CONVEX FUNCTIONS, AN ELEMENTARY APPROACH

August 23, 2018

JON JOHNSEN

Abstract. This note collects some basic facts on convex functions in one or several variables, with elementary proofs.

1. Convex sets

In these notes, $\mathbb{R}^n$ denotes the Euclidean space of dimension $n \geq 1$.

1.1. **Basics.**

By definition, a subset $U \subset \mathbb{R}^n$ is said to be *convex*, if for any two points $x, y \in U$ the line segment connecting x and y is contained in U, which means that

$$x + \theta(y - x) \in U \quad \text{for all} \quad \theta \in [0,1]. \tag{1.1}$$

Equivalently this can be written as

$$(1-\theta)x + \theta y \in U \quad \text{for all} \quad \theta \in [0,1]. \tag{1.2}$$

Or, in an even more symmetric manner, that

$$\begin{gathered} \theta_0 x + \theta_1 y \in U \quad \text{when} \\ \theta_0 + \theta_1 = 1,\ \theta_0 \geq 0,\ \theta_1 \geq 0. \end{gathered} \tag{1.3}$$

Anyhow, it is enough to check one of these conditions for $x \neq y$, and to do so *eg* for $0 < \theta < 1$, for otherwise there is nothing to show.

As examples of convex sets, there are open and closed balls $B(x,r)$ and $\bar{B}(x,r)$; and in $\mathbb{R}^3$ ellipsoids such as

$$(\frac{x}{a})^2 + (\frac{y}{b})^2 + (\frac{z}{c})^2 \leq 1 \tag{1.4}$$

and other barrel-shaped objects.

By definition a set $K \subset \mathbb{R}^n$ is said to be *concave* when its complement $\mathbb{R}^n \setminus K$ is a convex set. The unit sphere

$$\mathbb{S}^{n-1} = \left\{\, x \in \mathbb{R}^n \;\middle|\; \|x\| = 1 \,\right\} \tag{1.5}$$

is neither, but its complement $\mathbb{R}^n \setminus \mathbb{S}^{n-1}$ has two components of which the unbounded is concave.

1.2. **Elementary results.** From the definition above it follows at once that

$$\forall i \in I\colon U_i \text{ is convex } \implies \bigcap_{i\in I} U_i \text{ is convex.} \tag{1.6}$$

One can now introduce the *convex hull*, written $\operatorname{ch}(A)$, of an arbitrary set $A \subset \mathbb{R}^n$ as

$$\operatorname{ch}(A) = \bigcap \Big\{\, U \subset \mathbb{R}^n \;\Big|\; U \text{ is convex },\; A \subset U \,\Big\}. \tag{1.7}$$

This is by (1.6) convex and $A \subset \operatorname{ch}(A)$; therefore $\operatorname{ch}(A)$ is one of the sets in the intersection, hence the *smallest* convex subset containing A.

As a preparation it is recalled that the boundary of a set $M \subset \mathbb{R}^n$ is defined in terms of its interior M° and exterior $M^e = \mathbb{R}^n \setminus \bar{M}$ as

$$\partial M = \mathbb{R}^n \setminus (M^\circ \cup M^e). \tag{1.8}$$

It follows at once that $\partial M = \bar{M} \setminus M^\circ$. In general it holds true that

$$\partial(\bar{M}) \subset \partial M. \tag{1.9}$$

For convex sets there is always equality here, that is $\partial(\bar{U}) = \partial U$. This can be shown by using the next result (*cf* Exercise 1.1).

As expected, every point in the closure of a convex set can always be reached along a line segment in the interior U°, if this is non-empty:

Lemma 1.1. *Let U be convex with given points $x \in U^\circ$ and $y \in \bar{U}$. Then $(1-\theta)x + \theta y$ belongs to U° whenever $0 \le \theta < 1$.*

Proof. For such x, y, θ, it suffices to show that, when $\varepsilon > 0$ is suitably chosen, then $v = (1-\theta)x + \theta y + z$ belongs to U for all $z \in B(0, \varepsilon)$. Now

$$v = (1-\theta)(x + (1-\theta)^{-1} z) + \theta y, \tag{1.10}$$

where the vector in parenthesis is contained in U° for $\varepsilon > 0$ so small that $B(x, \frac{2\varepsilon}{1-\theta}) \subset U$. So in case $y \in U$ convexity yields $v \in U$.

For general $y \in \bar{U}$ there is some $w \in B(y, \varepsilon) \cap U$, and one may in (1.10) replace θy by θw and z by $z + \theta(y - w)$. Since the latter belongs to $B(0, 2\varepsilon)$, the first part of the proof yields $v \in U$. □

Using the lemma, one finds the next result (*cf* Exercise 1.2), which shows that convexity plays well together with topological notions:

Proposition 1.2. *If $U \subset \mathbb{R}^n$ is convex, so are U° and $\bar{U}$.*

Furthermore, a convex set U cannot have isolated points, unless it is a singleton; *ie* unless $U = \{x\}$ for some $x \in \mathbb{R}^n$.

1.3. **Enveloping halfspaces.** As a remarkable general property, it is observed that convex sets share the following property:

A non-trivial convex set U is contained in a half-space.

Whilst this is geometrically obvious (U lies "on one side" of each boundary point), a formal proof will be given below.

Recall first that the plane P in $\mathbb{R}^3$ through the point $x^0 = (x^0_1, x^0_2, x^0_3)$, and having the normal vector $\vec{n} = (a, b, c)$, consists of the points $x = (x_1, x_2, x_3)$ in $\mathbb{R}^3$ that fulfil the equation $\vec{n} \cdot (x - x^0) = 0$; that is,

$$a(x_1 - x^0_1) + b(x_2 - x^0_2) + c(x_3 - x^0_3) = 0. \tag{1.11}$$

The half-space $H_{x_0,\vec{n}}$ associated with P is therefore introduced as the set of points $x = (x_1, x_2, x_3)$ lying on the positive side of P in the sense that $\vec{n} \cdot (x - x^0) \geq 0$. More precisely,

$$H_{x_0,\vec{n}} = \left\{ (x_1, x_2, x_3) \in \mathbb{R}^3 \;\middle|\; a(x_1 - x^0_1) + b(x_2 - x^0_2) + c(x_3 - x^0_3) \geq 0 \right\}. \tag{1.12}$$

