are negative-semidefinite or negative-definite.

2.15 Example (quadratic functions). *A function f on $I\!R^n$ is quadratic if it's expressible as $f(x) = \frac{1}{2}\langle x, Ax\rangle + \langle a, x\rangle + \text{const.}$, where the matrix $A \in I\!R^{n\times n}$ is symmetric. Then $\nabla f(x) = Ax + a$ and $\nabla^2 f(x) \equiv A$, so f is convex if and only if A is positive-semidefinite. Moreover, a function f of this type is strictly convex if and only if A is positive-definite.*

Detail. Note that the positive definiteness of the Hessian is being asserted as necessary for the strict convexity of a quadratic function, even though it was only listed in Theorem 2.14 merely as sufficient in general. The reason is that if A is positive-semidefinite, but not positive-definite, there's a vector $z \neq 0$ such that $\langle z, Az\rangle = 0$, and then along the line through the origin in the direction of z it's impossible for f to be strictly convex. □

Because level sets of convex functions are convex by 2.7, it follows from Example 2.15 that every set of the form

$$C = \left\{x \,\middle|\, \tfrac{1}{2}\langle x, Ax\rangle + \langle a, x\rangle \leq \alpha\right\} \text{ with } A \text{ positive-semidefinite}$$

is convex. This class of sets includes all closed Euclidean balls as well as general ellipsoids, paraboloids, and 'cylinders' with ellipsoidal or paraboloidal base.

Algebraic functions of the form described in 2.15 are often called *quadratic*, but that name seems to carry the risk of suggesting sometimes that $A \neq 0$ is assumed, or that second-order terms—only—are allowed. Sometimes, we'll refer to them as *linear-quadratic* as a way of emphasizing the full generality.

2.16 Example (convexity of vector-max and log-exponential). *In terms of $x = (x_1, \ldots, x_n)$, the functions*

$$\text{vecmax}(x) := \max\{x_1, \ldots, x_n\}, \qquad \text{logexp}(x) := \log\big(e^{x_1} + \cdots + e^{x_n}\big),$$

are convex on $I\!R^n$ but not strictly convex.

Detail. Convexity of $f = \text{logexp}$ is established via 2.14(c) by calculating in terms of $\sigma(x) = \sum_{j=1}^n e^{x_j}$ that

$$\begin{aligned}\big\langle z, \nabla^2 f(x) z\big\rangle &= \frac{1}{\sigma(x)} \sum_{j=1}^n e^{x_j} z_j^2 - \frac{1}{\sigma(x)^2} \sum_{j=1}^n \sum_{i=1}^n e^{(x_i + x_j)} z_i z_j \\ &= \frac{1}{2\sigma(x)^2} \sum_{i=1}^n \sum_{j=1}^n e^{(x_i + x_j)} (z_i - z_j)^2 \geq 0.\end{aligned}$$

Strict convexity fails because $f(x + t1\!\!1) = f(x) + t$ for $1\!\!1 = (1, 1, \ldots, 1)$. As $f = \text{vecmax}$ is the pointwise max of the n linear functions $x \mapsto x_j$, it's convex by 2.9(c). It isn't strictly convex, because $f(\lambda x) = \lambda f(x)$ for $\lambda \geq 0$. □

2.17 Example (norms). *By definition, a norm on* $\mathbb{R}^n$ *is a real-valued function* $h(x) = \|x\|$ *such that*

$$\|\lambda x\| = |\lambda|\|x\|, \quad \|x + y\| \leq \|x\| + \|y\|, \quad \|x\| > 0 \ \text{ for } \ x \neq 0.$$

Any such a function h *is convex, but not strictly convex. The corresponding balls* $\{x \mid \|x - x_0\| \leq \rho\}$ *and* $\{x \mid \|x - x_0\| < \rho\}$ *are convex sets. Beyond the Euclidean norm* $|x| = (\sum_{j=1}^n |x_j|^2)^{1/2}$ *these properties hold for the* l_p *norms*

$$\|x\|_p := \Big(\sum\nolimits_{j=1}^n |x_j|^p\Big)^{1/p} \ \text{ for } \ 1 \leq p < \infty, \qquad \|x\|_\infty := \max_{j=1,\ldots,n} |x_j|. \qquad 2(9)$$

Detail. Any norm satisfies the convexity inequality 2(3), but the strict version fails when $x_0 = 0$, $x_1 \neq 0$. The associated balls are convex as translates of level sets of a convex function. For any $p \in [1, \infty]$ the function $h(x) = \|x\|_p$ obviously fulfills the first and third conditions for a norm. In light of the first condition, the second can be written as $h\big(\frac{1}{2}x + \frac{1}{2}y\big) \leq \frac{1}{2}h(x) + \frac{1}{2}h(y)$, which will follow from verifying that h is convex, or equivalently that $\operatorname{epi} h$ is convex. We have $\operatorname{epi} h = \{\lambda(x, 1) \mid x \in B,\ \lambda \geq 0\}$, where $B = \{x \mid \|x\|_p \leq 1\}$. The convexity of $\operatorname{epi} h$ can easily be derived from this formula once it is known that the set B is convex. For $p = \infty$, B is a box, while for $p \in [1, \infty)$ it is $\operatorname{lev}_{\leq 1} g$ for the convex function $g(x) = \sum_{j=1}^n |x_j|^p$, hence it is convex in that case too. □

D. Convexity in Operations

Many important convex functions lack differentiability. Norms can't ever be differentiable at the origin. The vector-max function in 2.16 fails to be differentiable at $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n)$ if two coordinates $\bar{x}_j$ and $\bar{x}_k$ tie for the max. (At such a point the function has 'kinks' along the lines parallel to the x_j-axis and the x_k axis.) Although derivative tests can't be used to establish the convexity of functions like these, other criteria can fill the need, for instance the fact in 2.9 that a pointwise supremum of convex functions is convex. (A pointwise infimum of convex functions is convex only in cases like a decreasing sequence of convex functions or a 'convexly parameterized' family as will be treated in 2.22(a).)

2.18 Exercise (addition and scalar multiplication). *For convex functions* $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ *and real coefficients* $\lambda_i \geq 0$, *the function* $\sum_{i=1}^m \lambda_i f_i$ *is convex. It is strictly convex if for at least one index* i *with* $\lambda_i > 0$, f_i *is strictly convex.*

Guide. Work from Definition 2.1. □

2.19 Exercise (set products and separable functions).

(a) *If* $C = C_1 \times \cdots \times C_m$ *where each* C_i *is convex in* $\mathbb{R}^{n_i}$, *then* C *is convex in* $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_m}$. *In particular, any box in* $\mathbb{R}^n$ *is a closed, convex set.*

(b) *If* $f(x) = f_1(x_1)+\cdots+f_m(x_m)$ *for* $x = (x_1, \ldots, x_m)$ *in* $I\!R^{n_1}\times\cdots\times I\!R^{n_m}$, *where each* f_i *is convex, then* f *is convex. If each* f_i *is strictly convex, then* f *is strictly convex.*

2.20 Exercise (convexity in composition)**.**

(a) *If* $f(x) = g(Ax+a)$ *for a convex function* $g : I\!R^m \to \overline{I\!R}$ *and some choice of* $A \in I\!R^{m\times n}$ *and* $a \in I\!R^m$*, then* f *is convex.*

(b) *If* $f(x) = \theta\big(g(x)\big)$ *for a convex function* $g : I\!R^n \to \overline{I\!R}$ *and a nondecreasing convex function* $\theta : I\!R \to \overline{I\!R}$*, the convention being used that* $\theta(\infty) = \infty$ *and* $\theta(-\infty) = \inf\theta$*, then* f *is convex. Furthermore,* f *is strictly convex in the case where* g *is strictly convex and* θ *is increasing.*

(c) *Suppose* $f(x) = g\big(F(x)\big)$ *for a convex function* $g : I\!R^m \to \overline{I\!R}$ *and a mapping* $F : I\!R^n \to I\!R^m$*, where* $F(x) = \big(f_1(x), \ldots, f_m(x)\big)$ *with* f_i *convex for* $i = 1, \ldots, s$ *and* f_i *affine for* $i = s+1, \ldots, m$*. Suppose that* $g(u_1, \ldots, u_m)$ *is nondecreasing in* u_i *for* $i = 1, \ldots, s$*. Then* f *is convex.*

An example of a function whose convexity follows from 2.20(a) is $f(x) = \|Ax - b\|$ for any matrix A, vector b, and any norm $\|\cdot\|$. Some elementary examples of convex and strictly convex functions constructed as in 2.20(b) are:

- $f(x) = e^{g(x)}$ is convex when g is convex, and strictly convex when g is strictly convex.
- $f(x) = -\log|g(x)|$ when $g(x) < 0$, $f(x) = \infty$ when $g(x) \geq 0$, is convex when g is convex, and strictly convex when g is strictly convex.
- $f(x) = g(x)^2$ when $g(x) \geq 0$, $f(x) = 0$ when $g(x) < 0$, is convex when g is convex.

As an example of the higher-dimensional composition in 2.20(c), a vector of convex functions $f_1, \ldots, f_m$ can be composed with vecmax or logexp, cf. 2.16. This confirms that convexity is preserved when a nonsmooth function, given as the pointwise maximum of finitely many smooth functions, is approximated by a smooth function as in Example 1.30. Other examples in the mode of 2.20(c) are obtained with $g(z)$ of the form

$$\big|z\big|_+ := \max\{z, 0\} \;\text{ for }\; z \in I\!R$$

in 2.20(b), or more generally

$$\big|z\big|_+ := \sqrt{\big|z_1\big|_+^2 + \cdots + \big|z_m\big|_+^2} \;\text{ for }\; z = (z_1, \ldots, z_m) \in I\!R^m \qquad 2(10)$$

in 2.20(c). This way we get the convexity of a function of the form $f(x) = \big|\big(f_1(x), \ldots, f_m(x)\big)\big|_+$ when f_i is convex. Then too, for instance, the pth power of such a function is convex for any $p \in [1, \infty)$, as seen from further composition with the nondecreasing, convex function $\theta : t \mapsto \big|t\big|_+^p$.

2.21 Proposition (images under linear mappings)**.** *If* $L : I\!R^n \to I\!R^m$ *is linear, then* $L(C)$ *is convex in* $I\!R^m$ *for every convex set* $C \subset I\!R^n$*, while* $L^{-1}(D)$ *is convex in* $I\!R^n$ *for every convex set* $D \subset I\!R^m$*.*

Proof. The first assertion is obtained from the definition of the convexity of C and the linearity of L. Specifically, if $u = (1-\tau)u_0 + \tau u_1$ for points $u_0 = L(x_0)$ and $u_1 = L(x_1)$ with x_0 and x_1 in C, then $u = L(x)$ for the point $x = (1-\tau)x_0 + \tau x_1$ in C. The second assertion can be proved quite as easily, but it may also be identified as a specialization of the composition rule in 2.20(a) to the case of g being the indicator δ_C. □

Projections onto subspaces are examples of mappings L to which 2.21 can be applied. If D is a subset of $\mathbb{R}^n \times \mathbb{R}^d$ and C consists of the vectors $x \in \mathbb{R}^n$ for which there's a vector $w \in \mathbb{R}^d$ with $(x,w) \in D$, then C is the image of D under the mapping $(x,w) \mapsto x$, so it follows that C is convex when D is convex.

2.22 Proposition (convexity in inf-projection and epi-composition).

(a) *If $p(u) = \inf_x f(x,u)$ for a convex function f on $\mathbb{R}^n \times \mathbb{R}^m$, then p is convex on $\mathbb{R}^m$. Also, the set $P(u) = \operatorname{argmin}_x f(x,u)$ is convex for each u.*

(b) *For any convex function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and matrix $A \in \mathbb{R}^{m\times n}$ the function $Af : \mathbb{R}^m \to \overline{\mathbb{R}}$ defined by $(Af)(u) := \inf\{f(x) \mid Ax = u\}$ is convex.*

Proof. In (a), the set $D = \{(u,\alpha) \mid p(u) < \alpha < \infty\}$ is the image of the set $E = \{(x,u,\alpha) \mid f(x,u) < \alpha < \infty\}$ under the linear mapping $(x,u,\alpha) \mapsto (u,\alpha)$. The convexity of E through 2.4 implies that of D through 2.21, and this ensures that p is convex. The convexity of $P(u)$ follows from 2.6.

Likewise in (b), the set $D' = \{(u,\alpha) \mid (Af)(u) < \alpha < \infty\}$ is the image of $E' = \{(x,\alpha) \mid f(x) < \alpha < \infty\}$ under the linear transformation $(x,\alpha) \mapsto (Ax,\alpha)$. Again, the claimed convexity is justified through 2.4 and 2.21. □

The notation Af in 2.22(b) is interchangeable with Lf for the mapping $L : x \mapsto Ax$, which fits with the general definition of epi-composition in 1(17).

2.23 Proposition (convexity in set algebra).

(a) *$C_1 + C_2$ is convex when C_1 and C_2 are convex.*

(b) *λC is convex for every $\lambda \in \mathbb{R}$ when C is convex.*

(c) *When C is convex, one has $(\lambda_1 + \lambda_2)C = \lambda_1 C + \lambda_2 C$ for $\lambda_1, \lambda_2 \geq 0$.*

Proof. Properties (a) and (b) are immediate from the definitions. In (c) it suffices through rescaling to deal with the case where $\lambda_1 + \lambda_2 = 1$. The equation is then merely a restatement of the definition of the convexity of C. □

As an application of 2.23, the 'fattened' set $C + \varepsilon\mathbb{B}$ is convex whenever C is convex; here $\mathbb{B}$ is the closed, unit Euclidean ball, cf. 1(15).

