Lecture 3

Convex functions

(Basic properties; Calculus; Closed functions; Continuity of convex functions; Subgradients; Optimality conditions)

3.1 First acquaintance

Definition 3.1.1 [Convex function] *A function $f : M \to \mathbf{R}$ defined on a nonempty subset M of $\mathbf{R}^n$ and taking real values is called convex, if*

- *the domain M of the function is convex;*
- *for any $x, y \in M$ and every $\lambda \in [0,1]$ one has*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \qquad (3.1.1)$$

If the above inequality is strict whenever $x \neq y$ and $0 < \lambda < 1$, f is called strictly convex.

A function f such that $-f$ is convex is called *concave*; the domain M of a concave function should be convex, and the function itself should satisfy the inequality opposite to (3.1.1):

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda) f(y), \quad x, y \in M, \lambda \in [0,1].$$

The simplest example of a convex function is an *affine function*

$$f(x) = a^T x + b$$

– the sum of a linear form and a constant. This function clearly is convex on the entire space, and the "convexity inequality" for it is equality; the affine function is also concave. It is easily seen that the function which is both convex and concave on the entire space is an affine function.

Here are several elementary examples of "nonlinear" convex functions of one variable:

- functions convex on the whole axis:

  x^{2p}, p being positive integer;

  $\exp\{x\}$;

- functions convex on the nonnegative ray:

  x^p, $1 \leq p$;

  $-x^p$, $0 \leq p \leq 1$;

  $x \ln x$;

- functions convex on the positive ray:

  $1/x^p$, $p > 0$;

  $-\ln x$.

To the moment it is not clear why these functions are convex; in the mean time we will derive a simple analytic criterion for detecting convexity which immediately demonstrates that the above functions indeed are convex.

More examples of convex functions: norms. Recall that a real-valued function $\pi(x)$ on $\mathbf{R}^n$ is called a norm, if it is nonnegative everywhere, positive outside of the origin, is homogeneous:

$$\pi(tx) = |t|p(x)$$

and satisfies the triangle inequality

$$\pi(x+y) \leq \pi(x) + \pi(y).$$

We are about to prove that every norm is convex:

Proposition 3.1.1 *Let $\pi(x)$ be a real-valued function on $\mathbf{R}^n$ which is positively homogeneous of degree 1:*

$$\pi(tx) = t\pi(x) \quad \forall x \in \mathbf{R}^n, t \geq 0.$$

π is convex if and only if it is sub-additive:

$$\pi(x+y) \leq \pi(x) + \pi(y) \quad \forall x, y \in \mathbf{R}^n.$$

In particular, a norm (which by definition is positively homogeneous of degree 1 and is sub-additive) is convex.

Proof is immediate:

$$\begin{array}{rcll} \pi(\lambda x + (1-\lambda)y) & \leq & \pi(\lambda x) + \pi((1-\lambda)y) & \text{triangle inequality} \\ & = & \lambda\pi(x) + (1-\lambda)\pi(y) & \text{homogeneity,} \end{array}$$

for any $x, y \in \mathbf{R}^n$ and $0 \leq \lambda \leq 1$.■

The most important (to us) examples of norms are so-called l_p-norms:

$$|x|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}, \quad 1 \leq p \leq \infty,$$

(the right hand side of the latter relation for $p = \infty$ is, *by definition*, $\max_i |x_i|$) though it is not clear in advance that these functions are indeed norms. However, you certainly know quite well three members of the above family of norms:

- The Euclidean norm $|x| = |x|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^n x_i^2}$. This is indeed a norm due to the Cauchy-Schwartz inequality.
- The l_1-norm $|x|_1 = \sum_{i=1}^n |x_i|$.
- The l_∞-norm or uniform norm (or Tchebychev norm) $|x|_\infty = \max_{1\leq i\leq n} |x_i|$.

Epigraph of a convex function. A very convenient equivalent definition of a convex function is in terms of its *epigraph*. Given a real-valued function f defined on a nonempty subset M of $\mathbf{R}^n$, we define its epigraph as the set

$$\mathrm{Epi}\,(f) = \{(t, x) \in \mathbf{R}^{n+1} \mid x \in M,\, t \geq f(x)\};$$

geometrically, to define the epigraph, you should take the *graph* of the function – the surface $\{t = f(x),\, x \in M\}$ in $\mathbf{R}^{n+1}$ – and add to this surface all points which are "above" it. The equivalent, more geometrical, definition of a convex function is given by the following simple

Proposition 3.1.2 $^{+}$ [Definition of convexity in terms of the epigraph]
A function f defined on a subset of $\mathbf{R}^n$ is convex if and only if its epigraph is a nonempty convex set in $\mathbf{R}^{n+1}$.

We will see later that the behavior of a "general" convex function on the boundary of its domain is sometimes out of control. The following definition will be useful in the future analysis.

Definition 3.1.2 [Closed convex function] *A convex function f is called* closed *if its epigraph is a closed set.*

Note that the function which is convex and continuous on a closed domain is a closed function. For instance, the norms are closed convex functions. Further, all *differentiable* convex functions are closed with $\mathrm{Dom}\, f = \mathbf{R}^n$. The function $f(x) = \frac{1}{x}$, $x \in [0, +\infty]$ is convex and closed, however, its domain, which is $(0, \infty)$ is open.

When speaking about *convex* functions, it is extremely convenient to think that the function outside its domain also has a value, namely, takes the value $+\infty$; with this convention, we can say that *a convex function f on $\mathbf{R}^n$ is a function taking values in the extended real axis $\mathbf{R}\cup\{+\infty\}$ such that the domain* $\mathrm{Dom}\, f$ *of the function – the set of those x's where $f(x)$ is finite – is nonempty, and for all $x, y \in \mathbf{R}^n$ and all $\lambda \in [0,1]$ one has*

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y). \tag{3.1.2}$$

If the expression in the right hand side involves infinities, it is assigned the value according to the standard and reasonable conventions on what are arithmetic operations in the "extended real axis" $\mathbf{R}\cup\{+\infty\}\cup\{-\infty\}$.