In analogy, a *half-space* is in $\mathbb{R}^n$, $n \geq 1$, defined in terms of a point $x_0 \in \mathbb{R}^n$ and a normal vector $\vec{n} \in \mathbb{R}^n$ as

$$H_{x_0,\vec{n}} = \left\{ x \in \mathbb{R}^n \;\middle|\; \vec{n} \cdot (x - x_0) \geq 0 \right\}. \tag{1.13}$$

Another preparation concerns the well-known distance from a point y to a given set $M \subset \mathbb{R}^n$:

$$d(y, M) = \inf\{\, d(y, x) \mid x \in M \,\}. \tag{1.14}$$

Lemma 1.3. *If a point $y_0 \in \mathbb{R}^n$ lies outside a closed subset F, then the distance $d(y_0, F)$ is a minimum attained at some x_0 in F.* Ie *there exists some $x_0 \in F$ such that*

$$\|x_0 - y_0\| \leq \|x - y_0\| \quad \textit{for all } x \in F. \tag{1.15}$$

Moreover, x_0 belongs to the boundary of F; that is, $x_0 \in F \setminus F^\circ$.

Proof. If F is compact continuity of $x \mapsto \|x - y_0\|$ yields the existence of x_0. Else $F \cap \bar{B}(y_0, N)$ is compact, say for $N = 2d(y_0, F)$, and has a point x_0 of minimal distance.

Had x_0 been an inner point, say $B(x_0, \varepsilon) \subset F$, then the function $t \mapsto y_0 + t(x_0 - y_0) =: x_t$ would have an open preimage of $B(x_0, \varepsilon)$, so that for some $\delta > 0$ it would hold that $x_t \in F$ for all $t \in \,]1 - \delta, 1 + \delta[$, whence $\|x_t - y_0\| = t\|x_0 - y_0\|$ would not attain a minimum at $t = 1$. $\square$

Now it is easy to verify a weak version of the claim on the half-space:

Proposition 1.4. *When a convex subset $U \subset \mathbb{R}^n$ is proper, then U is contained in a closed half-space H.*

Proof. Convexity and the property that $U \neq \mathbb{R}^n$ are both inherited by $\bar{U}$ (*cf* Exercise 1.3). Hence U may be assumed closed.

By assumption a point $y_0 \in \mathbb{R}^n \setminus U$ may be fixed; since U is closed, $d(y_0, U)$ is attained at some $x_0 \in U$ by Lemma 1.3. Writing $x = x_0 + z$ for arbitrary $x \in U$, and squaring the inequality of the lemma,

$$0 \leq \|z\|^2 + 2z \cdot (x_0 - y_0). \tag{1.16}$$

Since U is convex, every $x_0 + \theta z$ with $\theta \in [0,1]$ belongs to U, which by substitution into the above (as we may) gives

$$0 \leq \theta \cdot (\theta \|z\|^2 + 2z \cdot (x_0 - y_0)). \tag{1.17}$$

Because the factor in parenthesis after θ must remain non-negative for $\theta \to 0^+$, this implies $z \cdot (x_0 - y_0) \geq 0$. Setting $\vec{n} = x_0 - y_0$ this means that $(x - x_0) \cdot \vec{n} \geq 0$ for all $x \in U$, thence the inclusion $U \subset H_{x_0, \vec{n}}$. □

The proof above gave more than stated, for the point x_0 entering the half-space $H_{x_0,\vec{n}}$ is obviously a boundary point of $H_{x_0,\vec{n}}$ as well as in ∂U according to Lemma 1.3.

In fact, $H_{x_0,\vec{n}}$ can be so chosen, that x_0 equals any given boundary point of U. This sharpening will now be proved using a compactness argument.

Theorem 1.5. *For every boundary point x_0 of a convex set $U \subset \mathbb{R}^n$ there is an inclusion*

$$U \subset H \tag{1.18}$$

into the closed halfspace $H = H_{x_0,\vec{n}}$ determined by x_0 and a suitable unit vector $\vec{n}$.

Proof. Since x_0 is also a boundary point of $\bar{U}$, *cf* (1.9), there is for each $k \in N$ a point $y_k \in B(x_0, 1/k) \setminus \bar{U}$.

From the remarks after the proof of Proposition 1.4, it is seen that each distance $d(y_k, \bar{U})$ is attained at a point $x_k \in \partial(\bar{U})$. These x_k converge to x_0 since

$$0 \leq \|x_k - x_0\| \leq \|x_k - y_k\| + \|y_k - x_0\| \leq 2\|y_k - x_0\| \searrow 0. \tag{1.19}$$

Associated with these points there are (as $x_k \neq y_k$) unit vectors

$$\vec{n}_k = \frac{x_k - y_k}{\|x_k - y_k\|} \in \mathbb{S}^{n-1}. \tag{1.20}$$

Since $\vec{n}_k \in \mathbb{S}^{n-1}$ for each k (*cf* (1.5)), it may by the compactness of $\mathbb{S}^{n-1}$ be assumed that for a suitable $\vec{n}_0 \in \mathbb{S}^{n-1}$ there is convergence

$$\vec{n}_k \to \vec{n}_0 \quad \text{for } k \to \infty. \tag{1.21}$$

(Otherwise one may extract a convergent subsequence $\vec{n}_{k_m}$ that along with x_{k_m} can replace $\vec{n}_k$, x_k in the rest of the proof.)

Now it only remains to show that the half-space $H_{x_0,\vec{n}_0}$ contains U as claimed. However, by the construction of $\vec{n}_k$ it is seen from the previous proof that $U \subset H_{x_k,\vec{n}_k}$. So for any given $x \in U$,

$$0 \leq \vec{n}_k \cdot (x - x_k) \qquad \text{for all } k \in \mathbb{N}. \tag{1.22}$$

When combined with (1.19) and (1.21), continuity of the inner product therefore yields $0 \le \vec{n}_0 \cdot (x - x_0)$, as was to be shown. $\square$

The normal vector $\vec{n}_0$ of the halfspace H in the theorem is in general not uniquely determined. This is geometrically obvious, *eg* by taking for U a square in $\mathbb{R}^2$ and x_0 as a cornerpoint. In the proof above it is also clear that $(\vec{n}_k)$ might have several accumulation points.