2.24 Exercise (epi-sums and epi-multiples).

(a) *$f_1 \# f_2$ is convex when f_1 and f_2 are convex.*

(b) *$\lambda\star f$ is convex for every $\lambda > 0$ when f is convex.*

(c) *When f is convex, one has $(\lambda_1 + \lambda_2)\star f = (\lambda_1\star f) \# (\lambda_2\star f)$ for $\lambda_1, \lambda_2 \geq 0$.*

Guide. Get these conclusions from 2.23 through the geometry in 1.28. □

2.25 Example (distance functions and projections). *For a convex set C, the distance function d_C is convex, and as long as C is closed and nonempty, the* *projection mapping P_C is single-valued, i.e., for each point $x \in \mathbb{R}^n$ there is a unique point of C nearest to x. Moreover P_C is continuous.*

Detail. The distance function (from 1.20) arises through epi-addition, cf. 1(13), so its convexity is a consequence of 2.24. The elements of $P_C(x)$ minimize $|w-x|$ over all $w \in C$, but they can also be regarded as minimizing $|w-x|^2$ over all $w \in C$. Since the function $w \mapsto |w-x|^2$ is strictly convex (via 2.15), there can't be more than one such element (see 2.6), but on the other hand there's at least one (by 1.20). Hence $P_C(x)$ is a singleton. The cluster point property in 1.20 ensures then that, as a single-valued mapping, P_C is continuous. □

For a convex function f, the Moreau envelopes and proximal mappings defined in 1.22 have remarkable properties.

2.26 Theorem (proximal mappings and envelopes under convexity)**.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be lsc, proper, and convex. Then f is prox-bounded with threshold ∞, and the following properties hold for every $\lambda > 0$.*

(a) *The proximal mapping $P_\lambda f$ is single-valued and continuous. In fact $P_\lambda f(x) \to P_{\bar\lambda} f(\bar x)$ whenever $(\lambda, x) \to (\bar\lambda, \bar x)$ with $\bar\lambda > 0$.*

(b) *The envelope function $e_\lambda f$ is convex and continuously differentiable, the gradient being*

$$\nabla e_\lambda f(x) = \frac{1}{\lambda}\big[x - P_\lambda f(x)\big].$$

Proof. From Definition 1.22 we have $e_\lambda f(x) := \inf_w g_\lambda(x,w)$ and $P_\lambda f(x) := \operatorname{argmin}_w g_\lambda(x,w)$ for the function $g_\lambda(x,w) := f(w) + (1/2\lambda)|w-x|^2$, which our assumptions imply to be lsc, proper and convex in (x,w), even strictly convex in w. If the threshold for f is ∞ as claimed, then $e_\lambda f$ and $P_\lambda f$ have all the properties in 1.25, and in addition $e_\lambda f$ is convex by 2.22(a), while $P_\lambda f$ is single-valued by 2.6. This gives everything in (a) and (b) except for the differentiability in (b), which will need a supplementary argument. Before proceeding with that argument, we verify the prox-boundedness.

In order to show that the threshold for f is ∞, it suffices to show for arbitrary $\lambda > 0$ that $e_\lambda f(0) > -\infty$, which can be accomplished through Theorem 1.9 by demonstrating the boundedness of the level sets of $g_\lambda(0,\cdot)$. If the latter property were absent, there would exist $\alpha \in \mathbb{R}$ and points x^ν with $f(x^\nu) + (1/2\lambda)|x^\nu|^2 \le \alpha$ such that $1 < |x^\nu| \to \infty$. Fix any x_0 with $f(x_0) < \infty$. Then in terms of $\tau^\nu = 1/|x^\nu| \in (0,1)$ and $\bar x^\nu := (1-\tau^\nu)x_0 + \tau^\nu x^\nu$ we have $\tau^\nu \to 0$ and

$$\begin{aligned} f(\bar x^\nu) &\le (1-\tau^\nu) f(x_0) + \tau^\nu f(x^\nu) \\ &\le (1-\tau^\nu) f(x_0) + \tau^\nu \alpha - (1/2\lambda)|x^\nu| \to -\infty. \end{aligned}$$

The sequence of points $\bar x^\nu$ is bounded, so this is incompatible with f being proper and lsc (cf. 1.10). The contradiction proves the claim.

The differentiability claim in (b) is next. Continuous differentiability will follow from the formula for the gradient mapping, once that is established, since $P_\lambda f$ is already known to be a continuous, single-valued mapping. Consider any point $\bar{x}$, and let $\bar{w} = P_\lambda f(\bar{x})$ and $\bar{v} = (\bar{x} - \bar{w})/\lambda$. Our task is to show that $e_\lambda f$ is differentiable at $\bar{x}$ with $\nabla e_\lambda f(\bar{x}) = \bar{v}$, or equivalently in terms of $h(u) := e_\lambda f(\bar{x}+u) - e_\lambda f(\bar{x}) - \langle \bar{v}, u\rangle$ that h is differentiable at 0 with $\nabla h(0) = 0$. We have $e_\lambda f(\bar{x}) = f(\bar{w}) + (1/2\lambda)|\bar{w} - \bar{x}|^2$, whereas $e_\lambda f(\bar{x} + u) \le f(\bar{w}) + (1/2\lambda)|\bar{w} - (\bar{x} + u)|^2$, so that

$$h(u) \;\le\; \frac{1}{2\lambda}|\bar{w} - (\bar{x} + u)|^2 - \frac{1}{2\lambda}|\bar{w} - \bar{x}|^2 - \frac{1}{\lambda}\langle \bar{x} - \bar{w}, u\rangle \;=\; \frac{1}{2\lambda}|u|^2.$$

But h inherits the convexity of $e_\lambda f$ and therefore has $\frac{1}{2}h(u)+\frac{1}{2}h(-u) \ge h\big(\frac{1}{2}u + \frac{1}{2}(-u)\big) = h(0) = 0$, so from the inequality just obtained we also get

$$h(u) \ge -h(-u) \ge -\frac{1}{2\lambda}|-u|^2 = -\frac{1}{2\lambda}|u|^2.$$

Thus we have $\big|h(u)\big| \le (1/2\lambda)|u|^2$ for all u, and this obviously yields the desired differentiability property. □

Especially interesting in Theorem 2.26 is the fact that regardless of any nonsmoothness of f, the envelope approximations $e_\lambda f$ are always smooth.

E. Convex Hulls

Nonconvex sets can be 'convexified'. For $C \subset I\!R^n$, the *convex hull* of C, denoted by $\operatorname{con} C$, is the smallest convex set that includes C. Obviously $\operatorname{con} C$ is the intersection of all the convex sets $D \supset C$, this intersection being a convex set by 2.9(a). (At least one such set D always exists—the whole space.)

2.27 Theorem (convex hulls from convex combinations)**.** *For a set $C \subset I\!R^n$, $\operatorname{con} C$ consists of all the convex combinations of elements of C:*

$$\operatorname{con} C \;=\; \Big\{\sum\nolimits_{i=0}^{p} \lambda_i x_i \,\Big|\, x_i \in C,\ \lambda_i \ge 0,\ \sum\nolimits_{i=0}^{p} \lambda_i = 1,\ p \ge 0\Big\}.$$

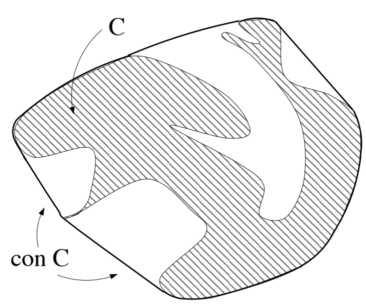

Fig. 2–8. The convex hull of a set.

Proof. Let D be the set of all convex combinations of elements of C. We have $D \subset \operatorname{con} C$ by 2.2, because $\operatorname{con} C$ is a convex set that includes C. But

D is itself convex: if x is a convex combination of $x_0, \ldots, x_p$ in C and x' is a convex combination of $x'_0, \ldots, x'_{p'}$ in C, then for any $\tau \in (0,1)$, the vector $(1-\tau)x + \tau x'$ is a convex combination of $x_0, \ldots, x_p$ and $x'_0, \ldots, x'_{p'}$ together. Hence $\operatorname{con} C = D$. ☐

Sets expressible as the convex hull of a *finite* subset of $I\!R^n$ are especially important. When $C = \{a_0, a_1, \ldots, a_p\}$, the formula in 2.27 simplifies: $\operatorname{con} C$ consists of all convex combinations $\lambda_0 a_0 + \lambda_1 a_1 + \cdots + \lambda_p a_p$. If the points $a_0, a_1, \ldots, a_p$ are *affinely independent*, the set $\operatorname{con}\{a_0, a_1, \ldots, a_p\}$ is called a *p-simplex* with these points as its *vertices*. The affine independence condition means that the only choice of coefficients $\lambda_i \in (-\infty, \infty)$ with $\lambda_0 a_0 + \lambda_1 a_1 + \cdots + \lambda_p a_p = 0$ and $\lambda_0 + \lambda_1 + \cdots + \lambda_p = 0$ is $\lambda_i = 0$ for all i; this holds if and only if the vectors $a_i - a_0$ for $i = 1, \ldots, p$ are linearly independent. A 0-simplex is a point, whereas a 1-simplex is a closed line segment joining a pair of distinct points. A 2-simplex is a triangle, and a 3-simplex a tetrahedron.

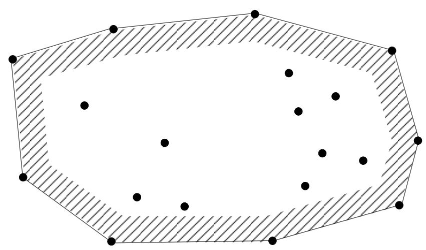

Fig. 2–9. The convex hull of finitely many points.

Simplices are often useful in technical arguments about convexity. The following are some of the facts commonly employed.

2.28 Exercise (simplex technology)**.**

(a) *Every simplex $S = \operatorname{con}\{a_0, a_1, \ldots, a_p\}$ is a polyhedral set, in particular closed and convex.*

(b) *When $a_0, \ldots, a_n$ are affinely independent in $I\!R^n$, every $x \in I\!R^n$ has a unique expression in barycentric coordinates: $x = \sum_{i=0}^n \lambda_i a_i$ with $\sum_{i=0}^n \lambda_i = 1$.*

(c) *The expression of each point of a p-simplex $S = \operatorname{con}\{a_0, a_1, \ldots, a_p\}$ as a convex combination $\sum_{i=0}^p \lambda_i a_i$ is unique: there is a one-to-one correspondence, continuous in both directions, between the points of S and the vectors $(\lambda_0, \lambda_1, \ldots, \lambda_p) \in I\!R^{p+1}$ such that $\lambda_i \geq 0$ and $\sum_{i=0}^p \lambda_i = 1$.*

(d) *Every n-simplex $S = \operatorname{con}\{a_0, a_1, \ldots, a_n\}$ in $I\!R^n$ has nonempty interior, and $x \in \operatorname{int} S$ if and only if $x = \sum_{i=0}^n \lambda_i a_i$ with $\lambda_i > 0$ and $\sum_{i=0}^n \lambda_i = 1$.*

(e) *Simplices can serve as neighborhoods: for every point $\bar{x} \in I\!R^n$ and neighborhood $V \in \mathcal{N}(\bar{x})$ there is an n-simplex $S \subset V$ with $\bar{x} \in \operatorname{int} S$.*

(f) *For an n-simplex $S = \operatorname{con}\{a_0, a_1, \ldots, a_n\}$ in $I\!R^n$ and sequences $a_i^\nu \to a_i$, the set $S^\nu = \operatorname{con}\{a_0^\nu, a_1^\nu, \ldots, a_n^\nu\}$ is an n-simplex once ν is sufficiently large. Furthermore, if $x^\nu \to x$ with barycentric representations $x^\nu = \sum_{i=0}^n \lambda_i^\nu a_i^\nu$ and $x = \sum_{i=0}^n \lambda_i a_i$, where $\sum_{i=0}^n \lambda_i^\nu = 1$ and $\sum_{i=0}^n \lambda_i = 1$, then $\lambda_i^\nu \to \lambda_i$.*

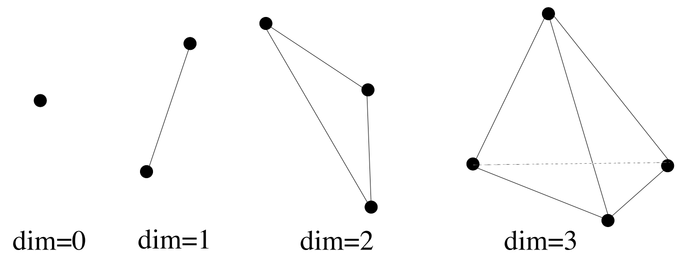

Fig. 2–10. Simplices of dimensions 0, 1, 2 and 3.

Guide. In (a)-(d) simplify by translating the points to make $a_0 = 0$; in (a) and (c) augment $a_1, \ldots, a_p$ by other vectors (if $p < n$) in order to have a basis for $I\!R^n$. Consider the linear transformation that maps a_i to $e_i = (0, \ldots, 1, 0, \ldots, 0)$ (where the 1 in e_i appears in ith position). For (e), the case of $\bar{x} = 0$ and convex neighborhoods $V = I\!B(0, \varepsilon)$ is enough. Taking any n-simplex S with $0 \in \operatorname{int} S$, show there's a $\delta > 0$ such that $\delta S \subset I\!B(0, \varepsilon)$. In (f) let A be the matrix having the vectors $a_i - a_0$ as its columns; similarly, A^ν. Identify the simplex assumption on S with the nonsingularity of A and argue (by way of determinants for instance) that then A^ν must eventually be nonsingular. For the last part, note that $x = Az + a_0$ for $z = (\lambda_1, \ldots, \lambda_n)$ and similarly $x^\nu = A^\nu z^\nu + a_0^\nu$ for ν sufficiently large. Establish that the convergence of A^ν to A entails the convergence of the inverse matrices. □

In the statement of Theorem 2.27, the integer p isn't fixed and varies over all possible choices of a convex combination. This is sometimes inconvenient, and it's valuable then to know that a fixed choice will suffice.