The following simple observation is maybe the most useful mathematical statement I know:

Proposition 3.1.3 [Jensen's inequality] *Let f be convex. Then for any convex combination*

$$\sum_{i=1}^N \lambda_i x_i$$

one has

$$f(\sum_{i=1}^N \lambda_i x_i) \le \sum_{i=1}^N \lambda_i f(x_i).$$

The proof is immediate: the points $(f(x_i), x_i)$ clearly belong to the epigraph of f; since f is convex, its epigraph is a convex set, so that the convex combination

$$\sum_{i=1}^N \lambda_i (f(x_i), x_i) = (\sum_{i=1}^N \lambda_i f(x_i), \sum_{i=1}^N \lambda_i x_i)$$

of the points also belongs to $\mathrm{Epi}\,(f)$. By definition of the epigraph, the latter means exactly that $\sum_{i=1}^N \lambda_i f(x_i) \ge f(\sum_{i=1}^N \lambda_i x_i)$. ■

Note that the definition of convexity of a function f is exactly the requirement on f to satisfy the Jensen inequality for the case of $N = 2$; we see that to satisfy this inequality for $N = 2$ is the same as to satisfy it for *all* N.

The following statement is a trivial corollary of the Jensen inequality:

Corollary 3.1.1 *Let f be a convex function and let x be a convex combination of the points $x_1, ..., x_N$. Then*

$$f(x) \le \max_{1\le i\le N} f(x_i).$$

In other words, if Δ is a convex hull of $x_1, ..., x_N$, i.e.

$$\Delta = \mathrm{Conv}\,\{x_1, ..., x_N\} \equiv \{x \in \mathbf{R}^n \mid x = \sum_{i=1}^N \lambda_i x_i,\ \alpha \ge 0,\ \sum_{i=1}^N \alpha_i = 1\},$$

then $\max_{x\in\Delta} f(x) \le \max_{1\le i\le N} f(x_i)$.

Convexity of level sets of a convex function

The following simple observation is also very useful:

Proposition 3.1.4 [Convexity of level sets] *Let f be a convex function with the domain M. Then, for any real α, the set*

$$\mathrm{lev}_{\alpha}(f) = \{x \in M \mid f(x) \leq \alpha\}$$

– the level set of f – is convex.

The proof takes one line: if $x, y \in \mathrm{lev}_{\alpha}(f)$ and $\lambda \in [0, 1]$, then $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y) \leq \lambda\alpha + (1-\lambda)\alpha = \alpha$, so that $\lambda x + (1-\lambda)y \in \mathrm{lev}_{\alpha}(f)$.

Note that the convexity of level sets does *not* characterize convex functions; there are nonconvex functions which share this property (e.g., any monotone function on the axis). The "proper" characterization of convex functions in terms of convex sets is given by Proposition 3.1.2 – convex functions are exactly the functions with convex epigraphs. Convexity of level sets specifies a wider family of functions, the so called *quasiconvex* ones.

Proposition 3.1.5 $^{+}$[Closeness of level sets] *If a convex function f is closed, then all its level sets are closed.*

Recall that an empty set is closed (and, by the way, is open).

Example 3.1.1 [$\|\cdot\|$-ball] *The unit ball of norm $\|\cdot\|$ – the set*

$$\{x \in \mathbf{R}^n \mid \|x\| \leq 1\},$$

same as any other $\|\cdot\|$-ball

$$\{x \mid \|x - a\| \leq r\}$$

($a \in \mathbf{R}^n$ and $r \geq 0$ are fixed) is convex and closed.

In particular, Euclidean balls ($|\cdot|$-balls associated with the standard Euclidean norm $|\cdot| = |\cdot|_2$) are convex and closed.

Let us look at the l_p-balls. For $p = 2$ we get the usual Euclidean norm; of course, you know how the Euclidean ball looks. When $p = 1$, we get

$$|x|_1 = \sum_{i=1}^{n} |x_i|,$$

and the unit ball is the *hyperoctahedron*

$$V = \{x \in \mathbf{R}^n \mid \sum_{i=1}^{n} |x_i| \leq 1\}$$

When $p = \infty$, we get

$$|x|_\infty = \max_{1 \leq i \leq n} |x_i|,$$

and the unit ball is the *hypercube*

$$V = \{x \in \mathbf{R}^n \mid -1 \leq x_i \leq 1,\ 1 \leq i \leq n\}.$$

It makes sense to draw the unit $|\cdot|_1$- and $|\cdot|_\infty$-balls in $\mathbf{R}^2$.

Example 3.1.2 [Ellipsoid] *Let Q be a $n \times n$ matrix which is symmetric ($Q = Q^T$) and positive definite ($x^T Q x \geq 0$, with $\geq$ being $=$ if and only if $x = 0$). Then, for any nonnegative r, the Q-ellipsoid of radius r centered at a – the set*

$$\{x \mid (x-a)^T Q (x-a) \leq r^2\}$$

is convex and closed.

The simplest way to prove that an ellipsoid is convex and closed is as follows: given a positive definite symmetric matrix Q, one can associate with it the Q-inner product

$$\langle x, y \rangle = x^T Q y$$

which, as it is immediately seen, satisfies the characteristic properties – bilinearity, symmetry and positivity – of the standard inner product $x^T y$ (in fact these three properties of a Q-inner product, taken together, are exactly equivalent to symmetry and positive definiteness of Q). It follows that the Q-*norm* – the function

$$|x|_Q = \sqrt{x^T Q x}$$

– is a norm. It is clearly seen that a Q-ellipsoid is nothing but a ball in the norm $|\cdot|_Q$, so that its convexity is given by Example 3.1.1.