However, one can of course let $N(x_0)$ denote the set of unit vectors $\vec{n}$ for which $U \subset H_{x_0,\vec{n}}$ holds true. Then one has the following result for a *closed* convex set U:

$$U = \bigcap_{x_0 \in \partial U} \bigcap_{\vec{n} \in N(x_0)} H_{x_0,\vec{n}}. \tag{1.23}$$

Indeed, the inclusion from the left to the right is obvious; and it cannot be strict, for when y_0 belongs to $\mathbb{R}^n \setminus U$, one may as in the proof of Proposition 1.4 minimise the distance from y_0 to U at some point $x_0 \in \partial U$, which gives $U \subset H_{x_0,\vec{n}}$ for $\vec{n}$ equal to $(x_0 - y_0)/\|x_0 - y_0\|$, so that $(y_0 - x_0) \cdot \vec{n} = -\|x_0 - y_0\| < 0$ implies $y_0 \notin H_{x_0,\vec{n}}$.

Because of (1.23), a closed convex set U is always *enveloped* in the halfspaces containing it.

Exercise 1.1. Show that the inclusion $\partial(\bar{M}) \subset \partial M$ may be strict. Verify that equality holds for convex M. (*Hint:* reduce to the case $M^\circ \neq \emptyset$ and obtain a contradiction by using Lemma 1.1 twice.)

Exercise 1.2. Give a proof of Proposition 1.2. (*Hint:* use Lemma 1.1.)

Exercise 1.3. Show that if U is convex and $U \neq \mathbb{R}^n$, then $\bar{U} \neq \mathbb{R}^n$. (*Hint:* Prolong the line segment in Lemma 1.1, and twist it).

Exercise 1.4. Find a set K which does not fulfil (1.23).

2. Convex and concave functions

Let $U \subset \mathbb{R}^n$ denote a fixed convex set in the following. The main subjects of these lecture notes are collected in

Definition 2.1. 1° A function $f \colon U \to \mathbb{R}$ is *convex* if for all x, $y \in U$ and $0 \le \theta \le 1$,

$$f((1-\theta)x + \theta y) \le (1-\theta)f(x) + \theta f(y). \tag{2.1}$$

If the inequality is strict when $x \neq y$ and $0 < \theta < 1$, then f said to be *strictly* convex.

2° f is called (strictly) *concave*, if $-f$ is (strictly) convex.

Obviously (2.1) is equivalent to

$$f(x + \theta(y - x)) \le f(x) + \theta(f(y) - f(x)). \tag{2.2}$$

Again it suffices to check this only for $x \neq y$ and for $0 < \theta < 1$.

As examples, it is clear that linear maps and affine maps $\mathbb{R}^n \to \mathbb{R}$ are both convex and concave, as (2.1) holds with equality. Consequently such functions are never strictly convex/concave.

Geometrically condition (2.1) means for $n = 1$ or $n = 2$ that the line segment $L_{x,y}$ that joins two arbitrary points $(x, f(x))$ and $(y, f(y))$ on the graph of f lies entirely above the graph. Indeed, this segment consists of the points

$$\begin{bmatrix} x \\ f(x) \end{bmatrix} + \theta \begin{bmatrix} y - x \\ f(y) - f(x) \end{bmatrix} \qquad \text{for } 0 \le \theta \le 1, \tag{2.3}$$

each of which is above the corresponding point on the graph of $f(x)$, *ie* above $(x + \theta(y - x), f(x + \theta(y - x)))$, because of (2.2).

For general $n \ge 1$ it is more meaningful to say that $L_{x,y}$ is contained in the so-called *epigraph* of f:

$$E(f) = \{\, (x, y) \in U \times \mathbb{R} \;\big|\; y \ge f(x) \,\}. \tag{2.4}$$

Using the definitions, one finds a classical characterisation of convexity:

Proposition 2.2. *A function $f\colon U \to \mathbb{R}$ is convex if, and only if, its epigraph $E(f)$ is a convex set.*

This geometric description shows that $|\cdot|$, $(\cdot)^2$, exp and $\exp(-\cdot)$ are convex on $\mathbb{R}$; while log and $\sqrt{\cdot}$ are concave on $]0,\infty[$ (the latter even on $[0,\infty[$). Whereas cos and sin are neither.

For later reference, it is convenient to observe that in dimension $n \ge 2$ a function $f\colon U \to \mathbb{R}$ is convex if and only if it holds for all $x, y \in U$ that the auxiliary function

$$g_{x,y}(t) = f(x + t(y - x)) \tag{2.5}$$

is convex for $t \in [0, 1]$, *ie* is convex as a function of one variable.

Exercise 2.1. Find the subintervals of $\mathbb{R}$ on which sin is convex.

Exercise 2.2. Deduce that when $f\colon I \to \mathbb{R}$ is convex for some *interval* $I \subset \mathbb{R}$, and when $x < y$ in (2.1), then one has for the intermediate point $z = (1 - \theta)x + \theta y$ that $\theta = \frac{z-x}{y-x}$ and $1 - \theta = \frac{y-z}{y-x}$, hence

$$f(z) \le \frac{y - z}{y - x} f(x) + \frac{z - x}{y - x} f(y) \qquad \text{for } x < z < y. \tag{2.6}$$

Conversely, does this property imply that f is convex on I ?

Exercise 2.3. Prove that when $g\colon \mathbb{R} \to \mathbb{R}$ is monotone increasing and convex, then $g \circ f$ is convex on $U \subset \mathbb{R}^n$ if f is so.

When will concavity of f carry over to $g \circ f$?

Exercise 2.4. Show that if f is strictly convex with a local minimum at $x \in U$, then $f(y) > f(x)$ for every $y \ne x$; ie x is uniquely determined.

Exercise 2.5. Suppose $f\colon [a,b] \to \mathbb{R}$ is convex on the open interval $]a,b[$ and continuous at $x=a$ and $x=b$. Show that f is convex on the closed interval $[a,b]$. Proceed to prove the same result for strict convexity. (*Hint:* Consider first the endpoint $y=b$, and use monotonicity of the slope function in (3.3) below.)

Exercise 2.6. Sketch $E(f)$ and explain why convexity of f follows if $E(f)$ is convex. Then deduce the converse; ie prove Proposition 2.2.