2.29 Theorem (convex hulls from simplices; Carathéodory)**.** *For a set $C \neq \emptyset$ in $I\!R^n$, every point of $\operatorname{con} C$ belongs to some simplex with vertices in C and thus can be expressed as a convex combination of $n+1$ points of C (not necessarily different). For every point in $\operatorname{bdry} C$, the boundary of C, n points suffice. When C is connected, then every point of $\operatorname{con} C$ can be expressend as the combination of no more than n points of C.*

Proof. First we show that $\operatorname{con} C$ is the union of all simplices formed from points of C: each $x \in \operatorname{con} C$ can be expressed not only as a convex combination $\sum_{i=0}^p \lambda_i x_i$ with $x_i \in C$, but as one in which $x_0, x_1, \ldots, x_p$ are affinely independent. For this it suffices to show that when the convex combination is chosen with p minimal (which implies $\lambda_i > 0$ for all i), the points x_i can't be affinely dependent. If they were, we would have coefficients μ_i, at least one of them positive, such that $\sum_{i=0}^p \mu_i x_i = 0$ and $\sum_{i=0}^p \mu_i = 0$. Then there is a largest $\tau > 0$ such that $\tau\mu_i \leq \lambda_i$ for all i, and by setting $\lambda_i' = \lambda_i - \tau\mu_i$ we would get a representation $x = \sum_{i=0}^p \lambda_i' x_i$ in which $\sum_{i=0}^p \lambda_i' = 1$, $\lambda_i' \geq 0$, and actually $\lambda_i' = 0$ for some i. This would contradict the minimality of p.

Of course when $x_0, x_1, \ldots, x_p$ are affinely independent, we have $p \leq n$. If $p < n$ we can choose additional points $x_{p+1}, \ldots, x_n$ arbitrarily from C and expand the representation $x = \sum_{i=0}^p \lambda_i x_i$ to $x = \sum_{i=0}^n \lambda_i x_i$ by taking $\lambda_i = 0$ for $i = p+1, \ldots, n$. This proves the first assertion of the theorem.

By passing to a lower dimensional space if necessary, one can suppose that $\operatorname{con} C$ is n-dimensional. Then $\operatorname{con} C$ is the union of all n-simplices whose vertices are in C. A point in $\operatorname{bdry} C$, while belonging to some such n-simplex, can't be in its interior and therefore must be on boundary of that n-simplex. But the boundary of an n-simplex is a union of $(n-1)$-simplices whose vertices are in C, i.e., every point in $\operatorname{bdry} C$ can be obtained as the convex combination of no more than n point in C.

Finally, let's show that if some point $\bar{x}$ in $\operatorname{con} C$, but not in C, has a minimal representation $\sum_{i=0}^{n} \bar{\lambda}_i x_i$ involving exactly $n+1$ points $x_i \in C$, then C can't be connected. By a translation if necessary, we can simplify to the case where $\bar{x} = 0$; then $\sum_{i=0}^{n} \bar{\lambda}_i x_i = 0$ with $0 < \bar{\lambda}_i < 1$, $\sum_{i=0}^{n} \bar{\lambda}_i = 1$, and $x_0, x_1, \ldots, x_n$ affinely independent. In this situation the vectors $x_1, \ldots, x_n$ are linearly independent, for if not we would have an expression $\sum_{i=1}^{n} \mu_i x_i = 0$ with $\sum_{i=1}^{n} \mu_i = 0$ (but not all coefficients 0), or one with $\sum_{i=1}^{n} \mu_i = 1$; the first case is precluded by $x_0, x_1, \ldots, x_n$ being affinely independent, while the second is impossible by the uniqueness of the coefficients $\bar{\lambda}_i$, cf. 2.28(b). Thus the vectors $x_1, \ldots, x_n$ form a basis for $\mathbb{R}^n$, and every $x \in \mathbb{R}^n$ can be expressed uniquely as a linear combination $\sum_{i=1}^{n} \alpha_i x_i$, where the correspondence $x \leftrightarrow (\alpha_1, \ldots, \alpha_n)$ is continuous in both directions. In particular, $x_0 \leftrightarrow (-\bar{\lambda}_1/\bar{\lambda}_0, \ldots, -\bar{\lambda}_n/\bar{\lambda}_0)$.

Let D be the set of all $x \in \mathbb{R}^n$ with $\alpha_i \leq 0$ for $i = 1, \ldots, n$; this is a closed set whose interior consists of all x with $\alpha_i < 0$ for $i = 1, \ldots, n$. We have $x_0 \in \operatorname{int} D$ but $x_i \notin D$ for all $i \neq 0$. Thus, the open sets $\operatorname{int} D$ and $\mathbb{R}^n \setminus D$ both meet C. If C is connected, it can't lie entirely in the union of these disjoint sets and must meet the boundary of D. There must be a point $x_0' \in C$ having an expression $x_0' = \sum_{i=1}^{n} \alpha_i x_i$ with $\alpha_i \leq 0$ for $i = 1, \ldots, n$ but also $\alpha_i = 0$ for some i, which we can take to be $i = n$. Then $x_0' + \sum_{i=1}^{n-1} |\alpha_i| x_i = 0$, and in dividing through by $1 + \sum_{i=1}^{n-1} |\alpha_i|$ we obtain a representation of 0 as a convex combination of only n points of C, which is contrary to the minimality that was assumed. $\square$

2.30 Corollary (compactness of convex hulls)**.** *For any compact set $C \subset \mathbb{R}^n$, $\operatorname{con} C$ is compact. In particular, the convex hull of a finite set of points is compact; thus, simplices are compact.*

Proof. Let $D \subset (\mathbb{R}^n)^{n+1} \times \mathbb{R}^{n+1}$ consist of all $w = (x_0, \ldots, x_n, \lambda_0, \ldots, \lambda_n)$ with $x_i \in C$, $\lambda_i \geq 0$, $\sum_{i=0}^{n} \lambda_i = 1$. From the theorem, $\operatorname{con} C$ is the image of D under the mapping $F : w \mapsto \sum_{i=0}^{n} \lambda_i x_i$. The image of a compact set under a continuous mapping is compact. $\square$

For a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, there is likewise a notion of convex hull: $\operatorname{con} f$ is the greatest convex function majorized by f. (Recall that a function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *majorized* by a function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ if $g(x) \leq h(x)$ for all x. With the opposite inequality, g is *minorized* by h.) Clearly, $\operatorname{con} f$ is the pointwise supremum of all the convex functions $g \leq f$, this pointwise supremum being a convex function by 2.9(b); the constant function $-\infty$ always serves as such a function g. Alternatively, $\operatorname{con} f$ is the function obtained by taking $\operatorname{epi}(\operatorname{con} f)$ to be the epigraphical closure of $\operatorname{con}(\operatorname{epi} f)$. The condition $f = \operatorname{con} f$ means that f is convex. For $f = \delta_C$ one has $\operatorname{con} f = \delta_D$, where $D = \operatorname{con} C$.

2.31 Proposition (convexification of a function)**.** *For $f : I\!R^n \to \overline{I\!R}$,*

$$\big(\operatorname{con} f\big)(x) = \inf\Big\{\sum\nolimits_{i=0}^{n} \lambda_i f(x_i) \,\Big|\, \sum\nolimits_{i=0}^{n} \lambda_i x_i = x,\ \lambda_i \geq 0, \sum\nolimits_{i=0}^{n} \lambda_i = 1\Big\}.$$

Proof. We apply Theorem 2.29 to $\operatorname{epi} f$ in $I\!R^{n+1}$: every point of $\operatorname{con}(\operatorname{epi} f)$ is a convex combination of at most $n+2$ points of $\operatorname{epi} f$. Actually, at most $n+1$ points x_i at a time are needed in determining the values of $\operatorname{con} f$, since a point $(\bar{x}, \bar{\alpha})$ of $\operatorname{con}(\operatorname{epi} f)$ not representable by fewer than $n+2$ points of $\operatorname{epi} f$ would lie in the interior of some $n+1$-simplex S in $\operatorname{con}(\operatorname{epi} f)$. The vertical line through $(\bar{x}, \bar{\alpha})$ would meet the boundary of S in a point $(\bar{x}, \tilde{\alpha})$ with $\bar{\alpha} > \tilde{\alpha}$, and such a boundary point is representable by $n+1$ of the points in question, cf. 2.28(d). Thus, $(\operatorname{con} f)(x)$ is the infimum of all numbers α such that there exist $n+1$ points $(x_i, \alpha_i) \in \operatorname{epi} f$ and scalars $\lambda_i \geq 0$, with $\sum_{i=0}^{n} \lambda_i (x_i, \alpha_i) = (x, \alpha)$, $\sum_{i=0}^{n} \lambda_i = 1$. This description translates to the formula claimed. □

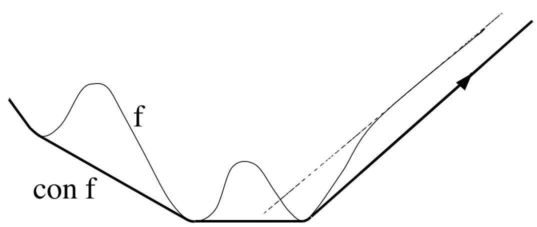

Fig. 2–11. The convex hull of a function.

F. Closures and Continuity

Next we consider the relation of convexity to some topological concepts like closures and interiors of sets and semicontinuity of functions.

2.32 Proposition (convexity and properness of closures)**.** *For a convex set $C \subset I\!R^n$, $\operatorname{cl} C$ is convex. Likewise, for a convex function $f : I\!R^n \to \overline{I\!R}$, $\operatorname{cl} f$ is convex. Moreover $\operatorname{cl} f$ is proper if and only if f is proper.*

Proof. First, for sequences of points x_0^ν and x_1^ν in C converging to x_0 and x_1 in $\operatorname{cl} C$, and for any $\tau \in (0,1)$, the points $x_\tau^\nu = (1-\tau)x_0^\nu + \tau x_1^\nu \in C$, converge to $x_\tau = (1-\tau)x_0 + \tau x_1$, which therefore lies in $\operatorname{cl} C$. This proves the convexity of $\operatorname{cl} C$. In the case of a function f, the epigraph of the function $\operatorname{cl} f$ (the lsc regularization of f introduced in 1(6)-1(7)) is the closure of the epigraph of f, which is convex, and therefore $\operatorname{cl} f$ is convex as well, cf. 2.4.

If f is improper, so is $\operatorname{cl} f$ by 2.5. Conversely, if $\operatorname{cl} f$ is improper it is $-\infty$ on the convex set $D = \operatorname{dom}(\operatorname{cl} f) = \operatorname{cl}(\operatorname{dom} f)$ by 2.5. Consider then a simplex $S = \text{con}\{a_0, a_1, \ldots, a_p\}$ of maximal dimension in D. By a translation if necessary, we can suppose $a_0 = 0$, so that the vectors $a_1, \ldots, a_p$ are linearly independent. The p-dimensional subspace M generated by these vectors must include D, for if not we could find an additional vector $a_{p+1} \in D$ linearly independent of the others, and this would contradict the maximality of p. Our analysis can therefore be reduced to M, which under a change of coordinates can be identified with $\mathbb{R}^p$. To keep notation simple, we can just as well assume that $M = \mathbb{R}^n$, $p = n$. Then S has nonempty interior; let $x \in \text{int}\, S$, so that $x = \sum_{i=0}^n \lambda_i a_i$ with $\lambda_i > 0$, $\sum_{i=0}^n \lambda_i = 1$, cf. 2.28(d). For each i we have $(\text{cl}\, f)(a_i) = -\infty$, so there is a sequence $a_i^\nu \to a_i$ with $f(a_i^\nu) \to -\infty$. Then for ν sufficiently large we have representations $x = \sum_{i=0}^n \lambda_i^\nu a_i^\nu$ with $\lambda_i^\nu > 0$, $\sum_{i=0}^n \lambda_i^\nu = 1$, cf. 2.28(f). From Jensen's inequality 2.2(b), we obtain $f(x) \leq \sum_{i=0}^n \lambda_i^\nu f(a_i^\nu) \leq \max\{f(a_0^\nu), f(a_1^\nu), \ldots, f(a_n^\nu)\} \to -\infty$. Therefore, $f(x) = -\infty$, and f is improper. □

The most important topological consequences of convexity can be traced to a simple fact about line segments which relates the closure $\text{cl}\, C$ to the interior $\text{int}\, C$ of a convex set C, when the interior is nonempty.

2.33 Theorem (line segment principle). *A convex set C has $\text{int}\, C \neq \emptyset$ if and only if $\text{int}(\text{cl}\, C) \neq \emptyset$. In that case, whenever $x_0 \in \text{int}\, C$ and $x_1 \in \text{cl}\, C$, one has $(1-\tau)x_0 + \tau x_1 \in \text{int}\, C$ for all $\tau \in (0,1)$. Thus, $\text{int}\, C$ is convex. Moreover,*

$$\text{cl}\, C = \text{cl}(\text{int}\, C), \qquad \text{int}\, C = \text{int}(\text{cl}\, C).$$

Proof. Leaving the assertions about $\text{int}(\text{cl}\, C)$ to the end, start just with the assumption that $\text{int}\, C \neq \emptyset$. Choose $\varepsilon_0 > 0$ small enough that the ball $\mathbb{B}(x_0, \varepsilon_0)$ is included in C. Writing $\mathbb{B}(x_0, \varepsilon_0) = x_0 + \varepsilon_0 \mathbb{B}$ for $\mathbb{B} = \mathbb{B}(0,1)$, note that our assumption $x_1 \in \text{cl}\, C$ implies $x_1 \in C + \varepsilon_1 \mathbb{B}$ for all $\varepsilon_1 > 0$. For arbitrary fixed $\tau \in (0,1)$, it's necessary to show that the point $x_\tau = (1-\tau)x_0 + \tau x_1$ belongs to $\text{int}\, C$. For this it suffices to demonstrate that $x_\tau + \varepsilon_\tau \mathbb{B} \subset C$ for some $\varepsilon_\tau > 0$. We do so for $\varepsilon_\tau := (1-\tau)\varepsilon_0 - \tau\varepsilon_1$, with ε_1 fixed at any positive value small enough that $\varepsilon_\tau > 0$ (cf. Figure 2–12), by calculating

$$\begin{aligned} x_\tau + \varepsilon_\tau \mathbb{B} &= (1-\tau)x_0 + \tau x_1 + \varepsilon_\tau \mathbb{B} \subset (1-\tau)x_0 + \tau(C + \varepsilon_1 \mathbb{B}) + \varepsilon_\tau \mathbb{B} \\ &= (1-\tau)x_0 + (\tau\varepsilon_1 + \varepsilon_\tau)\mathbb{B} + \tau C = (1-\tau)(x_0 + \varepsilon_0 \mathbb{B}) + \tau C \\ &\subset (1-\tau)C + \tau C = C, \end{aligned}$$

where the convexity of C and $\mathbb{B}$ has been used in invoking 2.23(c).