Example 3.1.3 $^+$ [ϵ-neighborhood of a convex set]
Let M be a convex set in $\mathbf{R}^n$, and let $\epsilon > 0$. Then, for any norm $\|\cdot\|$ on $\mathbf{R}^n$, the ϵ-neighborhood of M, i.e., the set

$$M_\epsilon = \{y \in \mathbf{R}^n \mid \mathrm{dist}_{\|\cdot\|}(y, M) \equiv \inf_{x \in M} \|y - x\| \leq \epsilon\}$$

is convex and closed.

3.2 How to detect convexity

In an optimization problem

$$f(x) \to \min \mid \ g_j(x) \leq 0, \ j = 1, ..., m$$

convexity of the objective f and the constraints g_i is crucial: it turns out that problems with this property possess nice theoretical properties (e.g., the local *necessary* optimality conditions for these problems are *sufficient for global optimality*); and what is much more important, convex problems can be efficiently (both in theoretical and, to some extent, in the practical meaning of the word) solved numerically, which is not, unfortunately, the case for general nonconvex problems. This is why it is so important to know how one can detect convexity of a given function. This is the issue we are coming to.

The scheme of our investigation is typical for mathematics. Let me start with the example which you know from Analysis. How do you detect continuity of a function? Of course, there is a definition of continuity in terms of ϵ and δ, but it would be an actual disaster if each time we need to prove continuity of a function, we were supposed to write down the proof that "for any positive ϵ there exists positive δ such that ...". In fact we use another approach: we list once for ever a number of standard operations which preserve continuity, like addition, multiplication, taking superpositions, etc., and point out a number of standard examples of continuous functions – like the power function, the exponent, etc. To prove that the operations in the list preserve continuity, same as to prove that the standard functions are continuous, this takes certain effort and indeed is done in $\epsilon - \delta$ terms; but after this effort is once invested, we normally have no difficulties with proving continuity of a given function: it suffices to demonstrate that the function can be obtained, in finitely many steps, from our "raw materials" – the standard functions which are known to be continuous – by applying our machinery – the combination rules which preserve continuity. Normally this demonstration is given by a single word "evident" or even is understood by default.

This is exactly the case with convexity. Here we also should point out the list of operations which preserve convexity and a number of standard convex functions.

3.2.1 Operations preserving convexity of functions

These operations are as follows:

- [stability under taking weighted sums] if f, g are convex (and closed) functions on $\mathbf{R}^n$, then their linear combination $\lambda f + \mu g$ with *nonnegative* coefficients again is convex (and closed), provided that it is finite at least at one point;

  [this is given by straightforward verification of the definition]

- [stability under affine substitutions of the argument] the superposition $\phi(x) = f(Ax+b)$ of a convex (and closed) function f on $\mathbf{R}^n$ and affine mapping $x \mapsto Ax + b$ from $\mathbf{R}^m$ into $\mathbf{R}^n$ is convex (and closed), provided that it is finite at least at one point.

  [we can prove it directly by verifying the definition: let x_1 and x_2 in $\mathbf{R}^m$ and $y_i = Ax_i+b$, $i = 1, 2$. Then for $0 \le \lambda \le 1$ we have:

$$\begin{aligned}\phi(\lambda x_1 + (1-\lambda)x_2) &= f(A(\lambda x_1 + (1-\lambda)x_2) + b) = f(\lambda y_1 + (1-\lambda)y_2) \\ &\le \lambda f(y_1) + (1-\lambda)f(y_2) = \lambda\phi(x_1) + (1-\lambda)\phi(x_2).\end{aligned}$$

  The closeness of the epigraph of ϕ follows from the continuity of the affine mapping.]

- [stability under taking pointwise sup] upper bound $\sup_\alpha f_\alpha(\cdot)$ of any family of convex (and closed) functions on $\mathbf{R}^n$ is convex (and closed), provided that this bound is finite at least at one point.

  [to understand it, note that the epigraph of the upper bound clearly is the intersection of epigraphs of the functions from the family; recall that the intersection of any family of convex (closed) sets is convex (closed)]

- ["Convex Monotone superposition"] Let $f(x) = (f_1(x), ..., f_k(x))$ be vector-function on $\mathbf{R}^n$ with convex components f_i, and assume that F is a convex function on $\mathbf{R}^k$ which is *monotone*, i.e., such that $z \leq z'$ always implies that $F(z) \leq F(z')$. Then the superposition
$$\phi(x) = F(f(x)) = F(f_1(x), ..., f_k(x))$$
is convex on $\mathbf{R}^n$, provided that it is finite at least at one point.

Remark 3.2.1 *The expression $F(f_1(x), ..., f_k(x))$ makes no evident sense at a point x where some of f_i's are $+\infty$. By definition, we assign the superposition at such a point the value $+\infty$.*

[To justify the rule, note that if $\lambda \in (0,1)$ and $x, x' \in \mathrm{Dom}\,\phi$, then $z = f(x), z' = f(x')$ are vectors from $\mathbf{R}^k$ which belong to $\mathrm{Dom}\,F$, and due to the convexity of the components of f we have
$$f(\lambda x + (1-\lambda)x') \leq \lambda z + (1-\lambda)z';$$
in particular, the left hand side is a vector from $\mathbf{R}^k$ – it has no "infinite entries", and we may further use the monotonicity of F:
$$\phi(\lambda x + (1-\lambda)x') = F(f(\lambda x + (1-\lambda)x')) \leq F(\lambda z + (1-\lambda)z')$$
and now use the convexity of F:
$$F(\lambda z + (1-\lambda)z') \leq \lambda F(z) + (1-\lambda)F(z')$$
to get the required relation
$$\phi(\lambda x + (1-\lambda)x') \leq \lambda\phi(x) + (1-\lambda)\phi(x').$$
]

Two more rules are as follows:

- [stability under partial minimization] if $f(x,y) : \mathbf{R}^n_x \times \mathbf{R}^m_y$ is convex (as a function of $z = (x,y)$; this is called *joint convexity*) and the function
$$g(x) = \inf_y f(x,y)$$
is *proper*, i.e., is $> -\infty$ everywhere and is finite at least at one point, then g is convex
[this can be proved as follows. We should prove that if $x, x' \in \mathrm{Dom}\,g$ and $x'' = \lambda x + (1-\lambda)x'$ with $\lambda \in [0,1]$, then $x'' \in \mathrm{Dom}\,g$ and $g(x'') \leq \lambda g(x) + (1-\lambda)g(x')$. Given positive ϵ, we can find y and y' such that $(x,y) \in \mathrm{Dom}\,f$, $(x', y') \in \mathrm{Dom}\,f$ and $g(x) + \epsilon \geq f(x,y)$, $g(y') + \epsilon \geq f(x', y')$. Taking weighted sum of these two inequalities, we get
$$\lambda g(x) + (1-\lambda)g(y) + \epsilon \geq \lambda f(x,y) + (1-\lambda)f(x',y') \geq$$

[since f is convex]

$$\geq f(\lambda x + (1-\lambda)x', \lambda y + (1-\lambda)y') = f(x'', \lambda y + (1-\lambda)y')$$

(the last $\geq$ follows from the convexity of f). The concluding quantity in the chain is $\geq g(x'')$, and we get $g(x'') \leq \lambda g(x) + (1-\lambda)g(x') + \epsilon$. In particular, $x'' \in \mathrm{Dom}\, g$ (recall that g is assumed to take only the values from $\mathbf{R}$ and the value $+\infty$). Moreover, since the resulting inequality is valid for all $\epsilon > 0$, we come to $g(x'') \leq \lambda g(x) + (1-\lambda)g(x')$, as required.]

- the "conic transformation" of a convex function f on $\mathbf{R}^n$ – the function $g(y,x) = yf(x/y)$ – is convex in the half-space $y > 0$ in $\mathbf{R}^{n+1}$.

We have just learned (cf Example 3.1.1) that a (unit) ball of the norm $\|\cdot\|$ is a convex set. The inverse is also true: any convex and closed set $M \subset \mathbf{R}^n$ which is also symmetric an spans the whole $\mathbf{R}^n$ can be used to "generate" a norm on $\mathbf{R}^n$:

Example 3.2.1 *Consider the support function of a convex set M:*

$$\psi_M(x) = \sup\{y^T x | \, y \in M\}.$$

Note that ψ_M is convex and closed as the maximum of linear functions. Further, this function is positively homogenous of degree 1:

$$\psi_M(tx) = t\psi_M(x), \quad x \in \mathrm{Dom}\, \psi_M, \quad t \geq 0,$$

and if the set M is bounded then $\mathrm{Dom}\, \psi_M = \mathbf{R}^n$.

Let us consider the function $\psi(x,\gamma) = sup_{y\in M}\phi(y,x,\gamma)$, where

$$\phi(y,x,\gamma) = y^T x - \frac{\gamma}{2}|y|_2^2.$$

This function is convex and closed. Let us look at its properties.

If M is bounded then $\mathrm{Dom}\, \psi = \mathbf{R}^n$. Consider the case $M = \mathbf{R}^n$. Clearly, $\mathrm{Dom}\, \psi$ contains only points with $\gamma \geq 0$. If $\gamma = 0$, the only possible value of x is zero, since otherwise the function $\phi(y,x,0)$ is unbounded. Finally, if $\gamma > 0$, then the point maximizing $\phi(y,x,\gamma)$ with respect to y is $y^* = \frac{x}{\gamma}$ and $\psi(x,\gamma) = \frac{|x|_2^2}{2\gamma}$.

When summing up,

$$\psi(x,\gamma) = \begin{cases} 0, & \text{if} \quad x = 0, \ \gamma = 0; \\ \frac{|x|_2^2}{2\gamma} & \text{if} \quad \gamma > 0, \end{cases}$$

and the domain of ψ is the set $\mathbf{R}^n \times \{\gamma > 0\} \cup \{0, 0\}$. This set is neither open nor close, nevertheless, ψ is a closed convex function. Note that this function is not continuous at the origin:

$$\lim_{\gamma \downarrow 0} \psi(\sqrt{\gamma} x, \gamma) = \frac{1}{2}|x|_2^2.$$

Now we know what are the basic operations preserving convexity. Let us look what are the standard functions these operations can be applied to. A number of examples was already given, but we still do not know why the functions in the examples are convex. The usual way to check convexity of a "simple" – given by a simple formula – function is based on *differential criteria of convexity*. Let us look what are these criteria.

3.2.2 Differential criteria of convexity

From the definition of convexity of a function it immediately follows that convexity is one-dimensional property: a proper (i.e., finite at least at one point) function f on $\mathbf{R}^n$ taking values in $\mathbf{R} \cup \{+\infty\}$ is convex if and only if its restriction on any line, i.e., any function of the type $g(t) = f(x + th)$ on the axis, is either convex, or is identically $+\infty$.

It follows that to detect convexity of a function, it, in principle, suffices to know how to detect convexity of functions of one variable. This latter question can be resolved by the standard Calculus tools. Namely, in the Calculus they prove the following simple

Proposition 3.2.1 [Necessary and Sufficient Convexity Condition for smooth functions on the axis] *Let (a, b) be an interval on the real axis (we do not exclude the case of $a = -\infty$ and/or $b = +\infty$). Then*

(i) *A differentiable everywhere on (a, b) function f is convex on (a, b) if and only if its derivative f' is monotonically nondecreasing on (a, b);*

(ii) *A twice differentiable everywhere on (a, b) function f is convex on (a, b) if and only if its second derivative f'' is nonnegative everywhere on (a, b).*

With the Proposition, you can immediately verify that the functions listed as examples of convex functions in Section 3.1 indeed are convex. The only difficulty which you may meet is that some of these functions (e.g., x^p, $p \geq 1$, and $-x^p$, $0 \leq p \leq 1$, were claimed to be convex on the half-interval $[0, +\infty)$, while the Proposition speaks about convexity of functions on intervals. To overcome this difficulty, we will use the following simple

Proposition 3.2.2 *Let M be a convex set and f be a function with $\mathrm{Dom}\, f = M$. Assume that f is convex on $\mathrm{ri}\, M$ and is continuous on M, i.e.,*

$$f(x_i) \to f(x),\ i \to \infty,$$

whenever $x_i, x \in M$ and $x_i \to x$ as $i \to \infty$. Then f is convex on M.