Exercise 2.7. Show that if φ is convex on U, then $\varphi(\theta_0 x + \theta_1 y) \leq \theta_0\varphi(x) + \theta_1\varphi(y)$ when $\theta_0, \theta_1 \geq 0$ fulfil $\theta_0 + \theta_1 = 1$. Generalise this to finite families $x_\nu \in U$ and numbers $a_\nu \geq 0$:

$$\varphi\left(\frac{\sum a_\nu x_\nu}{\sum a_\nu}\right) \leqq \frac{\sum a_\nu \varphi(x_\nu)}{\sum a_\nu}. \tag{2.7}$$

This is *J.L.W.V. Jensen's inequality* for convex functions (1906)—for decades used in the letter head of the Mathematics Department of Copenhagen University.

3. Geometry of Convex Epigraphs

3.1. **The one-dimensional case.** When graphing a convex function in dimension $n=1$, one expects at once that the slope of the secant between two points on the graph is increasing, as expressed by the inequality (3.1) below.

Indeed, this property is a classical characterisation of convex functions:

Proposition 3.1. *For a function $f\colon I \to \mathbb{R}$ on an interval $I \subset \mathbb{R}$, convexity is equivalent to the property that*

$$\frac{f(x)-f(y)}{x-y} \leq \frac{f(z)-f(x)}{z-x} \quad \textit{whenever} \quad y < x < z. \tag{3.1}$$

The strict inequality holds for all $y < x < z$ if and only f is strictly convex.

Proof. For elements $y < x < z$ in I it is clear that

$$\begin{aligned} &\frac{f(x)-f(y)}{x-y} \leq \frac{f(z)-f(x)}{z-x} \\ &\qquad\iff (y-z)f(x) \geq (x-z)f(y) + (y-x)f(z) \qquad (3.2)\\ &\qquad\iff f(x) \leq \frac{z-x}{z-y}f(y) + \frac{x-y}{z-y}f(z). \end{aligned}$$

The last inequality follows from convexity of f, since $\theta_0 := \frac{z-x}{z-y} > 0$, $\theta_1 := \frac{x-y}{z-y} > 0$, $\theta_0 + \theta_1 = 1$ and $x = \frac{z-x}{z-y}y + \frac{x-y}{z-y}z$. Conversely the last inequality also implies convexity. Sharp inequalities above are similarly seen to be equivalent to strict convexity. $\square$

In general it is convenient to introduce the slope function

$$S(x,y) = \frac{f(y)-f(x)}{y-x} \quad \text{for } x \neq y. \tag{3.3}$$

This is clearly symmetric, that is $S(x,y) = S(y,x)$, so for example it is an increasing function of x if and only if it is increasing with respect to y. In fact, its monotonicity characterises the convexity etc. of f:

Proposition 3.2. *A given function $f\colon I \to \mathbb{R}$ on an interval $I \subset \mathbb{R}$ is convex (respectively strictly convex) if and only if the slope function S is (strictly) monotone increasing in one of its arguments.*

Proof. By the symmetry of S, it is enough to show that convexity of f is equivalent to the property that $S(x,y) \le S(x,z)$ for $y \le z$ and $x \notin \{y,z\}$. The case $y < x < z$ was covered in the proof of Proposition 3.1. The two other cases, ie $x < y < z$ and $y < z < x$, can be treated analogously (do it!). □

The well-known fact that convex functions always are *left-* and *right-*differentiable can be shown by a nice application of Proposition 3.1–3.2.

Indeed, the left-hand side of (3.1) equals $S(x,y)$ hence is an increasing function of y; denoted $S(x,\cdot)$. Since $S(x,\cdot)$ is bounded from above by the right-hand side of (3.1), its supremum is necessarily a limit for $y \to x^-$. By definition this limit is the left-derivative $f'_-(x)$ of f at x, and for $z > x$ it fulfils

$$f'_-(x) \le \frac{f(z)-f(x)}{z-x}. \tag{3.4}$$

Repeating the argument, it follows from this inequality that the right-hand side has (its infimum as) a limit $f'_+(x)$ for $z \to x^+$ and that

$$f'_-(x) \le f'_+(x). \tag{3.5}$$

Therefore we have proved

Proposition 3.3. *A convex function $f\colon I \to \mathbb{R}$ on an interval $I \subset \mathbb{R}$ is differentiable both from the left and the right at every interior point x in I and* (3.5) *holds.*

As an addendum to this proposition, it also follows from the proof of the existence of the one-sided derivatives that

$$f(z) \ge f(x) + f'_+(x)(z-x) \qquad \text{for all } z \ge x \text{ in } I, \tag{3.6}$$

$$f(y) \ge f(x) + f'_-(x)(y-x) \qquad \text{for all } y \le x \text{ in } I. \tag{3.7}$$

Moreover, since $(f(z)-f(x))/(z-x)$ equals $f'_+(x)+o(1)$, and similarly for $f'_-(y)$,

$$f(z) = f(x) + f'_+(x)(z-x) + o(z-x) \qquad \text{for } x \le z \tag{3.8}$$

$$f(y) = f(x) + f'_-(x)(y-x) + o(y-x) \qquad \text{for } y \le x. \tag{3.9}$$

Therefore the graph of f has left and right *half-tangents* $T_\pm$ (at every inner point x), namely the graphs of

$$y \mapsto f(x) + f'_\pm(x)(y-x) \quad \text{for } y \gtrless x. \tag{3.10}$$

The geometric meaning of (3.6)–(3.7) is then that the graph of f lies entirely above T_- and T_+. *Cf* examples like $e^{|x|}$.

To combine (3.6)–(3.7) into a single inequality valid for all $y \in I$, we may multiply (3.5) by $y - x \geq 0$ and by $y - x < 0$, respectively, to see that both $a = f'_+(x)$ and $a = f'_-(x)$ fulfil that

$$f(y) \geq f(x) + a(y-x) \qquad \text{for } every\ y \in I. \tag{3.11}$$

Because of the inequality here, every $a \in \mathbb{R}$ with this property is called a *subgradient* of f at x; cf. Section 3.2.

Whenever a function f has a subgradient at x, i.e. fulfils (3.11) for some a, then f is called *subdifferentiable* at x. In the affirmative case, the set of all the possible subgradients a of f at x is usually denoted by $\partial f(x)$; it is the so-called *subdifferential* of f at x.

The above deduction of (3.11) shows that convex functions always are subdifferentiable in the interior of their domains.