As a special case of the argument so far, if $x_1 \in \text{int}\, C$ we get $x_\tau \in \text{int}\, C$; thus, $\text{int}\, C$ is convex. Also, any point of $\text{cl}\, C$ can be approached along a line segment by points of $\text{int}\, C$, so $\text{cl}\, C = \text{cl}(\text{int}\, C)$.

It's always true, on the other hand, that $\text{int}\, C \subset \text{int}(\text{cl}\, C)$, so the nonemptiness of $\text{int}\, C$ implies that of $\text{int}(\text{cl}\, C)$. To complete the proof of the theorem, no longer assuming outright that $\text{int}\, C \neq \emptyset$, we suppose $\bar{x} \in \text{int}(\text{cl}\, C)$ and aim at showing $\bar{x} \in \text{int}\, C$.

By 2.28(e), some simplex $S = \text{con}\{a_0, a_1, \ldots, a_n\} \subset \text{cl}\, C$ has $\bar{x} \in \text{int}\, S$.

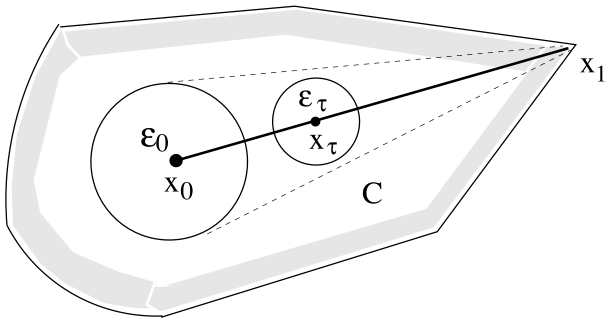

Fig. 2–12. Line segment principle for convex sets.

Then $\bar{x} = \sum_{i=0}^{n} \lambda_i a_i$ with $\sum_{i=0}^{n} \lambda_i = 1$, $\lambda_i > 0$, cf. 2.28(d). Consider in C sequences $a_i^\nu \to a_i$, and let $S^\nu = \mathrm{con}\{a_0^\nu, a_1^\nu, \ldots, a_n^\nu\} \subset C$. For large ν, S^ν is an n-simplex too, and $\bar{x} = \sum_{i=0}^{n} \lambda_i^\nu a_i^\nu$ with $\sum_{i=0}^{n} \lambda_i^\nu = 1$ and $\lambda_i^\nu \to \lambda_i$; see 2.28(f). Eventually $\lambda_i^\nu > 0$, and then $\bar{x} \in \mathrm{int}\, S^\nu$ by 2.28(d), so $\bar{x} \in \mathrm{int}\, C$. □

2.34 Proposition (interiors of epigraphs and level sets)**.** *For f convex on $I\!R^n$,*

$$\mathrm{int}(\mathrm{epi}\, f) = \big\{(x, \alpha) \in I\!R^n \times I\!R \,\big|\, x \in \mathrm{int}(\mathrm{dom}\, f),\ f(x) < \alpha\big\},$$
$$\mathrm{int}(\mathrm{lev}_{\leq\alpha}\, f) = \big\{x \in \mathrm{int}(\mathrm{dom}\, f) \,\big|\, f(x) < \alpha\big\} \ \textit{for}\ \ \alpha \in (\inf f, \infty).$$

Proof. Obviously, if $(\bar{x}, \bar{\alpha}) \in \mathrm{int}(\mathrm{epi}\, f)$ there's a ball around $(\bar{x}, \bar{\alpha})$ within $\mathrm{epi}\, f$, so that $\bar{x} \in \mathrm{int}(\mathrm{dom}\, f)$ and $f(\bar{x}) < \bar{\alpha}$.

On the other hand, if the latter properties hold there's a simplex $S = \mathrm{con}\{a_0, a_1, \ldots, a_n\}$ with $\bar{x} \in \mathrm{int}\, S \subset \mathrm{dom}\, f$, cf. 2.28(e). Each $x \in S$ is a convex combination $\sum_{i=0}^{n} \lambda_i a_i$ and satisfies $f(x) \leq \sum_{i=0}^{n} \lambda_i f(a_i)$ by Jensen's inequality in 2.2(b) and therefore also satisfies $f(x) \leq \max\big\{f(a_0), \ldots, f(a_n)\big\}$. For $\tilde{\alpha} := \max\big\{f(a_0), \ldots, f(a_n)\big\}$ the open set $\mathrm{int}\, S \times (\tilde{\alpha}, \infty)$ lies then within $\mathrm{epi}\, f$. The vertical line through $(\bar{x}, \bar{\alpha})$ thus contains a point $(\bar{x}, \alpha_0) \in \mathrm{int}(\mathrm{epi}\, f)$ with $\alpha_0 > \bar{\alpha}$, but it also contains a point $(\bar{x}, \alpha_1) \in \mathrm{epi}\, f$ with $\alpha_1 < \bar{\alpha}$, inasmuch as $f(\bar{x}) < \bar{\alpha}$. The set $\mathrm{epi}\, f$ is convex because f is convex, so by the line segment principle in 2.33 all the points between $(\bar{x}, \alpha_0)$ and $(\bar{x}, \alpha_1)$ must belong to $\mathrm{int}(\mathrm{epi}\, f)$. This applies to $(\bar{x}, \bar{\alpha})$ in particular.

Consider now a level set $\mathrm{lev}_{\leq\bar{\alpha}}\, f$. If $\bar{x} \in \mathrm{int}(\mathrm{dom}\, f)$ and $f(\bar{x}) < \bar{\alpha}$, then some ball around $(\bar{x}, \bar{\alpha})$ lies in $\mathrm{epi}\, f$, so $f(x) \leq \bar{\alpha}$ for all x in some neighborhood of $\bar{x}$. Then $\bar{x} \in \mathrm{int}(\mathrm{lev}_{\leq\bar{\alpha}}\, f)$. Conversely, if $\bar{x} \in \mathrm{int}(\mathrm{lev}_{\leq\bar{\alpha}}\, f)$ and $\inf f < \bar{\alpha} < \infty$, we must have $\bar{x} \in \mathrm{int}(\mathrm{dom}\, f)$, but also there's a point $x_0 \in \mathrm{dom}\, f$ with $f(x_0) < \bar{\alpha}$. For $\varepsilon > 0$ sufficiently small the point $x_1 = \bar{x} + \varepsilon(\bar{x} - x_0)$ still belongs to $\mathrm{lev}_{\leq\bar{\alpha}}\, f$. Then $\bar{x} = (1 - \tau)x_0 + \tau x_1$ for $\tau = 1/(1 + \varepsilon)$, and we get $f(\bar{x}) \leq (1 - \tau)f(x_0) + \tau f(x_1) < \bar{\alpha}$, which is the desired inequality. □

2.35 Theorem (continuity properties of convex functions)**.** *A convex function $f : I\!R^n \to \overline{I\!R}$ is continuous on* $\mathrm{int}(\mathrm{dom}\, f)$ *and therefore agrees with* $\mathrm{cl}\, f$ *on* $\mathrm{int}(\mathrm{dom}\, f)$*, this set being the same as* $\mathrm{int}(\mathrm{dom}(\mathrm{cl}\, f))$*. (Also, f agrees with* $\mathrm{cl}\, f$ *as having the value ∞ outside of* $\mathrm{dom}(\mathrm{cl}\, f)$*, hence in particular outside of* $\mathrm{cl}(\mathrm{dom}\, f)$*.) Moreover,*

$$(\mathrm{cl}\, f)(x) \;=\; \lim_{\tau \nearrow 1} f\big((1-\tau)x_0 + \tau x\big) \;\; \textit{for all } x \textit{ if } x_0 \in \mathrm{int}(\mathrm{dom}\, f). \qquad 2(11)$$

If f is lsc, it must in addition be continuous relative to the convex hull of any finite subset of $\mathrm{dom}\, f$*, in particular any line segment in* $\mathrm{dom}\, f$*.*

Proof. The closure formula will be proved first. We know from the basic expression for $\mathrm{cl}\, f$ in 1(7) that $(\mathrm{cl}\, f)(x) \leq \liminf_{\tau \nearrow 1} f((1-\tau)x_0+\tau x)$, so it will suffice to show that if $(\mathrm{cl}\, f)(x) \leq \alpha \in I\!R$ then $\limsup_{\tau \nearrow 1} f((1-\tau)x_0+\tau x) \leq \alpha$. The assumption on α means that $(x,\alpha) \in \mathrm{cl}(\mathrm{epi}\, f)$, cf. 1(6). On the other hand, for any real number $\alpha_0 > f(x_0)$ we have $(x_0,\alpha_0) \in \mathrm{int}(\mathrm{epi}\, f)$ by 2.34. Then by the line segment principle in Theorem 2.33, as applied to the convex set $\mathrm{epi}\, f$, the points $(1-\tau)(x_0,\alpha_0) + \tau(x,\alpha)$ for $\tau \in (0,1)$ belong to $\mathrm{int}(\mathrm{epi}\, f)$. In particular, then, $f((1-\tau)x_0 + \tau x) < (1-\tau)\alpha_0 + \tau\alpha$ for $\tau \in (0,1)$. Taking the upper limit on both sides as $\tau \nearrow 1$, we get the inequality needed.

When the closure formula is applied with $x = x_0$ it yields the fact that $\mathrm{cl}\, f$ agrees with f on $\mathrm{int}(\mathrm{dom}\, f)$. Hence f is lsc on $\mathrm{int}(\mathrm{dom}\, f)$. But f is usc there by 1.13(b) in combination with the characterization of $\mathrm{int}(\mathrm{epi}\, f)$ in 2.34. It follows that f is continuous on $\mathrm{int}(\mathrm{dom}\, f)$. We have $\mathrm{int}(\mathrm{dom}\, f) = \mathrm{int}(\mathrm{dom}(\mathrm{cl}\, f))$ by 2.33, because $\mathrm{dom}\, f \subset \mathrm{dom}(\mathrm{cl}\, f) \subset \mathrm{cl}(\mathrm{dom}\, f)$. The same inclusions also yield $\mathrm{cl}(\mathrm{dom}\, f) = \mathrm{cl}(\mathrm{dom}(\mathrm{cl}\, f))$.

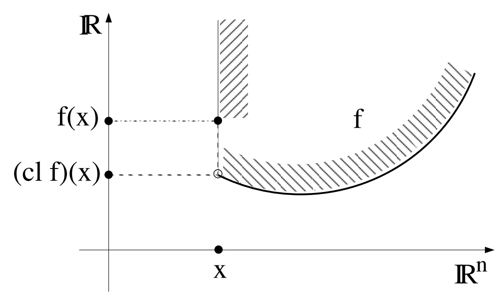

Fig. 2–13. Closure operation on a convex function.

Consider now a finite set $C \subset \mathrm{dom}\, f$. Under the assumption that f is lsc, we wish to show that f is continuous relative to $\mathrm{con}\, C$. But $\mathrm{con}\, C$ is the union of the simplices generated by the points of C, as shown in the first part of the proof of Theorem 2.29, and there are only finitely many of these. If a function is continuous on a finite family of sets, it is also continuous on their union. It suffices therefore to show that f is continuous relative to any simplex $S = \mathrm{con}\{a_0, a_1, \ldots, a_p\} \subset \mathrm{dom}\, f$. For simplicity of notation we can translate so that $0 \in S$ and investigate continuity at 0; we have a unique representation of 0 as a convex combination $\sum_{i=0}^{p} \bar{\lambda}_i a_i$.

We argue next that S is the union of the finitely many other simplices having 0 as one vertex and certain a_i's as the other vertices. This will enable us to reduce further to the study of continuity relative to such a simplex.

Consider any point $\tilde{x} \neq 0$ in S and represent it as a convex combination $\sum_{i=0}^{p} \tilde{\lambda}_i a_i$. The points $x_\tau = (1-\tau)0 + \tau\tilde{x}$ on the line through 0 and $\tilde{x}$ can't all be in S, which is compact (by 2.30), so there must be a highest τ with $x_\tau \in S$. For this we have $\tau \geq 1$ and $x_\tau = \sum_{i=0}^p \mu_i a_i$ for $\mu_i = (1-\tau)\bar{\lambda}_i + \tau\tilde{\lambda}_i$ with $\sum_{i=0}^p \mu_i = 1$, where $\mu_i \geq 0$ for all i but $\mu_i = 0$ for at least one i such that $\bar{\lambda}_i > \tilde{\lambda}_i$ (or τ wouldn't be the highest). We can suppose $\mu_0 = 0$; then in particular $\bar{\lambda}_0 > 0$. Since $\tilde{x}$ lies on the line segment joining 0 with x_τ, it belongs to $\operatorname{con}\{0, a_1, \ldots, a_p\}$. This is a simplex, for if not the vectors $a_1, \ldots, a_p$ would be linearly dependent: we would have $\sum_{i=1}^p \eta_i a_i = 0$ for certain coefficients η_i, not all 0. It's impossible that $\sum_{i=1}^p \eta_i = 0$, because $\{a_0, a_1, \ldots, a_p\}$ are affinely independent, so if this were the case we could, by rescaling, arrange that $\sum_{i=1}^p \eta_i = 1$. But then in defining $\eta_0 = 0$ we would be able to conclude from the affine independence that $\eta_i = \bar{\lambda}_i$ for $i = 0, \ldots, p$, because $0 = \sum_{i=0}^p (\eta_i - \bar{\lambda}_i) a_i$ with $\sum_{i=0}^p (\eta_i - \bar{\lambda}_i) = 0$. This would contradict $\bar{\lambda}_0 > 0$.