In fact, for functions of one variable there is a differential criterion of convexity which does not "preassume" any smoothness (we will not prove this criterion):

Proposition 3.2.3 [convexity criterion for univariate functions]

Let $g : \mathbf{R} \to \mathbf{R} \cup \{+\infty\}$ be a function. Let the domain $\Delta = \{t \mid g(t) < \infty\}$ of the function be a convex set which is not a singleton, i.e., let it be an interval (a, b) with possibly added one or both endpoints ($-\infty \leq a < b \leq \infty$). g is convex if and only if it satisfies the following 3 requirements:

1) g is continuous on (a, b);

2) g is differentiable everywhere on (a, b), excluding, possibly, a countable set of points, and the derivative $g'(t)$ is nondecreasing on its domain;

3) at each endpoint u of the interval (a, b) which belongs to Δ g is upper semicontinuous:

$$g(u) \geq \limsup_{t \in (a,b), t \to u} g(t).$$

Proof of Proposition 3.2.2: Let $x, y \in M$ and $z = \lambda x + (1-\lambda)y$, $\lambda \in [0,1]$, and let us prove that

$$f(z) \leq \lambda f(x) + (1-\lambda)f(y).$$

As we know, there exist sequences $x_i \in \mathrm{ri}\, M$ and $y_i \in \mathrm{ri}\, M$ converging, respectively to x and to y. Then $z_i = \lambda x_i + (1-\lambda)y_i$ converges to z as $i \to \infty$, and since f is convex on $\mathrm{ri}\, M$, we have

$$f(z_i) \leq \lambda f(x_i) + (1-\lambda)f(y_i);$$

passing to limit and taking into account that f is continuous on M and x_i, y_i, z_i converge, as $i \to \infty$, to $x, y, z \in M$, respectively, we obtain the required inequality. ∎

From Propositions 3.2.1.(ii) and 3.2.2 we get the following convenient *necessary and sufficient* condition for convexity of a *smooth* function of n variables:

Corollary 3.2.1 [Convexity Criterion for smooth functions on $\mathbf{R}^n$]

Let $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ be a function. Assume that the domain M of f is a convex set and that f is

- *continuous on M*

  and

- *twice differentiable on* $\mathrm{ri}\, M$.

Then f is convex if and only if its Hessian is positive semidefinite on $\mathrm{ri}\, M$*:*

$$h^T f''(x) h \geq 0 \;\; \forall x \in \mathrm{ri}\, M \;\; \forall h \in \mathbf{R}^n.$$

Proof. The "only if" part is evident: if f is convex and $x \in M' = \mathrm{ri}\, M$, then the function of one variable

$$g(t) = f(x + th)$$

(h is an arbitrary fixed direction in $\mathbf{R}^n$) is convex in certain neighborhood of the point $t = 0$ on the axis (recall that affine substitutions of argument preserve convexity). Since f is twice differentiable in a neighborhood of x, g is twice differentiable in a neighborhood of $t = 0$, so that $g''(0) = h^T f''(x) h \geq 0$ by Proposition 3.2.1.∎

Now let us prove the "if" part, so that we are given that $h^T f''(x) h \geq 0$ for every $x \in \mathrm{ri}\, M$ and every $h \in \mathbf{R}^n$, and we should prove that f is convex.

Let us first prove that f is convex on the relative interior M' of the domain M. First, clearly, M' is a convex set. Since, as it was already explained, the convexity of a function on a convex set is one-dimensional fact, all we should prove is that every one-dimensional function

$$g(t) = f(x + t(y - x)), \;\; 0 \leq t \leq 1$$

(x and y are from M') is convex on the segment $0 \leq t \leq 1$. Since f is continuous on $M \supset M'$, g is continuous on the segment; and since f is twice continuously differentiable on M', g is continuously differentiable on $(0,1)$ with the second derivative

$$g''(t) = (y - x)^T f''(x + t(y - x))(y - x) \geq 0.$$

Consequently, g is convex on $[0,1]$ (Propositions 3.2.1.(ii) and 3.2.2). Thus, f is convex on M'. It remains to note that f, being convex on M' and continuous on M, is convex on M by Proposition 3.2.2. ■

Applying the combination rules preserving convexity to simple functions which pass the "infinitesimal' convexity tests, we can prove convexity of many complicated functions. Consider, e.g., an *exponential posynomial* – a function

$$f(x) = \sum_{i=1}^{N} c_i \exp\{a_i^T x\}$$

with positive coefficients c_i (this is why the function is called *posy*nomial). How could we prove that the function is convex? This is immediate:

$\exp\{t\}$ is convex (since its second order derivative is positive and therefore the first derivative is monotone, as required by the infinitesimal convexity test for smooth functions of one variable);

consequently, all functions $\exp\{a_i^T x\}$ are convex (stability of convexity under affine substitutions of argument);

consequently, f is convex (stability of convexity under taking linear combinations with nonnegative coefficients).

And if we were supposed to prove that the maximum of three posynomials is convex? Ok, we could add to our three steps the fourth, which refers to stability of convexity under taking pointwise supremum.

3.3 Lipschitz continuity of convex functions

Convex functions possess very nice local properties.

Theorem 3.3.1 * [Boundedness and Lipschitz continuity of convex function]

Let f be a convex function and let K be a closed and bounded set contained in the relative interior of the domain $\mathrm{Dom}\, f$ of f. Then f is Lipschitz continuous on K – there exists constant L – the Lipschitz constant of f on K – such that

$$|f(x) - f(y)| \le L|x - y| \quad \forall x, y \in K. \tag{3.3.3}$$

In particular, f is bounded on K.