It is straightforward to show from (3.11) that the subdifferential, as a subset $\partial f(x) \subset \mathbb{R}$, is always closed and *convex*. When f is convex , then the subdifferential $\partial f(x)$ is therefore a non-empty interval containing both $a = f'_\pm(x)$, and in fact

$$\partial f(x) = [f'_-(x), f'_+(x)]. \tag{3.12}$$

Hence $\partial f(x) = f'(x)$ at each point x of differentiability of convex f.

As an appetizing example, consider $f(x) = |x|$ for $x \in \mathbb{R}$. At $x_0 = 0$, the subdifferential is simply determined from (3.12) as

$$\partial f(x_0) = [-1, 1]. \tag{3.13}$$

Although f is not differentiable at $x_0 = 0$ (so that $f'(x) = 0$ cannot be solved for $x_0 \dots$), this convex function f has a minimum at $x_0 = 0$ precisely because $0 \in [-1, 1]$!

3.2. **Subgradients in higher dimensions.** Let f denote a fixed convex function on a convex set $U \subset \mathbb{R}^n$ in this section.

As a definition, a vector $a \in \mathbb{R}^n$ is called a *subgradient* of f at $x \in U$ if

$$f(y) \geq f(x) + a \cdot (y-x) \qquad \text{for all } y \in U. \tag{3.14}$$

It is a general result that f has a subgradient a for each interior point x of U. (Notice that $a = \nabla f(x)$ is possible when f is differentiable at x; cf Theorem 5.1.)

First of all Proposition 3.3 generalises to directional derivatives:

Proposition 3.4. *A convex function $f\colon U \to \mathbb{R}$ has at every interior point $x \in U$ a one-sided directional derivative $f'_v(x)$ along any unit vector v, which is given by*

$$f'_v(x) = Df(x;v) = \lim_{t\to 0^+} \frac{f(x+tv)-f(x)}{t}. \tag{3.15}$$

Proof. As x is an interior point, $g(t) = f(x+tv)$ is at least defined on an open interval $]-\delta,\delta[$ containing 0. Since g is convex, it has by Proposition 3.3 a derivative from the right at $t=0$, namely $g'_+(0)$. But this is given by the same limit as the one defining $Df(x;v)$. □

As a sharpening of the proof above, note that $L(t) = x + tv$ is linear, so the pre-image of U° is a set $J \subset \mathbb{R}$, which is necessarily convex; hence J is an open *interval* on which the above $g(t)$ is defined; $0 \in J$. So for $t \in J$ it follows from (3.11) that

$$g(t) \ge g(0) + g'_+(0)t. \tag{3.16}$$

Returning to f itself, one has for *every* $y \in U$ of the form $y = x + tv$, since $t = tv \cdot v = (y-x)\cdot v$, that

$$f(y) \ge f(x) + f'_v(x)v\cdot(y-x) \qquad \text{for } y = x+tv,\ y \in U. \tag{3.17}$$

This strongly suggests that the one-sided directional derivatives give rise to the possible subgradients of f at x. However, this will not be pursued here.

Instead we proceed directly to the following main result:

Theorem 3.5. *When $f\colon U \to \mathbb{R}$ is a convex function, then f is* sub-differentiable *at every interior point x_0 of U. That is, to each $x_0 \in U^\circ$ there exists some $a \in \mathbb{R}^n$ such that*

$$f(x) - f(x_0) \ge a\cdot(x-x_0) \qquad \textit{for all } x \in U. \tag{3.18}$$

Proof. For the convex set $E(f) \subset \mathbb{R}^{n+1}$ the element $(x_0, f(x_0))$ is not an inner point, hence lies at the boundary. So according to Theorem 1.5 there is a unit vector $\vec{n} \in \mathbb{R}^{n+1}$ such that

$$E(f) \subset H_{(x_0,f(x_0)),\vec{n}}. \tag{3.19}$$

Here $\vec{n}$ may be written in the form $\vec{n} = (-v,s) \neq (0,0)$ with $v \in \mathbb{R}^n$, $s \in \mathbb{R}$. In particular the inclusion gives that for every $(x, f(x))$ with $x \in U$,

$$(-v,s)\cdot((x,f(x)) - (x_0,f(x_0))) \ge 0, \tag{3.20}$$

which clearly is equivalent to

$$s(f(x) - f(x_0))) \ge v\cdot(x-x_0). \tag{3.21}$$

It now suffices to show that $s > 0$, for then the claim holds with $a = \frac{1}{s}v$.

It is clear that $s = 0$ is impossible, for else $v \neq 0$ so that the resulting inequality $0 \ge v\cdot(x-x_0)$ must be false for some $x \in U$, because x_0 is an interior point of U: some $r > 0$ fulfils $B(x_0,r) \subset U$, and then $x = x_0 + tv$ belongs to this ball for a suitably small $t > 0$ (ie $x - x_0$

points in the same direction as v) which results in the contradiction $0 \geq tv \cdot v$.

Similarly $s < 0$ can be ruled out. However, the simpler way is to make full use of the above inclusion, which instead of (3.21) means that for every (x, y) in $E(f)$, ie for every $x \in U$ and $y \geq f(x)$,

$$s(y - f(x_0))) \geq v \cdot (x - x_0). \tag{3.22}$$

Hence $\vec{n} = (-v, s)$ cannot hold for $s < 0$, for in case $v = 0$ a contradiction results by inserting $(x, y) \in E(f)$ with $x = x_0$ and $y > f(x_0)$; if $v \neq 0$ one can modify by taking first $x \in U$ such that the right-hand side of (3.22) is positive (cf the treatment of $s = 0$), and then some $y > \max(f(x_0), f(x))$; clearly $(x, y) \in E(f)$, whence (3.22) yields the contradiction $s \geq 0$. □

For a *concave* function $g \colon U \to \mathbb{R}$, it is obvious from the above theorem that $-g$ has a subgradient $a \in \mathbb{R}^n$ at every interior point $x_0 \in U$. The vector $b = -a$ therefore fulfils

$$g(x) \leq g(x_0) + b \cdot (x - x_0) \quad \text{for every } x \in U. \tag{3.23}$$

Because of this inequality, b is said to be a *supergradient* of g at x_0.

Exercise 3.1. Find convex functions $f(x)$ on $\mathbb{R}$ for which the half-tangents $T_\pm$ are not parallel at some point.