We have gotten to where a simplex $S_0 = \operatorname{con}\{0, a_1, \ldots, a_p\}$ lies in $\operatorname{dom} f$ and we need to prove that f is not just lsc relative to S_0 at 0, as assumed, but also usc. Any point of S_0 has a unique expression $\sum_{i=1}^p \lambda_i a_i$ with $\lambda_i \geq 0$ and $\sum_{i=1}^p \lambda_i \leq 1$, and as the point approaches 0 these coefficients vanish, cf. 2.28(c). The corresponding value of f is bounded above by $\lambda_0 f(0) + \sum_{i=1}^p \lambda_i f(a_i)$ through Jensen's inequality 2.2(b), where $\lambda_0 = 1 - \sum_{i=1}^p \lambda_i$, and in the limit this bound is $f(0)$. Thus, $\limsup_{x \to 0} f(x) \leq f(0)$ for $x \in S_0$. □

2.36 Corollary (finite convex functions)**.** *A finite, convex function f on an open, convex set $O \neq \emptyset$ in $\mathbb{R}^n$ is continuous on O. Such a function has a unique extension to a proper, lsc, convex function f on $\mathbb{R}^n$ with* $\operatorname{dom} f \subset \operatorname{cl} O$.

Proof. Apply the theorem to the convex function g that agrees with f on O but takes on ∞ everywhere else. Then $\operatorname{int}(\operatorname{dom} g) = O$. □

Finite convex functions will be seen in 9.14 to have the even stronger property of Lipschitz continuity locally.

2.37 Corollary (convex functions of a single real variable)**.** *Any lsc, convex function $f : \mathbb{R} \to \overline{\mathbb{R}}$ is continuous with respect to* $\operatorname{cl}(\operatorname{dom} f)$.

Proof. This is clear from 2(11) when $\operatorname{int}(\operatorname{dom} f) \neq \emptyset$; otherwise it's trivial. □

Not everything about the continuity properties of convex functions is good news. The following example provides clear insight into what can go wrong.

2.38 Example (discontinuity and unboundedness)**.** *On $\mathbb{R}^2$, the function*

$$f(x_1, x_2) = \begin{cases} x_1^2/2x_2 & \text{if } x_2 > 0, \\ 0 & \text{if } x_1 = 0 \text{ and } x_2 = 0, \\ \infty & \text{otherwise,} \end{cases}$$

is lsc, proper, convex and positively homogeneous. Nonetheless, f fails to be continuous relative to the compact, convex set $C = \{(x_1, x_2) \mid x_1^4 \leq x_2 \leq 1\}$ in $\operatorname{dom} f$*, despite f being continuous relative to every line segment in C. In fact f is not even bounded above on C.*

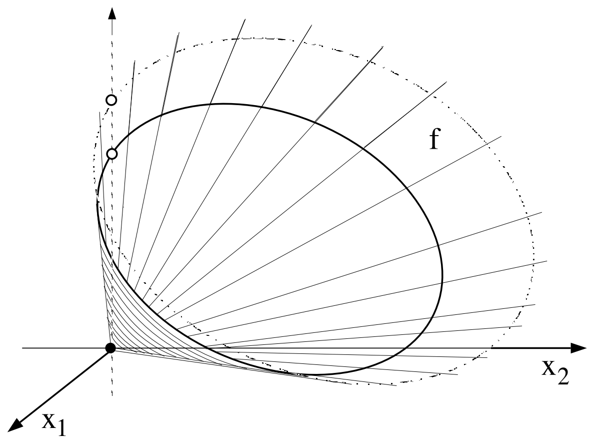

Fig. 2–14. Example of discontinuity.

Detail. The convexity of f on the open half-plane $H = \{(x_1, x_2) \mid x_2 > 0\}$ can be verified by the second-derivative test in 2.14, and the convexity relative to all of $I\!R^n$ follows then via the extension procedure in 2.36. The epigraph of f is actually a circular cone whose axis is the ray through $(0,1,1)$ and whose boundary contains the rays through $(0,1,0)$ and $(0,0,1)$, see Figure 2–14. The unboundedness and lack of continuity of f relative to H are seen from its behavior along the boundary of C at 0, as indicated from the piling up of the level sets of this function as shown in Figure 2–15. □

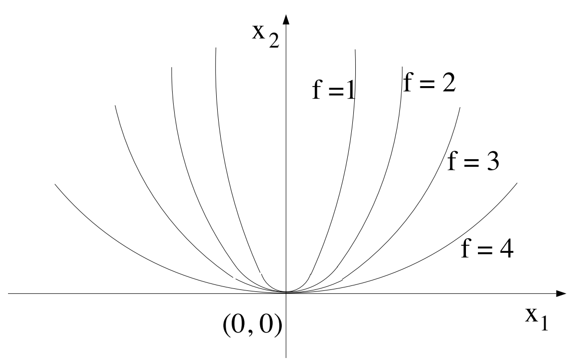

Fig. 2–15. Nest of level sets illustrating discontinuity.

G.* Separation

Properties of closures and interiors of convex sets lead also to a famous principle of separation. A hyperplane $\{x \mid \langle a, x\rangle = \alpha\}$, where $a \neq 0$ and $\alpha \in I\!R$, is said to *separate* two sets C_1 and C_2 in $I\!R^n$ if C_1 is included in one of the corresponding closed half-spaces $\{x \mid \langle a, x\rangle \leq \alpha\}$ or $\{x \mid \langle a, x\rangle \geq \alpha\}$, while C_2 is included in the other. The separation is said to be *proper* if the hyperplane itself doesn't actually include *both* C_1 and C_2. As a related twist in wording, we say that C_1 and C_2 *can, or cannot, be separated* according to whether a separating hyperplane exists; similarly for whether they can, or cannot, be separated *properly*. Additional variants are *strict* separation, where the two sets lie in complementary *open* half-spaces, and *strong* separation, where they lie in different half-spaces $\{x \mid \langle a, x\rangle \le \alpha_1\}$ and $\{x \mid \langle a, x\rangle \ge \alpha_2\}$ with $\alpha_1 < \alpha_2$. Strong separation implies strict separation, but not conversely.

2.39 Theorem (separation)**.** *Two nonempty, convex sets C_1 and C_2 in $I\!R^n$ can be separated by a hyperplane if and only if $0 \notin \operatorname{int}(C_1 - C_2)$. The separation must be proper if also $\operatorname{int}(C_1 - C_2) \neq \emptyset$. Both conditions certainly hold when $\operatorname{int} C_1 \neq \emptyset$ but $C_2 \cap \operatorname{int} C_1 = \emptyset$, or when $\operatorname{int} C_2 \neq \emptyset$ but $C_1 \cap \operatorname{int} C_2 = \emptyset$.*

Strong separation is possible if and only if $0 \notin \operatorname{cl}(C_1 - C_2)$. This is ensured in particular when $C_1 \cap C_2 = \emptyset$ with both sets closed and one of them bounded.

Proof. In the case of a separating hyperplane $\{x \mid \langle a, x\rangle = \alpha\}$ with $C_1 \subset \{x \mid \langle a, x\rangle \le \alpha\}$ and $C_2 \subset \{x \mid \langle a, x\rangle \ge \alpha\}$, we have $0 \ge \langle a, x_1 - x_2\rangle$ for all $x_1 - x_2 \in C_1 - C_2$, in which case obviously $0 \notin \operatorname{int}(C_1 - C_2)$. This condition is therefore necessary for separation. Moreover if the separation weren't proper, we would have $0 = \langle a, x_1 - x_2\rangle$ for all $x_1 - x_2 \in C_1 - C_2$, and this isn't possible if $\operatorname{int}(C_1 - C_2) \neq \emptyset$. In the case where $\operatorname{int} C_1 \neq \emptyset$ but $C_2 \cap \operatorname{int} C_1 = \emptyset$, we have $0 \notin (\operatorname{int} C_1) - C_2$ where the set $(\operatorname{int} C_1) - C_2$ is convex (by 2.23 and the convexity of interiors in 2.33) as well as open (since it is the union of translates $(\operatorname{int} C_1) - x_2$, each of which is an open set). Then $(\operatorname{int} C_1) - C_2 \subset C_1 - C_2 \subset \operatorname{cl}\left[(\operatorname{int} C_1) - C_2\right]$ (through 2.33), so actually $\operatorname{int}(C_1 - C_2) = (\operatorname{int} C_1) - C_2$ (once more through the relations in 2.33). In this case, therefore, we have $0 \notin \operatorname{int}(C_1 - C_2) \neq \emptyset$. Similarly, this holds when $\operatorname{int} C_2 \neq \emptyset$ but $C_1 \cap \operatorname{int} C_2 = \emptyset$.

The sufficiency of the condition $0 \notin \operatorname{int}(C_1 - C_2)$ for the existence of a separating hyperplane is all that remains to be established. Note that because $C_1 - C_2$ is convex (cf. 2.23), so is $C := \operatorname{cl}(C_1 - C_2)$. We need only produce a vector $a \neq 0$ such that $\langle a, x\rangle \le 0$ for all $x \in C$, for then we'll have $\langle a, x_1\rangle \le \langle a, x_2\rangle$ for all $x_1 \in C_1$ and $x_2 \in C_2$, and separation will be achieved with any α in the nonempty interval between $\sup_{x_1 \in C_1}\langle a, x_1\rangle$ and $\inf_{x_2 \in C_2}\langle a, x_2\rangle$.

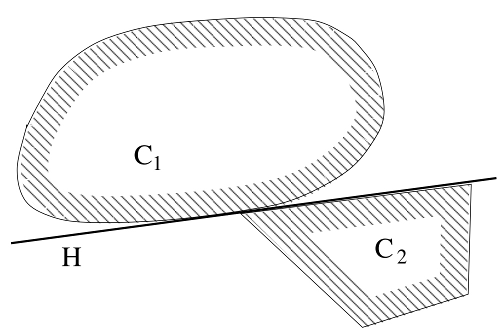

Fig. 2–16. Separation of convex sets.

Let $\bar{x}$ denote the unique point of C nearest to 0 (cf. 2.25). For any $x \in C$ the segment $[\bar{x}, x]$ lies in C, so the function $g(\tau) = \frac{1}{2}|(1-\tau)\bar{x} + \tau x|^2$ satisfies

$g(\tau) \geq g(0)$ for $\tau \in (0,1)$. Hence $0 \leq g'(0) = \langle \bar{x}, x - \bar{x} \rangle$. If $0 \notin C$, so that $\bar{x} \neq 0$, we can take $a = -\bar{x}$ and be done. A minor elaboration of this argument confirms that *strong* separation is possible *if and only if* $0 \notin C$. Certainly we have $0 \notin C$ in particular when $C_1 \cap C_2 = \emptyset$ and $C_1 - C_2$ is closed, i.e., $C_1 - C_2 = C$; it's easy to see that $C_1 - C_2$ is closed if both C_1 and C_2 are closed and one is actually compact.

When $0 \in C$, so that $\bar{x} = 0$, we need to work harder to establish the sufficiency of the condition $0 \notin \operatorname{int}(C_1 - C_2)$ for the existence of a separating hyperplane. From $C = \operatorname{cl}(C_1 - C_2)$ and $0 \notin \operatorname{int}(C_1 - C_2)$, we have $0 \notin \operatorname{int} C$ by Theorem 2.33. Then there is a sequence $x^\nu \to 0$ with $x^\nu \notin C$. Let $\bar{x}^\nu$ be the unique point of C nearest to x^ν; then $\bar{x}^\nu \neq x^\nu$, $\bar{x}^\nu \to 0$. Applying to x^ν and $\bar{x}^\nu$ the same argument we earlier applied to 0 and its projection on C when this wasn't necessarily 0 itself, we verify that $\langle x - \bar{x}^\nu,\, x^\nu - \bar{x}^\nu \rangle \leq 0$ for every $x \in C$. Let a be any cluster point of the sequence of vectors $a^\nu := (x^\nu - \bar{x}^\nu)/|x^\nu - \bar{x}^\nu|$, which have $|a^\nu| = 1$ and $\langle x - \bar{x}^\nu, a^\nu \rangle \leq 0$ for every $x \in C$. Then $|a| = 1$, so $a \neq 0$, and in the limit we have $\langle x, a \rangle \leq 0$ for every $x \in C$, as required. □

H.* Relative Interiors

Every nonempty affine set has a well determined *dimension*, which is the dimension of the linear subspace of which it is a translate, cf. 2.11. Singletons are 0-dimensional, lines are 1-dimensional, and so on. The hyperplanes in $I\!R^n$ are the affine sets that are $(n-1)$-dimensional.

For any convex set C in $I\!R^n$, the *affine hull* of C is the smallest affine set that includes C (it's the intersection of all the affine sets that include C). The interior of C relative to its affine hull is the *relative interior* of C, denoted by $\operatorname{rint} C$. This coincides with the true interior when the affine hull is all of $I\!R^n$, but is able to serve as a robust substitute for $\operatorname{int} C$ when $\operatorname{int} C = \emptyset$.