Remark 3.3.1 All three assumptions on K – (1) closeness, (2) boundedness and the assumption (3) $K \subset \mathrm{ri}\ \mathrm{Dom}\, f$ – are essential, as it is seen from the following four examples:

- $f(x) = 1/x$, $\mathrm{Dom}\, F = (0, +\infty)$,$K = (0, 1]$. We have (2), (3) and not (1); f is neither bounded, nor Lipschitz continuous on K.
- $f(x) = x^2$, $\mathrm{Dom}\, f = \mathbf{R}$, $K = \mathbf{R}$. We have (1), (3) and not (2); f is neither bounded nor Lipschitz continuous on K.

- $f(x) = -\sqrt{x}$, $\mathrm{Dom}\, f = [0, +\infty)$, $K = [0, 1]$. We have (1), (2) and not (3); f is not Lipschitz continuous on K [1)], although is bounded.

- consider the following function of two variables:

$$f(x, y) = \begin{cases} 0 & \text{if} \quad x^2 + y^2 < 1, \\ \phi(x, y) & \text{if} \quad x^2 + y^2 = 1, \end{cases}$$

where ϕ is an *arbitrary* nonnegative function on the unit circle. The domain of this function is the unit disk, which is closed and convex. However, $f(x, y)$ is neither bounded nor Lipschitz continuous on $\mathrm{Dom}\, f$[2)].

Proof of Theorem 3.3.1. We will start with the following local version of the Theorem.

Proposition 3.3.1 *Let f be a convex function, and let $\bar{x}$ be a point from the relative interior of the domain* $\mathrm{Dom}\, f$ *of f. Then*

(i) *f is bounded at $\bar{x}$: there exists a positive r such that f is bounded in the r-neighborhood $U_r(\bar{x})$ of $\bar{x}$ in the affine hull of* $\mathrm{Dom}\, f$*:*

$$\exists r > 0, C : \quad |f(x)| \le C \quad \forall x \in U_r(\bar{x}) = \{x \in \mathrm{Aff}\,(\mathrm{Dom}\, f) \mid |x - \bar{x}| \le r\};$$

(ii) *f is Lipschitz continuous at $\bar{x}$, i.e., there exists a positive ρ and a constant L such that*

$$|f(x) - f(x')| \le L|x - x'| \;\forall x, x' \in U_\rho(\bar{x}).$$

Implication "Proposition 3.3.1 ⇒ Theorem 3.3.1" is given by standard Analysis reasoning. All we need is to prove that if K is a bounded and closed (i.e., a compact) subset of $\mathrm{ri}\, \mathrm{Dom}\, f$, then f is Lipschitz continuous on K (the boundedness of f on K is an evident consequence of its Lipschitz continuity on K and boundedness of K). Assume, on contrary, that f is not Lipschitz continuous on K; then for every integer i there exists a pair of points $x_i, y_i \in K$ such that

$$f(x_i) - f(y_i) \ge i|x_i - y_i|. \tag{3.3.4}$$

Since K is compact, passing to a subsequence we can ensure that $x_i \to x \in K$ and $y_i \to y \in K$. By Proposition 3.3.1 the case $x = y$ is impossible – by Proposition f is Lipschitz continuous in a neighborhood B of $x = y$; since $x_i \to x, y_i \to y$, this neighborhood should contain all x_i and y_i with large enough indices i; but then, from the Lipschitz continuity of f in B, the ratios $(f(x_i) - f(y_i))/|x_i - y_i|$ form a bounded sequence, which we know is not the case. Thus, the case $x = y$ is impossible. The case $x \ne y$ is "even less possible" – since, by Proposition, f is continuous on $\mathrm{Dom}\, f$ at both the points x and y (note that Lipschitz continuity at a point clearly implies the usual continuity at it), so that we would have $f(x_i) \to f(x)$ and $f(y_i) \to f(y)$ as $i \to \infty$.

[1)] indeed, we have $\lim_{t \to +0} \frac{f(0) - f(t)}{t} = \lim_{t \to +0} t^{-1/2} = +\infty$, while for a Lipschitz continuous f the ratios $t^{-1}(f(0) - f(t))$ should be bounded

[2)] note that in this case the function f is not closed. It is clear that to "close" the function f we should put $\phi(x, y) \equiv 0$

Thus, the left hand side in (3.3.4) remains bounded as $i \to \infty$. In the right hand side one factor – i – tends to ∞, and the other one has a nonzero limit $|x-y|$, so that the right hand side tends to ∞ as $i \to \infty$; this is the desired contradiction. ■

Proof of Proposition 3.3.1.

1^0. We start with proving the *above boundedness* of f in a neighborhood of $\bar{x}$. This is immediate: we know that there exists a neighborhood $U_{\bar{r}}(\bar{x})$ which is contained in $\mathrm{Dom}\, f$ (since, by assumption, $\bar{x}$ is a relative interior point of $\mathrm{Dom}\, f$). Now, we can find a small simplex Δ of the dimension $m = \dim \mathrm{Aff}\,(\mathrm{Dom}\, f)$ with the vertices $x_0, ..., x_m$ in $U_{\bar{r}}(\bar{x})$ in such a way that $\bar{x}$ will be a convex combination of the vectors x_i with *positive* coefficients, even with the coefficients $1/(m+1)$:

$$\bar{x} = \sum_{i=0}^{m} \frac{1}{m+1} x_i \quad {}^{3)}.$$

We know that $\bar{x}$ is the point from the relative interior of Δ (see the proof of Theorem 1.1.1.(ii)); since Δ spans the same affine set as $\mathrm{Dom}\, f$ (m is the dimension of $\mathrm{Aff}\,(\mathrm{Dom}\, f)$!), it means that Δ contains $U_r(\bar{x})$ with certain $r > 0$. Now, in

$$\Delta = \{\sum_{i=0}^{m} \lambda_i x_i \mid \lambda_i \geq 0, \sum_i \lambda_i = 1\}$$

f is bounded from above by the quantity $\max_{0 \leq i \leq m} f(x_i)$ by Jensen's inequality:

$$f(\sum_{i=0}^{m} \lambda_i x_i) \leq \sum_{i=0}^{m} \lambda_i f(x_i) \leq \max_i f(x_i).$$

Consequently, f is bounded from above, by the same quantity, in $U_r(\bar{x})$.