Exercise 3.2. Complete the last part of the proof of Proposition 3.2.

Exercise 3.3. Prove for a convex function f on an open interval I that both $f'_\pm(x)$ are increasing. Is it even true that $y < z$ implies $f'_+(y) < f'_-(z)$?

Exercise 3.4. Show that the set of subgradients, ie $\partial f(x)$, is convex and closed. Prove for $n = 1$ that $\partial f(x)$ for convex f equals the interval $[f'_-(x), f'_+(x)]$. (*Hint:* Why is (3.11) impossible for $a > f'_+(x)$?)

Exercise 3.5. Show for $n = 1$ that the subdifferential of a convex function f on an open interval is a singleton, say $\partial f(x) = \{a\}$ if and only if f is differentiable at x with $f'(x) = a$.

Exercise 3.6. Verify the claims on $|x|$ and its subdifferential made above. (*Hint:* Read on in Section 5 below to see how the subgradient may be exploited.)

4. Convexity and Continuity

Continuity of a convex function $f \colon I \to \mathbb{R}$ will not in general hold on the boundary of I: defining $f(x) = 1$ for $x > 0$ while $f(0) = 2$, clearly f is convex on $I = [0, \infty[$, but discontinuous at $x = 0$.

But a convex function is automatically continuous on the interior of its domain. In dimension $n = 1$, this follows easily by using the slope function to invoke a 'sandwich trick':

Proposition 4.1. *When $f\colon I\to\mathbb{R}$ is convex on an interval $I\subset\mathbb{R}$, then f is continuous at every inner point $x\in I^\circ$.*

Proof. To show continuity at a given inner point x, one may fix $z>0$ so that $x\pm z\in I$. For $0<\theta\le 1$ one insert $y=x\pm z$ and θ into (2.2), which yields

$$f(x+\theta z)-f(x)\le\theta(f(x+z)-f(x)) \tag{4.1}$$
$$f(x-\theta z)-f(x)\le\theta(f(x-z)-f(x)). \tag{4.2}$$

Dividing these inequalities by θz and $-\theta z$, respectively, and using that the slope function is increasing, *cf* Proposition 3.2, it follows that

$$\begin{aligned} f(x)-f(x-z)&\le\frac{f(x)-f(x-\theta z)}{\theta}\\ &\le\frac{f(x+\theta z)-f(x)}{\theta}\le f(x+z)-f(x). \end{aligned} \tag{4.3}$$

Substituting $h=\theta$ or $h=-\theta$ this entails

$$f(x)-f(x-z)\le\frac{f(x+hz)-f(x)}{h}\le f(x+z)-f(x). \tag{4.4}$$

Now, by setting $M=\max(|f(x\pm z)-f(x)|)$, the above yields

$$0\le|f(x+hz)-f(x)|\le M|h|\qquad\text{for } -1\le h\le 1, \tag{4.5}$$

so $f(x+h)\to f(x)$ for $h\to 0$. Thence continuity of f at x. $\square$

An extension of the proof to dimensions $n\ge 2$ is not immediate, for if $B(x,\varepsilon)\subset U$, the corresponding M will be a supremum over infinitely many z-dependent constants as z runs through the sphere $|z|=\varepsilon$.

But the corresponding result *is* true for all $n\ge 1$. To circumvent the difficulties in dimension $n\ge 2$, one can use the banal observation that, for a convex function g on an interval $I\subset\mathbb{R}$,

$$g((1-\theta)x+\theta y)\le\max(g(x),g(y)). \tag{4.6}$$

In n dimensions this applies on a cube $Q=x+[-\delta,\delta]^n$ as follows, when y denotes an arbitrary point in Q,

$$\begin{aligned} f(y)&\le\max f(x_1\pm\delta,y_2,\dots,y_n)\\ &\le\max f(x_1\pm\delta,x_2\pm\delta,y_3,\dots,y_n)\\ &\le\dots\\ &\le\max f(x_1\pm\delta,\dots,x_n\pm\delta)=:M_x. \end{aligned} \tag{4.7}$$

Therefore f is bounded from above on every closed cube $Q\subset U$. In fact, this even shows that f attains its maximum on every such cube.

Thus prepared one can give an elementary proof of

Theorem 4.2. *When $f\colon U\to\mathbb{R}$ is a convex function on a convex set $U\subset\mathbb{R}^n$, then f is continuous at every inner point of U.*

Proof. Given an inner point $x \in U$, one may fix $\delta \in]0,1[$ so small that $Q = x + [-\delta,\delta]^n$ is contained in U. Then $B(x, \delta/\sqrt{n}) \subset U$ too.

Now one has whenever $|z| = \delta/\sqrt{n}$ and $0 \le |h| \le 1$ that

$$|f(x+hz) - f(x)| \le |h|(M_x + |f(x)|). \tag{4.8}$$

This may be seen by following the proof of Proposition 4.1, where the argument based on the slope function now applies to the convex auxiliary function $g(t) = f(x+tz)$, resulting in (4.4) in the present situation. But by using (4.7),

$$f(x+z) - f(x) \le M_x + |f(x)| \tag{4.9}$$

$$f(x) - f(x-z) \ge -M_x - |f(x)|. \tag{4.10}$$

Hence (4.8) follows from (4.4). So continuity of f at x is immediate. □

Exercise 4.1. Derive from Theorem 4.2 and its proof that a convex function $f\colon U \to \mathbb{R}$ is locally Lipschitz continuous. Give examples in which "locally" cannot be omitted.

Exercise 4.2. Prove that $E(f)$ is closed in $U \times \mathbb{R}$ whenever $f\colon U \to \mathbb{R}$ is convex and the convex set $U \subset \mathbb{R}^n$ is open. Is $E(f)$ closed in $\mathbb{R}^n \times \mathbb{R}$? (*Hint:* f is continuous.)

5. Convexity and Differentiability

In addition to the geometric property in (3.1), one should envisage that the graph of a convex function must lie above each of its tangents. This is confirmed in

Theorem 5.1. *If $f\colon U \to \mathbb{R}$ is convex and differentiable at some point $x \in U^\circ$, then*

$$f(y) \ge f(x) + \nabla f(x) \cdot (y-x) \quad \textit{for all} \quad y \in U. \tag{5.1}$$

When f is strictly convex, this inequality is strict for every $y \neq x$.