2.40 Proposition (relative interiors of convex sets)**.** *For $C \subset I\!R^n$ nonempty and convex, the set $\operatorname{rint} C$ is nonempty and convex with $\operatorname{cl}(\operatorname{rint} C) = \operatorname{cl} C$ and $\operatorname{rint}(\operatorname{cl} C) = \operatorname{rint} C$. If $x_0 \in \operatorname{rint} C$ and $x_1 \in \operatorname{cl} C$, then $\operatorname{rint}[x_0, x_1] \subset \operatorname{rint} C$.*

Proof. Through a translation if necessary, we can suppose that $0 \in C$. We can suppose further that C contains more than just 0, because otherwise the assertion is trivial. Let $a_1, \ldots, a_p$ be a set of linearly independent vectors chosen from C with p as high as possible, and let M be the p-dimensional subspace of $I\!R^n$ generated by these vectors. Then M has to be the affine hull of C, because the affine hull has to be an affine set containing M, and yet there can't be any point of C outside of M or the maximality of p would be contradicted. The p-simplex $\operatorname{con}\{0, a_1, \ldots, a_p\}$ in C has nonempty interior relative to M (which can be identified with $I\!R^p$ through a change of coordinates), so $\operatorname{rint} C \neq \emptyset$. The relations between $\operatorname{rint} C$ and $\operatorname{cl} C$ follow then from the ones in 2.33. □

2.41 Exercise (relative interior criterion)**.** *For a convex set $C \subset \mathbb{R}^n$, one has $x \in \operatorname{rint} C$ if and only if $x \in C$ and, for every $x_0 \neq x$ in C, there exists $x_1 \in C$ such that $x \in \operatorname{rint}[x_0, x_1]$.*

Guide. Reduce to the case where C is n-dimensional. □

The properties in Proposition 2.40 are crucial in building up a calculus of closures and relative interiors of convex sets.

2.42 Proposition (relative interiors of intersections)**.** *For a family of convex sets $C_i \subset \mathbb{R}^n$ indexed by $i \in I$ and such that $\bigcap_{i\in I} \operatorname{rint} C_i \neq \emptyset$, one has $\operatorname{cl} \bigcap_{i\in I} C_i = \bigcap_{i\in I} \operatorname{cl} C_i$. If I is finite, then also $\operatorname{rint} \bigcap_{i\in I} C_i = \bigcap_{i\in I} \operatorname{rint} C_i$.*

Proof. On general grounds, $\operatorname{cl} \bigcap_{i\in I} C_i \subset \bigcap_{i\in I} \operatorname{cl} C_i$. For the reverse consider any $x_0 \in \bigcap_{i\in I} \operatorname{rint} C_i$ and $x_1 \in \bigcap_{i\in I} \operatorname{cl} C_i$. We have $[x_0, x_1) \subset \bigcap_{i\in I} \operatorname{rint} C_i \subset \bigcap_{i\in I} C_i$ by 2.40, so $x_1 \in \operatorname{cl} \bigcap_{i\in I} C_i$. This argument has the by-product that $\operatorname{cl} \bigcap_{i\in I} C_i = \operatorname{cl} \bigcap_{i\in I} \operatorname{rint} C_i$. In taking the relative interior on both sides of this equation, we obtain via 2.40 that $\operatorname{rint} \bigcap_{i\in I} C_i = \operatorname{rint} \bigcap_{i\in I} \operatorname{rint} C_i$. The fact that $\operatorname{rint} \bigcap_{i\in I} \operatorname{rint} C_i = \bigcap_{i\in I} \operatorname{rint} C_i$ when I is finite is clear from 2.41. □

2.43 Proposition (relative interiors in product spaces)**.** *For a convex set G in $\mathbb{R}^n \times \mathbb{R}^m$, let X be the image of G under the projection $(x, u) \mapsto x$, and for each $x \in X$ let $S(x) = \{u \mid (x, u) \in G\}$. Then X and $S(x)$ are convex, and*

$$(x, u) \in \operatorname{rint} G \iff x \in \operatorname{rint} X \text{ and } u \in \operatorname{rint} S(x).$$

Proof. We have X convex by 2.21 and $S(x)$ convex by 2.9(a). We begin by supposing that $(x, u) \in \operatorname{rint} G$ and proving $x \in \operatorname{rint} X$ and $u \in \operatorname{rint} S(x)$. Certainly $x \in X$ and $u \in S(x)$. For any $x_0 \in X$ there exists $u_0 \in S(x_0)$, and we have $(x_0, u_0) \in G$. If $x_0 \neq x$, then $(x_0, u_0) \neq (x, u)$ and we can obtain from the criterion in 2.41 as applied to $\operatorname{rint} G$ the existence of $(x_1, u_1) \neq (x, u)$ in G such that $(x, u) \in \operatorname{rint}\big[(x_0, u_0), (x_1, u_1)\big]$. Then $x_1 \in X$ and $x \in \operatorname{rint}[x_0, x_1]$. This proves by 2.41 (as applied to X) that $x \in \operatorname{rint} X$. At the same time, from the fact that $u \in S(x)$ we argue via 2.41 (as applied to G) that if $u_0' \in S(x)$ and $u_0' \neq u$, there exists $(x_1', u_1') \in G$ for which $(x, u) \in \operatorname{rint}\big[(x, u_0'), (x_1', u_1')\big]$. Then necessarily $x_1' = x$, and we conclude $u \in \operatorname{rint}[u_0', u_1']$, thereby establishing by criterion 2.41 (as applied to $S(x)$) that $u \in \operatorname{rint} S(x)$.

We tackle the converse now by assuming $x \in \operatorname{rint} X$ and $u \in \operatorname{rint} S(x)$ (so in particular $(x, u) \in G$) and aiming to prove $(x, u) \in \operatorname{rint} G$. Relying yet again on 2.41, we suppose $(x_0, u_0) \in G$ with $(x_0, u_0) \neq (x, u)$ and reduce the argument to showing that for some choice of $(x_1, u_1) \in G$ the pair (x, u) lies in $\operatorname{rint}\big[(x_0, u_0), (x_1, u_1)\big]$. If $x_0 = x$, this follows immediately from criterion 2.41 (as applied to $S(x)$ with $x_1 = x$). We assume therefore that $x_0 \neq x$. Since $x \in \operatorname{rint} X$, there then exists by 2.41 a vector $\bar{x}_1 \in X$, $\bar{x}_1 \neq x_0$, with $x \in \operatorname{rint}[x_0, \bar{x}_1]$, i.e., $x = (1 - \tau)x_0 + \tau\bar{x}_1$ for a certain $\tau \in (0, 1)$. We have $(\bar{x}_1, \bar{u}_1) \in G$ for some $\bar{u}_1$, and consequently $(1 - \tau)(x_0, u_0) + \tau(\bar{x}_1, \bar{u}_1) \in G$. If the point $u_0' := (1 - \tau)u_0 + \tau\bar{u}_1$ coincides with u, we get the desired sort of representation of (x, u) as a relative interior point of the line segment $\big[(x_0, u_0), (\bar{x}_1, \bar{u}_1)\big]$, whose

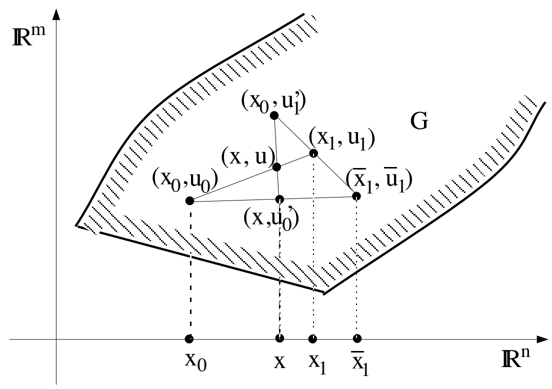

Fig. 2–17. Relative interior argument in a product space.

endpoints lie in G. If not, u_0' is a point of $S(x)$ different from u and, because $u \in \operatorname{rint} S(x)$, we get from 2.41 the existence of some $u_1' \in S(x)$, $u_1' \neq u_0'$, with $u \in \operatorname{rint}[u_0', u_1']$, i.e., $u = (1-\mu)u_0' + \mu u_1'$ for some $\mu \in (0,1)$. This yields

$$\begin{aligned}
(x,u) &= (1-\mu)(x,u_0') + \mu(x,u_1') \\
&= (1-\mu)\big[(1-\tau)(x_0,u_0) + \tau(\bar{x}_1,\bar{u}_1)\big] + \mu(x,u_1') \\
&= \tau_0(x_0,u_0) + \tau_1(\bar{x}_1,\bar{u}_1) + \tau_2(x,u_1'), \\
&\qquad\qquad\qquad \text{where } \ 0 < \tau_i < 1, \ \tau_0+\tau_1+\tau_2 = 1.
\end{aligned}$$

We can write this as $(x,u) = (1-\tau')(x_0,u_0) + \tau'(x_1,u_1)$ with $\tau' := \tau_1 + \tau_2 = 1 - \tau_0 \in (0,1)$ and $(x_1,u_1) := [\tau_1/(\tau_1+\tau_2)](\bar{x}_1,\bar{u}_1) + [\tau_2/(\tau_1+\tau_2)](x,u_1') \in G$. Here $(x_1,u_1) \neq (x_0,u_0)$, because otherwise the definition of x_1 would imply $x = \bar{x}_1$ in contradiction to our knowledge that $x \in \operatorname{rint}[x_0,\bar{x}_1]$. Thus, $(x,u) \in \operatorname{rint}\big[(x_0,u_0),(x_1,u_1)\big]$, and we can conclude that $(x,u) \in \operatorname{rint} G$. □

2.44 Proposition (relative interiors of set images)**.** *For linear $L : I\!R^n \to I\!R^m$,*

(a) $\operatorname{rint} L(C) = L(\operatorname{rint} C)$ *for any convex set* $C \subset I\!R^n$*,*

(b) $\operatorname{rint} L^{-1}(D) = L^{-1}(\operatorname{rint} D)$ *for any convex set* $D \subset I\!R^m$ *such that* $\operatorname{rint} D$ *meets the range of L, and in this case also* $\operatorname{cl} L^{-1}(D) = L^{-1}(\operatorname{cl} D)$.

Proof. First we prove the relative interior equation in (b). In $I\!R^n \times I\!R^m$ let $G = \big\{(x,u) \,\big|\, u = L(x) \in D\big\} = M \cap (I\!R^n \times D)$, where M is the graph of L, a certain linear subspace of $I\!R^n \times I\!R^m$. We'll apply 2.43 to G, whose projection on $I\!R^n$ is $X = L^{-1}(D)$. Our assumption that $\operatorname{rint} D$ meets the range of L means $M \cap \operatorname{rint}[I\!R^n \times D] \neq \emptyset$, where $M = \operatorname{rint} M$ because M is an affine set. Then $\operatorname{rint} G = M \cap \operatorname{rint}(I\!R^n \times D)$ by 2.42; thus, (x,u) belongs to $\operatorname{rint} G$ if and only if $u = L(x)$ and $u \in \operatorname{rint} D$. But by 2.43 (x,u) belongs to $\operatorname{rint} G$ if and only if $x \in \operatorname{rint} X$ and $u \in \operatorname{rint}\{L(x)\} = \{L(x)\}$. We conclude that $x \in \operatorname{rint} L^{-1}(D)$ if and only if $L(x) \in \operatorname{rint} D$, as claimed in (b). The closure assertion follows similarly from the fact that $\operatorname{cl} G = M \cap \operatorname{cl}(I\!R^n \times D)$ by 2.42.

The argument for (a) just reverses the roles of x and u. We take $G = \big\{(x,u) \,\big|\, x \in C, \ u = L(x)\big\} = M \cap (C \times I\!R^m)$ and consider the projection U of G in $I\!R^m$, once again applying 2.42 and 2.43. □

2.45 Exercise (relative interiors in set algebra).

(a) *If* $C = C_1 \times C_2$ *for convex sets* $C_i \subset \mathbb{R}^{n_i}$, *then* $\operatorname{rint} C = \operatorname{rint} C_1 \times \operatorname{rint} C_2$.

(b) *For convex sets* C_1 *and* C_2 *in* $\mathbb{R}^n$, $\operatorname{rint}(C_1 + C_2) = \operatorname{rint} C_1 + \operatorname{rint} C_2$.

(c) *For a convex set* C *and any scalar* $\lambda \in \mathbb{R}$, $\operatorname{rint}(\lambda C) = \lambda(\operatorname{rint} C)$.

(d) *For convex sets* C_1 *and* C_2 *in* $\mathbb{R}^n$, *the condition* $0 \in \operatorname{int}(C_1 - C_2)$ *holds if and only if* $0 \in \operatorname{rint}(C_1 - C_2)$ *but there is no hyperplane* $H \supset C_1 \cup C_2$. *This is true in particular if either* $C_1 \cap \operatorname{int} C_2 \neq \emptyset$ *or* $C_2 \cap \operatorname{int} C_1 \neq \emptyset$.

(e) *Convex sets* $C_1 \neq \emptyset$ *and* $C_2 \neq \emptyset$ *in* $\mathbb{R}^n$ *can be separated properly if and only if* $0 \notin \operatorname{rint}(C_1 - C_2)$. *This condition is equivalent to* $\operatorname{rint} C_1 \cap \operatorname{rint} C_2 = \emptyset$.

Guide. Get (a) from 2.41 or 2.43. Get (b) and (c) from 2.44(a) through special choices of a linear transformation. In (d) verify that $C_1 \cup C_2$ lies in a hyperplane if and only if $C_1 - C_2$ lies in a hyperplane. Observe too that when $C_1 \cap \operatorname{int} C_2 \neq \emptyset$ the set $C_1 - \operatorname{int} C_2$ is open and contains 0. For (e), work with Theorem 2.39. Get the equivalent statement of the relative interior condition out of the algebra in (b) and (c). □

2.46 Exercise (closures of functions).

(a) *If* f_1 *and* f_2 *are convex on* $\mathbb{R}^n$ *with* $\operatorname{rint}(\operatorname{dom} f_1) = \operatorname{rint}(\operatorname{dom} f_2)$, *and on this common set* f_1 *and* f_2 *agree, then* $\operatorname{cl} f_1 = \operatorname{cl} f_2$.

(b) *If* f_1 *and* f_2 *are convex on* $\mathbb{R}^n$ *with* $\operatorname{rint}(\operatorname{dom} f_1) \cap \operatorname{rint}(\operatorname{dom} f_2) \neq \emptyset$, *then* $\operatorname{cl}(f_1 + f_2) = \operatorname{cl} f_1 + \operatorname{cl} f_2$.