[3] As an exercise, let us verify that the required Δ exists. Let us act as follows: first, the case of $\mathrm{Dom}\, f$ being a singleton is evident, so that we can assume that $\mathrm{Dom}\, f$ is a convex set of dimension $m \geq 1$. Let us take arbitrary affine basis $y_0, ..., y_m$ in $M = \mathrm{Aff}\,(\mathrm{Dom}\, f)$ and then pass from this basis to the set $z_0 = y_0, z_1 = y_0 + \epsilon(y_1 - y_0), z_2 = y_0 + \epsilon(y_2 - y_0), ..., z_m = y_0 + \epsilon(y_m - y_0)$ with some $\epsilon > 0$. The vectors z_i clearly belong to M and form an affine basis (the latter follows from the fact that the vectors $z_i - z_0$, $i = 1, ..., m$, are ϵ times the vectors $y_i - y_0$; the latter vectors form a basis in the linear subspace L such that $M = y_0 + L$; consequently, the vectors $z_i - z_0$, $i = 1, ..., m$, also form a basis in L, whence, by the same Corollary, $z_0, ..., z_m$ form am affine basis in M). Choosing $\epsilon > 0$ small enough, we may enforce all the vectors $z_0, ..., z_m$ be in the $(\bar{r}/10)$-neighbourhood of the vector z_0. Now let Δ' be the convex hull of $z_0, ..., z_m$; this is a simplex with the vertices contained in the neighbourhood of z_0 of the radius $\bar{r}/10$ (of course, we are speaking about the ball in M). This neighbourhood is an intersection of a Euclidean ball, which is a convex set, and M, which also is convex; therefore the neighbourhood is convex. Since the vertices of Δ' are contained in it, the entire Δ' is contained in the neighbourhood. Now let $\bar{z} = (m+1)^{-1} \sum_{i=0}^{m} z_i$; Δ' clearly is contained in the $2 \times (\bar{r}/10) = \bar{r}/5$ neighbourhood of $\bar{z}$ in M. Setting $\Delta = [\bar{x} - \bar{z}] + \Delta'$, we get the simplex with the vertices $x_i = z_i + \bar{x} - \bar{z}$ which is contained in the $\bar{r}/5$-neighbourhood of $\bar{x}$ in M and is such that $(m+1)^{-1} \sum_{i=0}^{m} x_i \equiv (m+1)^{-1} \sum_{i=0}^{m} [z_i + \bar{x} - \bar{z}] = \bar{z} + \bar{x} - \bar{z} = \bar{x}$, as required.

I gave this awful "explanation" to demonstrate how many words we need to make rigorous "evident" recommendations like "let us take a small simplex with the average of vertices equal to $\bar{x}$". The "explanations" of this type should (and will) be omitted, since they kill even the most clear reasoning. Yet, you are supposed to be able to do this routine work by yourselves; and to this end you should remember what is the exact meaning of the words we are using and what are the basic relations between the corresponding concepts.

2^0. Now let us prove that if f is above bounded, by some C, in $U_r(\bar{x})$, then it in fact is below bounded in this neighborhood (and, consequently, is bounded in U_r). Indeed, let $x \in U_r$, so that $x \in \mathrm{Aff}\,(\mathrm{Dom}\, f)$ and $|x - \bar{x}| \leq r$. Setting $x' = \bar{x} - [x - \bar{x}] = 2\bar{x} - x$, we get $x' \in \mathrm{Aff}\,(\mathrm{Dom}\, f)$ and $|x' - \bar{x}| = |x - \bar{x}| \leq r$, so that $x' \in U_r$. Since $\bar{x} = \frac{1}{2}[x + x']$, we have

$$2f(\bar{x}) \leq f(x) + f(x'),$$

whence

$$f(x) \geq 2f(\bar{x}) - f(x') \geq 2f(\bar{x}) - C, \;\; x \in U_r(\bar{x}),$$

and f indeed is below bounded in U_r.

(i) is proved.

3^0. (ii) is an immediate consequence of (i). Indeed, let us prove that f is Lipschitz continuous in the neighborhood $U_{r/2}(\bar{x})$, where $r > 0$ is such that f is bounded in $U_r(\bar{x})$ (we already know from (i) that the required r does exist). Let $|f| \leq C$ in U_r, and let $x, x' \in U_{r/2}$, $x \neq x'$. Let us extend the segment $[x, x']$ through the point x' until it reaches, at certain point x'', the (relative) boundary of U_r; then we will get

$$x' \in (x, x''); \quad |x'' - \bar{x}| = r.$$

By the convexity of f we have for $\lambda = \frac{|x'-x|}{|x''-x|} \in (0,1)$,

$$f(x') - f(x) \leq \lambda(f(x'') - f(x)) \leq |x' - x| \frac{f(x'') - f(x)}{|x'' - x|}.$$

The second factor in the right hand side does not exceed the quantity $(2C)/(r/2) = 4C/r$; indeed, the numerator is, in absolute value, at most $2C$ (since $|f|$ is bounded by C in U_r and both x, x'' belong to U_r), and the denominator is at least $r/2$ (indeed, x is at the distance at most $r/2$ from $\bar{x}$, and x'' is at the distance exactly r from $\bar{x}$, so that the distance between x and x'', by the triangle inequality, is at least $r/2$). Thus, we have

$$f(x') - f(x) \leq (4C/r)|x' - x|, \;\; x, x' \in U_{r/2};$$

swapping x and x', we come to

$$f(x) - f(x') \leq (4C/r)|x' - x|,$$

whence

$$|f(x) - f(x')| \leq (4C/r)|x - x'|, \;\; x, x' \in U_{r/2},$$

as required in (ii). ■

Let us show that convex functions possess a kind of differentiability. In what follows, for the sake of conciseness, we will assume that the the domain of f is full=dimensional, meaning that $\mathrm{Aff}\,(\mathrm{Dom}\, f) = \mathbf{R}^n$. This said, you are expected to be able to translate the corresponding statements for the case when $\mathrm{Aff}\,(\mathrm{Dom}\, f)$ is smaller than the whole $\mathbf{R}^n$ (so that interior becomes relative interior, $h \in \mathbf{R}^n$ transforms into $h \in \mathcal{L}$, such that $\mathrm{Aff}\,(\mathrm{Dom}\, f) = a + \mathcal{L}$, etc.) .