Proof. There is only something to show in (5.1) when $y \neq x$. Convexity implies that for $0 < \theta \le 1$,

$$\tfrac{1}{\theta}(f(x+\theta(y-x)) - f(x)) \le f(y) - f(x), \tag{5.2}$$

and differentiability at x shows that the left-hand side equals

$$\nabla f(x) \cdot (y-x) + \frac{o(\|\theta(y-x)\|)}{\theta\|y-x\|}\|y-x\|. \tag{5.3}$$

By definition of o, the second term tends to 0 for $\theta \to 0$, so the left-hand side of (5.2) tends to $\nabla f(x) \cdot (y-x)$. This gives (5.1).

When (5.1) is an identity for some $y_0 \neq x$, and $y_0 \in U$, then a successive application of (5.1), (5.2) and the property of y_0 gives

$$\begin{aligned} \nabla f(x) \cdot \theta(y_0 - x) &\le f(x + \theta(y_0 - x)) - f(x) \\ &\le \theta(f(y_0) - f(x)) = \theta \nabla f(x) \cdot (y_0 - x). \end{aligned} \tag{5.4}$$

Obviously this implies that $f(x+\theta(y_0-x)) = f(x)+\theta\nabla f(x)\cdot(y_0-x)$. This means that f behaves linearly along the line segment from x to y_0. Hence f is not strictly convex, when such y_0 exists. $\square$

From the above theorem one can read off a result, which makes it easy to find extrema of convex functions:

Corollary 5.2. *When a convex function $f\colon U \to \mathbb{R}$ is differentiable at an interior critical point x^*, i.e. $\nabla f(x^*) = 0$ for some $x^* \in U^\circ$, then the point x^* is a global minimum of f.*

If f is strictly convex, a local minimum $x^* \in U$ is global and unique.

Proof. When $\nabla f(x^*) = 0$ at some interior point, then (5.1) reduces to the inequality

$$f(y) \geq f(x^*) \quad \text{for all } y \in U, \tag{5.5}$$

which means that x^* is a global minimum. See also Exercise 2.4. $\square$

The corresponding result is valid for concave functions, of course, though their critical points are necessarily maxima.

In addition to the geometric property stated prior to Theorem 5.1, differentiable convex functions have for $n = 1$ a closely related analytic property:

Theorem 5.3. *A differentiable function $f\colon I \to \mathbb{R}$ on an open interval $I \subset \mathbb{R}$ is convex if and only if $f'(x)$ is monotone increasing on I.*

Proof. When f' exists and is monotone increasing on I, then the inequality (3.1) follows at once from the Mean Value Theorem, which reduces it to the statement that $f'(s) \leq f'(t)$ for some $s \in]y,x[$ and $t \in]x,z[$ (as is true by assumption on f').

Conversely, if f is a given differentiable convex function and $y < z$ for y, $z \in I$, Proposition 3.2 implies that the numbers

$$a_k := \frac{f(y+1/k)-f(y)}{1/k}, \qquad \frac{f(z)-f(z-1/k)}{1/k} =: b_k, \tag{5.6}$$

considered for k so large that $\frac{1}{k} < (z-y)/2$, always fulfil

$$a_k = S(y,y+1/k) \leq S(y,z) \leq S(z-1/k,z) = b_k. \tag{5.7}$$

In the limit $k \to \infty$ one has $a_k \to f'(y)$ while $b_k \to f'(z)$, so the inequality $f'(y) \leq f'(z)$ results. $\square$

Exercise 5.1. Can you prove for a differentiable function $f\colon U \to \mathbb{R}$ on an *open* convex set $U \subset \mathbb{R}^n$ that convexity of f is equivalent to validity of (5.1) at every $x \in U$? (*Hint:* Try first for $n = 1$.)

6. Convexity and Hessian Matrices

It was seen above that when f is convex and C^2 on an interval, then f' is monotone increasing. Therefore $f''(x) \geq 0$ follows from convexity. This is actually a characterisation (for C^2-functions):

Theorem 6.1. *When $f \in C^2(I, \mathbb{R})$, for an open interval $I \subset \mathbb{R}$, then $f''(x) \geq 0$ holds for all $x \in I$ if and only if f is a convex function.*
If $f'' > 0$ on I, then f is strictly *convex.*

Proof. That convexity implies the positivity of f'' was seen prior to the theorem. To prove sufficiency of $f'' \geq 0$, consider first a function $g \in C^1(I, \mathbb{R})$. The fundamental theorem of calculus gives, for $z \geq 0$ so small that both x and $x + z$ are in I,

$$g(x+z) - g(x) = \int_0^1 \tfrac{d}{dt} g(x+tz)\,dt = z \int_0^1 g'(x+tz)\,dt.$$

If $g' \geq 0$ on I the integral is positive, so then g is monotone increasing.
For $g = f'$ this gives that $f'(x+z) \geq f'(x)$, hence that f' is increasing, and from the above with $g = f$,

$$(f(x+z) - f(x))/z = \int_0^1 f'(x+tz)\,dt, \qquad z > 0.$$

Here the right-hand side is an increasing function of z, for the integrand is so, as just shown; whence the left-hand side is increasing with respect to z. Since $tz \leq z$ for $t \leq 1$ this gives that, for $z \geq 0$ and $x, x+z \in I$,

$$f(x+tz) - f(x) \leq t(f(x+z) - f(x)) \quad \text{for} \quad 0 \leq t \leq 1.$$

Since eg $x + tz = (1-t)x + t(x+z)$, the choice $t = \theta$, $y = x + z$ proves (2.1) for points $x < y$ in I. Now, $x < y$ can be assumed without loss of generality in (2.1), so this proves the convexity.
When $f''(x) > 0$ for all $x \in I$, then f' is strictly increasing so that also the above integral is strictly increasing with respect to z; hence the inequality above is strict for $0 < t < 1$, $z > 0$. Therefore (2.1) is a strict inequality for $x \neq y$ and $0 < \theta < 1$, so that f is strictly convex. The proof is complete. □

According to the theorem strict positivity always suffices for strict convexity. But this implication cannot be reversed: $f(x) = x^4$ is strictly convex on $\mathbb{R}$, yet $f''(x) = 0$ holds for $x = 0$.
In dimensions $n \geq 1$ there are similar results, which characterises the convex C^2-functions f as those for which the Hessian matrix $Hf(x)$ is positive semidefinite for all x. This should be evident from the fact $Hf(x)$ gives rise to the quadratic form given on $y \in \mathbb{R}^n$ by

$$y^T Hf(x) y = \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(x) y_j y_k. \tag{6.1}$$