(c) *If* $f = g \circ L$ *for a linear mapping* $L : \mathbb{R}^n \to \mathbb{R}^m$ *and a convex function* $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ *such that* $\operatorname{rint}(\operatorname{dom} g)$ *meets the range of* L, *then* $\operatorname{rint}(\operatorname{dom} f) = L^{-1}\big(\operatorname{rint}(\operatorname{dom} g)\big)$ *and* $\operatorname{cl} f = (\operatorname{cl} g) \circ L$.

Guide. Adapt Theorem 2.35 to relative interiors. In (b) use 2.42. In (c) argue that $\operatorname{epi} f = L_0^{-1}(\operatorname{epi} g)$ with $L_0 : (x, \alpha) \mapsto (L(x), \alpha)$, and apply 2.44(b). □

I.* Piecewise Linear Functions

Polyhedral sets are important not only in the geometric analysis of systems of linear constraints, as in 2.10, but also in piecewise linearity.

2.47 Definition (piecewise linearity).

(a) *A mapping* $F : D \to \mathbb{R}^m$ *for a set* $D \subset \mathbb{R}^n$ *is piecewise linear on* D *if* D *can be represented as the union of finitely many polyhedral sets, relative to each of which* $F(x)$ *is given by an expression of the form* $Ax + a$ *for some matrix* $A \in \mathbb{R}^{m \times n}$ *and vector* $a \in \mathbb{R}^m$.

(b) *A function* $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ *is piecewise linear if it is piecewise linear on* $D = \operatorname{dom} f$ *as a mapping into* $\mathbb{R}$.

Piecewise linear functions and mappings should perhaps be called 'piecewise affine', but the popular term is retained here. Definition 2.47 imposes no conditions on the extent to which the polyhedral sets whose union is D may overlap or be redundant, although if such a representation exists a refinement with better properties may well be available; cf. 2.50 below. Of course, on the intersection of any two sets in a representation the two formulas must agree, since both give $f(x)$, or as the case may be, $F(x)$. This, along with the finiteness of the collection of sets, ensures that the function or mapping in question must be continuous relative to D. Also, D must be closed, since polyhedral sets are closed, and the union of finitely many closed sets is closed.

2.48 Exercise (graphs of piecewise linear mappings). *For $F : D \to \mathbb{R}^m$ to be piecewise linear relative to D, it is necessary and sufficient that its graph, the set $G = \{(x, u) \mid x \in D,\ u = F(x)\}$, be expressible as the union of a finite collection of polyhedral sets in $\mathbb{R}^n \times \mathbb{R}^m$.*

Here we're interested primarily in what piecewise linearity means for convex functions. An example of a convex, piecewise linear function is the vector-max function in 1.30 and 2.16. This function f on $\mathbb{R}^n$ has the linear formula $f(x_1, \ldots, x_n) = x_k$ on $C_k := \{(x_1, \ldots, x_n) \mid x_j - x_k \leq 0 \text{ for } j = 1, \ldots, n\}$, and the union of these polyhedral sets C_k is the set $\operatorname{dom} f = \mathbb{R}^n$.

The graph of any piecewise linear function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ must have a representation like that in 2.48 relative to $\operatorname{dom} f$. In the convex case, however, a more convenient representation appears in terms of $\operatorname{epi} f$.

2.49 Theorem (convex piecewise linear functions). *A proper function f is both convex and piecewise linear if and only if $\operatorname{epi} f$ is polyhedral.*

In general for a function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the set $\operatorname{epi} f \subset \mathbb{R}^{n+1}$ is polyhedral if and only if f has a representation of the form

$$f(x) = \begin{cases} \max\{l_1(x), \ldots, l_p(x)\} & \text{when } x \in D, \\ \infty & \text{when } x \notin D, \end{cases}$$

where D is a polyhedral set in $\mathbb{R}^n$ and the functions l_i are affine on $\mathbb{R}^n$; here $p = 0$ is admitted and interpreted as giving $f(x) = -\infty$ when $x \in D$. Any function f of this type is convex and lsc, in particular.

Proof. The case of $f \equiv \infty$ being trivial, we can suppose $\operatorname{dom} f$ and $\operatorname{epi} f$ to be nonempty. Of course when $\operatorname{epi} f$ is polyhedral, $\operatorname{epi} f$ is in particular convex and closed, so that f is convex and lsc (cf. 2.4 and 1.6).

To say that $\operatorname{epi} f$ is polyhedral is to say that $\operatorname{epi} f$ can be expressed as the set of points $(x, \alpha) \in \mathbb{R}^n \times \mathbb{R}$ satisfying a finite system of linear inequalities (cf. 2.10 and the details after it); we can write this system as

$$\gamma_i \geq \langle (c_i, \eta_i), (x, \alpha) \rangle = \langle c_i, x \rangle + \eta_i \alpha \text{ for } i = 1, \ldots, m$$

for certain vectors $c_i \in \mathbb{R}^n$ and scalars $\eta_i \in \mathbb{R}$ and $\gamma_i \in \mathbb{R}$. Because $\operatorname{epi} f \neq \emptyset$, and $(x, \alpha') \in \operatorname{epi} f$ whenever $(x, \alpha) \in \operatorname{epi} f$ and $\alpha' \geq \alpha$, it must be true that $\eta_i \leq 0$ for all i. Arrange the indices so that $\eta_i < 0$ for $i = 1, \ldots, p$ but $\eta_i = 0$ for $i = p + 1, \ldots, m$. (Possibly $p = 0$, i.e., $\eta_i = 0$ for all i.) Taking $b_i = c_i / |\eta_i|$ and $\beta_i = \gamma_i / |\eta_i|$ for $i = 1, \ldots p$, we get $\operatorname{epi} f$ expressed as the set of points

$(x, \alpha) \in I\!R^n \times I\!R$ satisfying

$$\langle b_i, x\rangle - \beta_i \leq \alpha \;\text{ for }\; i = 1, \ldots, p, \quad \langle c_i, x\rangle \leq \gamma_i \;\text{ for }\; i = p+1, \ldots, m.$$

This gives a representation of f of the kind specified in the theorem, with D the polyhedral set in $I\!R^n$ defined by the constraints $\langle c_i, x\rangle \leq \gamma_i$ for $i = p+1, \ldots, m$ and l_i the affine function with formula $\langle b_i, x\rangle - \beta_i$ for $i = 1, \ldots, p$.

Under this kind of representation, if f is proper (i.e., $p \neq 0$), $f(x)$ is given by $\langle b_k, x\rangle - \beta_k$ on the set C_k consisting of the points $x \in D$ such that

$$\langle b_i, x\rangle - \beta_i \leq \langle b_k, x\rangle - \beta_k \;\text{ for }\; i = 1 \ldots, p,\; i \neq k.$$

Clearly C_k is polyhedral and $\bigcup_{k=1}^p C_k = D$, so f is piecewise linear, cf. 2.47.

Suppose now on the other hand that f is proper, convex and piecewise linear. Then $\operatorname{dom} f = \bigcup_{k=1}^r C_k$ with C_k polyhedral, and for each k there exist $a_k \in I\!R^n$ and $\alpha_k \in I\!R$ such that $f(x) = \langle a_k, x\rangle - \alpha_k$ for all $x \in C_k$. Let

$$E_k := \big\{(x, \alpha) \,\big|\, x \in C_k,\; \langle a_k, x\rangle - \alpha_k \leq \alpha < \infty\big\}.$$

Then $\operatorname{epi} f = \bigcup_{k=1}^r E_k$. Each set E_k is polyhedral, since an expression for C_k by a system of linear constraints in $I\!R^n$ readily extends to one for E_k in $I\!R^{n+1}$. The union of the polyhedral sets E_k is the convex set $\operatorname{epi} f$. The fact that $\operatorname{epi} f$ must then itself be polyhedral is asserted by the next lemma, which we state separately for its independent interest. $\square$

2.50 Lemma (polyhedral convexity of unions)**.** *If a convex set C is the union of a finite collection of polyhedral sets C_k, it must itself be polyhedral.*

Moreover if $\operatorname{int} C \neq \emptyset$, the sets C_k with $\operatorname{int} C_k = \emptyset$ are superfluous in the representation. In fact C can then be given a refined expression as the union of a finite collection of polyhedral sets $\{D_j\}_{j\in J}$ such that

(a) *each set D_j is included in one of the sets C_k,*
(b) $\operatorname{int} D_j \neq \emptyset$, *so* $D_j = \operatorname{cl}(\operatorname{int} D_j)$,
(c) $\operatorname{int} D_{j_1} \cap \operatorname{int} D_{j_2} = \emptyset$ *when* $j_1 \neq j_2$.

Proof. Let's represent the sets C_k for $k = 1, \ldots, r$ in terms of a single family of affine functions $l_i(x) = \langle a_i, x\rangle - \alpha_i$ indexed by $i = 1, \ldots, m$: for each k there is a subset I_k of $\{1, \ldots, m\}$ such that $C_k = \big\{x \,\big|\, l_i(x) \leq 0 \;\text{ for all }\; i \in I_k\big\}$. Let I denote the set of indices $i \in \{1, \ldots, m\}$ such that $l_i(x) \leq 0$ for all $x \in C$. We'll prove that

$$C = \big\{x \,\big|\, l_i(x) \leq 0 \;\text{ for all }\; i \in I\big\}.$$

Trivially the set on the right includes C, so our task is to demonstrate that the inclusion cannot be strict. Suppose $\bar{x}$ belongs to the set on the right but not to C. We'll argue this to a contradiction.

For each index $k \in \{1, \ldots, r\}$ let C'_k be the set of points $x \in C$ such that the line segment $[x, \bar{x}]$ meets C_k, this set being closed because C_k is closed. For each $x \in C$, a set which itself is closed (because it is the union of finitely many closed sets C_k), let $K(x)$ denote the set of indices k such that $x \in C'_k$.

Select a point $\hat{x} \in C$ for which the number of indices in $K(\hat{x})$ is as low as possible. Any point x in $C \cap [\hat{x}, \bar{x}]$ has $K(x) \subset K(\hat{x})$ by the definition of $K(x)$ and consequently $K(x) = K(\hat{x})$ by the minimality of $K(\hat{x})$. We can therefore suppose (by moving to the 'last' point that still belongs to C, if necessary) that the segment $[\hat{x}, \bar{x}]$ touches C only at $\hat{x}$.

For any k, the set of $x \in C$ with $k \notin K(x)$ is complementary to C'_k and therefore open relative to C. Hence the set of $x \in C$ satisfying $K(x) \subset K(\hat{x})$ is open relative to C. The minimality of $K(\hat{x})$ forbids strict inclusion, so there has to be a neighborhood V of $\hat{x}$ such that $K(x) = K(\hat{x})$ for all $x \in C \cap V$.

Take any index $k_0 \in K(\hat{x})$. Because $\hat{x}$ is the only point of $[\hat{x}, \bar{x}]$ in C, it must also be the only point of $[\hat{x}, \bar{x}]$ in C_{k_0}, and accordingly there must be an index $i_0 \in I_{k_0}$ such that $l_{i_0}(\hat{x}) = 0 < l_{i_0}(\bar{x})$. For each $x \in C \cap V$ the line segment $[x, \bar{x}]$ meets C_{k_0} and thus contains a point x' satisfying $l_{i_0}(x') \leq 0$ as well as the point $\bar{x}$ satisfying $l_{i_0}(\bar{x}) > 0$, so necessarily $l_{i_0}(x) \leq 0$ (because l_{i_0} is affine). More generally then, for any $x \neq \hat{x}$ in C (but not necessarily in V) the line segment $[x, \hat{x}]$, which lies entirely in C by convexity, contains a point $x^* \neq \hat{x}$ in V, therefore satisfying $l_{i_0}(x^*) \leq 0$. Since $l_{i_0}(\hat{x}) = 0$, this implies $l_{i_0}(x) \leq 0$ (again because l_{i_0} is affine). Thus, i_0 is an index with the property that $C \subset \{x \mid l_{i_0}(x) \leq 0\}$, or in other words, $i_0 \in I$. But since $l_{i_0}(\bar{x}) > 0$ this contradicts the choice of $\bar{x}$ as satisfying $l_i(\bar{x}) \leq 0$ for every $i \in I$.

In passing now to the case where $\operatorname{int} C \neq \emptyset$, there's no loss of generality in supposing that none of the affine functions l_i is a constant function. Then

$$\operatorname{int} C_k = \{x \mid l_i(x) < 0 \text{ for } i \in I_k\}, \qquad \operatorname{int} C = \{x \mid l_i(x) < 0 \text{ for } i \in I\}.$$

Let K_0 denote the set of indices k such that $\operatorname{int} C_k \neq \emptyset$, and K_1 the set of indices k such that $\operatorname{int} C_k = \emptyset$. We wish to show next that $C = \bigcup_{k \in K_0} C_k$. It will be enough to show that $\operatorname{int} C \subset \bigcup_{k \in K_0} C_k$, since $C \supset \bigcup_{k \in K_0} C_k$ and $C = \operatorname{cl}(\operatorname{int} C)$ (cf. 2.33). If $\operatorname{int} C \not\subset \bigcup_{k \in K_0} C_k$, the open set $\operatorname{int} C \setminus \bigcup_{k \in K_0} C_k \neq \emptyset$ would be contained in $\bigcup_{k \in K_1} C_k$, the union of the sets with empty interior. This is impossible for the following reason. If the latter union had nonempty interior, there would be a minimal index set $K \subset K_1$ such that $\operatorname{int} \bigcup_{k \in K} C_k \neq \emptyset$. Then K would have to be more than a singleton, and for any $k^* \in K$ the open set $\left[\operatorname{int} \bigcup_{k \in K} C_k\right] \setminus \bigcup_{k \in K \setminus \{k^*\}} C_k$ would be nonempty and included in C_{k^*}, in contradiction to $\operatorname{int} C_{k^*} = \emptyset$.

To construct a refined representation meeting conditions (a), (b), and (c), consider as index elements j the various partitions of $\{1, \ldots, m\}$; each partition j can be identified with a pair (I_+, I_-), where I_+ and I_- are disjoint subsets of $\{1, \ldots, m\}$ whose union is all of $\{1, \ldots, m\}$. Associate with such each partition j the set D_j consisting of all the points x (if any) such that

$$l_i(x) \leq 0 \text{ for all } i \in I_-, \qquad l_i(x) \geq 0 \text{ for all } i \in I_+.$$

Let J_0 be the index set consisting of the partitions $j = (I_+, I_-)$ such that $I_- \supset I_k$ for some k, so that $D_j \subset C_k$. Since every $x \in C$ belongs to at least one C_k, it also belongs to a set D_j for at least one $j \in J_0$. Thus, C is the union of the sets D_j for $j \in J_0$. Hence also, by the argument given earlier, it is the union of the ones with $\operatorname{int} D_j \neq \emptyset$, the partitions j with this property comprising a possibly smaller index set J within J_0. For each $j \in J$ the nonempty set $\operatorname{int} D_j$ consists of the points x satisfying

$$l_i(x) < 0 \text{ for all } i \in I_-, \qquad l_i(x) > 0 \text{ for all } i \in I_+.$$

For any two different partitions j_1 and j_2 in J there must be some $i^* \in \{1, \ldots, m\}$ such that one of the sets D_{j_1} and D_{j_2} is contained in the open half-space $\{x \mid l_{i^*}(x) < 0\}$, while the other is contained in the open half-space $\{x \mid l_{i^*}(x) > 0\}$. Hence $\operatorname{int} D_{j_1} \cap \operatorname{int} D_{j_2} = \emptyset$ when $j_1 \neq j_2$. The collection $\{D_j\}_{j \in J}$ therefore meets all the stipulations. □

J.* Other Examples

For any convex function f, the sets $\operatorname{lev}_{\leq \alpha} f$ are convex, as seen in 2.7. But a nonconvex function can have that property as well; e.g. $f(x) = \sqrt{|x|}$.

2.51 Exercise (functions with convex level sets). *For a differentiable function $f : I\!R^n \to I\!R$, the sets $\operatorname{lev}_{\leq \alpha} f$ are convex if and only if*

$$\langle \nabla f(x_0), x_0 - x_1 \rangle \geq 0 \text{ whenever } f(x_1) \leq f(x_0).$$

Furthermore, if such a function f is twice differentiable at $\bar{x}$ with $\nabla f(\bar{x}) = 0$, then $\nabla^2 f(\bar{x})$ must be positive-semidefinite.

The convexity of the log-exponential function in 2.16 is basic to the treatment of a much larger class of functions often occurring in applications, especially in problems of engineering design. These functions are described next.

2.52 Example (posynomials). *A* posynomial *in the variables $y_1, \ldots, y_n$ is an expression of the form*

$$g(y_1, \ldots, y_n) = \sum\nolimits_{i=1}^{r} c_i y_1^{a_{i1}} y_2^{a_{i2}} \cdots y_n^{a_{in}}$$

where (1) only positive *values of the variables y_j are admitted, (2) all the coefficients c_i are* positive, *but (3) the exponents a_{ij} can be* arbitrary real numbers. *Such a function may be far from convex, yet convexity can be achieved through a logarithmic change of variables. Setting $b_i = \log c_i$ and*

$$f(x_1, \ldots, x_n) = \log g(y_1, \ldots, y_n) \text{ for } x_j = \log y_j,$$

one obtains a convex function $f(x) = \operatorname{logexp}(Ax + b)$ defined for all $x \in I\!R^n$, where b is the vector in $I\!R^r$ with components b_i, and A is the matrix in $I\!R^{r \times n}$ with components a_{ij}. Still more generally, any function

$$g(y) = g_1(y)^{\lambda_1} \cdots g_p(y)^{\lambda_p} \text{ with } g_k \text{ posynomial}, \ \lambda_k > 0,$$

can be converted by the logarithmic change of variables into a convex function of form $f(x) = \sum_{k=1}^{p} \lambda_k \operatorname{logexp}(A_k x + b_k)$.

Detail. The convexity of f follows from that of logexp (in 2.16) and the composition rule in 2.20(a). The integer r is called the *rank* of the posynomial. When $r = 1$, f is merely an affine function on $\mathbb{R}^n$. □

The wide scope of example 2.52 is readily appreciated when one considers for instance that the following expression is a posynomial:

$$g(y_1, y_2, y_3) = 5y_1^7/y_3^4 + \sqrt{y_2 y_3} = 5y_1^7 y_2^0 y_3^{-4} + y_1^0 y_2^{1/2} y_3^{1/2} \quad \text{for} \quad y_i > 0.$$

The next example illustrates the approach one can take in using the derivative conditions in 2.14 to verify the convexity of a function defined on more than just an open set.

2.53 Example (weighted geometric means)**.** *For any choice of weights* $\lambda_j > 0$ *satisfying* $\lambda_1 + \cdots + \lambda_n \leq 1$ *(for instance* $\lambda_j = 1/n$ *for all* j*), one gets a proper, lsc, convex function* f *on* $\mathbb{R}^n$ *by defining*

$$f(x) = \begin{cases} -x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n} & \text{for } x = (x_1, \ldots, x_n) \text{ with } x_j \geq 0, \\ \infty & \text{otherwise.} \end{cases}$$

Detail. Clearly f is finite and continuous relative to $\operatorname{dom} f$, which is nonempty, closed, and convex. Hence f is proper and lsc. On $\operatorname{int}(\operatorname{dom} f)$, which consists of the vectors $x = (x_1, \ldots, x_n)$ with $x_k > 0$, the convexity of f can be verified through condition 2.14(c) and the calculation that

$$\langle z, \nabla^2 f(x) z \rangle = |f(x)| \left[\sum_{j=1}^{n} \lambda_j (z_j/x_j)^2 - \left(\sum_{j=1}^{n} \lambda_j (z_j/x_j) \right)^2 \right];$$

this expression is nonnegative by Jensen's inequality 2.2(b) as applied to the function $\theta(t) = t^2$. (Specifically, $\theta\big(\sum_{j=0}^{n} \lambda_j t_j\big) \leq \sum_{j=0}^{n} \lambda_j \theta(t_j)$ in the case of $t_j = z_j/x_j$ for $j = 1, \ldots, n$, $t_0 = 0$, $\lambda_0 = 1 - \sum_{j=1}^{n} \lambda_j$.) The convexity of f relative to all of $\operatorname{dom} f$ rather than merely $\operatorname{int}(\operatorname{dom} f)$, is obtained then by taking limits in the convexity inequality. □

A number of interesting convexity properties hold for spaces of matrices. The space $\mathbb{R}^{n \times n}$ of all square matrices of order n is conveniently treated in terms of the inner product

$$\big\langle A, B \big\rangle := \sum_{i,j=1}^{n,n} a_{ij} b_{ji} = \operatorname{tr} AB, \tag{2(12)}$$

where $\operatorname{tr} C$ denotes the *trace* of a matrix $C \in \mathbb{R}^{n \times n}$, which is the sum of the diagonal elements of C. (The rule that $\operatorname{tr} AB = \operatorname{tr} BA$ is obvious from this inner product interpretation.) Especially important is

$$\mathbb{R}^{n \times n}_{\text{sym}} := \text{ space of symmetric real matrices of order } n, \tag{2(13)}$$

which is a linear subspace of $\mathbb{R}^{n\times n}$ having dimension $n(n+1)/2$. This can be treated as a Euclidean vector space in its own right relative to the trace inner product 2(12). In $\mathbb{R}^{n\times n}_{\rm sym}$, the positive-semidefinite matrices form a closed, convex set whose interior consists of the positive-definite matrices.

2.54 Exercise (eigenvalue functions)**.** *For each matrix $A \in \mathbb{R}^{n\times n}_{\rm sym}$ denote by $\operatorname{eig} A = (\lambda_1, \ldots, \lambda_n)$ the vector of eigenvalues of A in descending order (with eigenvalues repeated according to their multiplicity). Then for $k = 1, \ldots, n$ one has the convexity on $\mathbb{R}^{n\times n}_{\rm sym}$ of the function*

$$\Lambda_k(A) := \text{ sum of the first } k \text{ components of } \operatorname{eig} A.$$

Guide. Show that $\Lambda_k(A) = \max_{P\in\mathcal{P}_k} \operatorname{tr}[PAP]$, where $\mathcal{P}_k$ is the set of all matrices $P \in \mathbb{R}^{n\times n}_{\rm sym}$ such that P has rank k and $P^2 = P$. (These matrices are the orthogonal projections of $\mathbb{R}^n$ onto its linear subspaces of dimension k.) Argue in terms of diagonalization. Note that $\operatorname{tr}[PAP] = \langle A, P\rangle$. □

2.55 Exercise (matrix inversion as a gradient mapping)**.** *On $\mathbb{R}^{n\times n}_{\rm sym}$, the function*

$$j(A) := \begin{cases} \log(\det A) & \text{if } A \text{ is positive-definite,} \\ -\infty & \text{if } A \text{ is not positive-definite,} \end{cases}$$

is concave and usc, and, where finite, differentiable with gradient $\nabla j(A) = A^{-1}$.

Guide. Verify first that for any symmetric, positive-definite matrix C one has $\operatorname{tr} C - \log(\det C) \geq n$, with strict inequality unless $C = I$; for this consider a diagonalization of C. Next look at arbitrary symmetric, positive-definite matrices A and B, and by taking $C = B^{1/2}AB^{1/2}$ establish that

$$\log(\det A) + \log(\det B) \leq \big\langle A, B\big\rangle - n$$

with equality holding if and only if $B = A^{-1}$. Show that in fact

$$j(A) = \inf_{B\in\mathbb{R}^{n\times n}_{\rm sym}} \big\{\big\langle A, B\big\rangle - j(B) - n\big\} \text{ for all } A \in \mathbb{R}^{n\times n}_{\rm sym},$$

where the infimum is attained if and only if A is positive-definite and $B = A^{-1}$. Using the differentiability of j at such A, along with the inequality $j(A') \leq \big\langle A', A^{-1}\big\rangle - j(A^{-1}) - n$ for all A', deduce then that $\nabla j(A) = A^{-1}$. □

The inversion mapping for symmetric matrices has still other remarkable properties with respect to convexity. To describe them, we'll use for matrices $A, B \in \mathbb{R}^{n\times n}_{\rm sym}$ the notation

$$A \succeq B \iff A - B \text{ is positive-semidefinite.} \qquad 2(14)$$

2.56 Proposition (convexity property of matrix inversion)**.** *Let P be the open, convex subset of $\mathbb{R}^{n\times n}_{\rm sym}$ consisting of the positive-definite matrices, and let $J : P \to P$ be the inverse-matrix mapping: $J(A) = A^{-1}$. Then J is convex with respect to $\succeq$ in the sense that*

$$J\big((1-\lambda)A_0 + \lambda A_1\big) \preceq (1-\lambda)J(A_0) + \lambda J(A_1)$$

for all A_0, $A_1 \in P$ *and* $\lambda \in (0,1)$. *In addition,* J *is order-inverting:*

$$A_0 \preceq A_1 \iff A_0^{-1} \succeq A_1^{-1}.$$

Proof. We know from linear algebra that any two positive-definite matrices A_0 and A_1 in $I\!R^{n\times n}_{\rm sym}$ can be diagonalized simultaneously through some choice of a basis for $I\!R^n$. We can therefore suppose without loss of generality that A_0 and A_1 are diagonal, in which case the assertions reduce to properties that are to hold component by component along the diagonal, i.e. in terms of the eigenvalues of the two matrices. We come down then to the one-dimensional case: for the mapping $J : (0,\infty) \to (0,\infty)$ defined by $J(\alpha) = \alpha^{-1}$, is J convex and order-inverting? The answer is evidently yes. □

2.57 Exercise (convex functions of positive-definite matrices)**.** *On the convex subset* P *of* $I\!R^{n\times n}_{\rm sym}$ *consisting of the positive-definite matrices, the relations*

$$f(A) = g(A^{-1}), \qquad g(B) = f(B^{-1}),$$

give a one-to-one correspondence between the convex functions $f : P \to \overline{I\!R}$ *that are nondecreasing (with* $f(A_0) \le f(A_1)$ *if* $A_0 \preceq A_1$*) and the convex functions* $g : P \to \overline{I\!R}$ *that are nonincreasing (with* $g(B_0) \ge g(B_1)$ *if* $B_0 \preceq B_1$*). In particular, for any subset* S *of* $I\!R^{n\times n}_{\rm sym}$ *the function*

$$g(B) = \sup_{C\in S} \operatorname{tr}\big[CB^{-1}C^*\big] = \sup_{C\in S} \big\langle C^*C, B^{-1}\big\rangle \qquad 2(15)$$

is convex on P *and nonincreasing.*

Guide. This relies on the properties of the inversion mapping J in 2.56, composing it with f and g in analogy with 2.20(c). The example at the end corresponds to $f(A) = \sup_{C\in S} \operatorname{tr}[CAC^*]$. □

Functions of the kind in 2(15) are important in theoretical statistics. As a special case of 2.57, we see from 2.54 that the function $g(B) = \Lambda_k(B^{-1})$, giving the sum of the first k eigenvalues of B^{-1}, is convex and nonincreasing on the set of positive-definite matrices $B \in I\!R^{n\times n}_{\rm sym}$. This corresponds to taking S to be the set of all projection matrices of rank k.

Commentary

The role of convexity in inspiring many of the fundamental strategies and ideas of variational analysis, such as the emphasis on extended-real-valued functions and the geometry of their epigraphs, has been described at the end of Chapter 1. The lecture notes of Fenchel [1951] were instrumental, as were the cited works of Moreau and Rockafellar that followed.

Extensive references to early developments in convex analysis and their historical background can be found in the book of Rockafellar [1970a]. The two-volume opus