Theorem 6.2. *When $f \in C^2(U, \mathbb{R})$, for an open, convex set $U \subset \mathbb{R}^n$, then f is convex if and only if it holds true for every $x \in U$ that*

$$0 \leq \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(x) y_j y_k \qquad \text{for all } y \in \mathbb{R}^n. \tag{6.2}$$

If this inequality is strict for all $y \neq 0$ and $x \in U$, then f is strictly *convex.*

Proof. When x, $x + y \in U$, then (6.2) implies via the chain rule that

$$\frac{d^2 f(x+ty)}{dt^2} \geq 0 \quad \text{ for } t \in \,]-\delta, 1+\delta[\,, \tag{6.3}$$

for some $\delta > 0$ since U is open. Conversely, if this is true whenever x, $x+y \in U$, it is seen for $t = 0$ that (6.2) holds for y with $\|y\| < \varepsilon$ when $B(x, \varepsilon) \subset U$; but then (6.2) also holds for every $y \in \mathbb{R}^n$ by scaling.

Now, (6.3) is equivalent to convexity of $g(t) = f(x + ty)$ on the interval $]-\delta, 1+\delta[$, according to Theorem 6.1. However, convexity of g immediately implies (2.2), hence that f is convex; in fact,

$$f(x+ty) = g(t) \leq g(0)+t(g(1)-g(0)) = f(x)+t(f(x+y)-f(x)). \tag{6.4}$$

Conversely, it is geometrically clear that g is convex if f is so.

When (6.2) is strict for each $y \neq 0$, then (6.3) is so; whence g is strictly convex by Theorem 6.1. Therefore the above inequality is strict for $0 < t < 1$, so f is strictly convex. $\square$

Notice that, according to the last part of the theorem, when the Hessian $Hf(x)$ is positive definite for all x, then f is strictly convex.

Since $Hf(x)$ is symmetric for every C^2-function f, it has n real eigenvalues $\lambda_1, \dots, \lambda_n$ (counted with multiplicity) by the Spectral Theorem. Diagonalisation of the quadratic form in (6.1) therefore yields

Corollary 6.3. *When $f \in C^2(U, \mathbb{R})$, for an open, convex set $U \subset \mathbb{R}^n$, then f is convex if and only if the Hessian matrix $Hf(x)$ for every $x \in U$ fulfils the eigenvalue condition*

$$\lambda_1 \geq 0, \dots, \lambda_n \geq 0. \tag{6.5}$$

If these inequalities are all strict, then f is strictly convex.

Exercise 6.1. Give an analytical proof, by means of inequalities, that g inherits convexity from f in the proof of Theorem 6.2.

Exercise 6.2. Prove Corollary 6.3 in details.

7. Other Notions

Sometimes only a part of the properties of convex or concave functions are needed. It is therefore convenient to introduce the following generalisation:

Definition 7.1. A *quasi-concave* function $f\colon U \to \mathbb{R}$ on a convex set $U \subset \mathbb{R}^n$ is a function with the property that for all $x, y \in U$

$$f(y) \geq f(x) \implies \forall\theta \in [0,1]\colon f(x+\theta(y-x)) \geq f(x). \tag{7.1}$$

First of all, every concave function f on U is quasi-concave; this follows at once from the opposite inequality of (2.2) that holds for concave functions.

Secondly, it should be noted that the condition for quasi-concavity simply means that whenever $f(y)$ is larger than $f(x)$, then f should be larger than $f(x)$ on the entire line segment between x and y. Therefore one has that

- e^x and x^2 are quasi-concave on $\mathbb{R}$ (although they are convex!);
- the function $f\colon [0,\infty] \to \mathbb{R}$ given by

$$f(x) = \begin{cases} x^2 & \text{for } 0 \leq x < 1 \\ 2+\sqrt{x-1} & \text{for } 1 \leq x < \infty \end{cases} \tag{7.2}$$

  is quasi-concave (though pieced together of a convex and a concave function, with a discontinuity at $x = 1$!).

It is therefore clear that quasi-concave functions constitute a much more general class than, say concave functions.

Quasi-concavity can also be expressed in terms of the gradient, when it exists:

Theorem 7.2. *When f is quasi-concave on U and differentiable at $x \in U$, then it holds for all $y \in U$ that*

$$f(y) \geq f(x) \implies \nabla f(x) \cdot (y-x) \geq 0. \tag{7.3}$$

Proof. Given that $f(y) \geq f(x)$, then $f(z) - f(x) \geq 0$ for every $z = x + \theta(y-x)$ with $\theta \in [0,1]$. Therefore

$$(\nabla f(x) \cdot \frac{z-x}{\|z-x\|} + o(1))\|z-x\| \geq 0 \tag{7.4}$$

For z in a small neighbourhood of x this shows that $\nabla f(x) \cdot \frac{z-x}{\|z-x\|} < 0$ is impossible. Since U is convex, it is then seen by scaling that the claimed inequality holds. □

Exercise 7.1. Find functions $f\colon \mathbb{R} \to \mathbb{R}$ which are not quasi-concave.

8. Final remarks

The subject of convex functions is classical, and it has been analysed much further in the literature. Cf the elegant, but rather abstract exposition in R. T. Rockafellar's classic text [Roc97], for example. A rather more elementary and illustrated presentation is available in the lecture notes of E. Christensen [Chr04]. An easily read account of classical convexity as well as the many related notions, and their applications, can be found in the book [NP06] of C. Niculescu and L.-E. Persson.

References

[Chr04] E. Christensen, *Konvekse mængder. Matematik 2OK*, University of Copenhagen, Denmark; ISBN 87-91180-22-8 (2004). (Available from `http://www.math.ku.dk/noter/filer/matematik.htm`)

[NP06] C. P. Niculescu and L.-E. Persson, *Convex functions and their applications. a contemporary approach*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, vol. 23, Springer, New York, 2006.

[Roc97] R. T. Rockafellar, *Convex analysis*, Princeton University Press, Princeton, NJ, 1997, Reprint of the 1970 original, Princeton Paperbacks.