Definition 3.3.1 [Directional derivative] *Let $x \in \mathrm{Dom}\, f$. We call the function f differentiable in the direction h at x if the following limit exists:*

$$f'_h(x) = \lim_{t \downarrow 0} \frac{f(x+ht) - f(x)}{t}. \tag{3.3.5}$$

Theorem 3.3.2 [Differentiability of convex functions] *Convex function f is differentiable in any direction $h \in \mathbf{R}^n$ at any point $x \in \mathrm{int}\, \mathrm{Dom}\, f$.*

Proof: Let $x \in \mathrm{int}\, \mathrm{Dom}\, f$. Consider the function

$$\phi(t) = \frac{f(x+ht) - f(x)}{t}, \quad t > 0.$$

Let $\lambda \in (0,1]$ and let $t \in (0,\epsilon]$ be small enough to have $x + \epsilon h \in \mathrm{int}\, \mathrm{Dom}\, f$. Then by convexity of f,

$$f(x + \lambda ht) = f((1-\lambda)x + \lambda(x+ht)) \le (1-\lambda)f(x) + \lambda f(x+ht).$$

Therefore

$$\phi(\lambda t) = \frac{1}{\lambda t}[f(x+\lambda ht) - f(x)] \le \frac{1}{t}[f(x+ht) - f(x)],$$

so that $\phi(t)$ decreases as $t \downarrow 0$ and due to the local Lipschitz property is below bounded so that the limit in (3.3.5) exists. ∎

Theorem 3.3.1 says that a convex function f is bounded on every compact (i.e., closed and bounded) subset of the interior of $\mathrm{Dom}\, f$. In fact there is much stronger statement on the *below* boundedness of f: f is below bounded on *any* bounded subset of $\mathbf{R}^n$! This results is an immediate consequence of the following lower bound:

Proposition 3.3.2 [Global lower bound] *Let f be a convex function and $x \in \mathrm{int}\, \mathrm{Dom}\, f$. Then $f'_h(x)$ is convex positive homogenous (of degree 1) function of h, and for any $y \in \mathrm{Dom}\, f$*

$$f(y) \ge f(x) + f'_{y-x}(x). \tag{3.3.6}$$

Proof: Let us prove first that the directional derivative is homogenous. Indeed, for any $h \in \mathbf{R}^n$ and $\tau > 0$

$$f'_{\tau h}(x) = \lim_{t \downarrow 0} \frac{f(x+\tau ht) - f(x)}{t} = \tau \lim_{\alpha \downarrow 0} \frac{f(x+h\alpha) - f(x)}{\alpha} = \tau f'_h(x).$$

Further, for any h_1, $h_2 \in \mathbf{R}^n$, and $\lambda \in [0,1]$, by the convexity of f we get

$$\begin{aligned} f'_{\lambda h_1 + (1-\lambda)h_2}(x) &= \lim_{t \downarrow 0} \frac{1}{t}[f(x + (\lambda h_1 + (1-\lambda)h_2)t) - f(x)] \\ &\le \lim_{t \downarrow 0} \frac{1}{t}\left\{\lambda[f(x+th_1) - f(x)] + (1-\lambda)[f(x+th_2) - f(x)]\right\} \\ &= \lambda f'_{h_1}(x) + (1-\lambda) f'_{h_2}(x). \end{aligned}$$

Thus $f'_h(x)$ is convex in h. Finally, let $t \in (0,1]$, $y \in \mathrm{Dom}\, f$ and

$$y_t = x + t(y - x) = (1-t)x + ty.$$

Then

$$f(y) = f\left(y_t + \frac{1}{t}(1-t)(y_t - x)\right) \geq f(y_t) + \frac{1-t}{t}[f(y_t) - f(x)],$$

(the latter inequality is nothing but the convexity property of f: $f(y_t) \leq tf(y) + (1-t)f(x)$) and we conclude (3.3.6) taking the limit as $t \downarrow 0$.■

3.4 Subgradients of convex functions

We are now ready to introduce a "good surrogate" of the notion of the gradient for a convex function. Let f be a convex function, and let $x \in \mathrm{Dom}\, f$. It may happen that there exists an affine minorant $d^T x - a$ of f which coincides with f at x:

$$f(y) \geq d^T y - a \quad \forall y, \quad f(x) = d^T x - a.$$

From the equality in the latter relation we get $a = d^T x - f(x)$, and substituting this representation of a into the first inequality, we get

$$f(y) \geq f(x) + d^T(y - x) \quad \forall y. \tag{3.4.7}$$

Thus, if f admits an affine minorant which is exact at x, then there exists d which gives rise to inequality (3.4.7). Vice versa, if d is such that (3.4.7) takes place, then the right hand side of (3.4.7), regarded as a function of y, is an affine minorant of f which is exact at x.

Now note that (3.4.7) express certain property of a vector d.

Definition 3.4.1 [Subgradient of convex function] *Let f be a convex function. A vector h is called* subgradient *of function f at a point $x \in \mathrm{Dom}\, f$ if for any $y \in Dom f$ we have*

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

The set $\partial f(x)$ of all subgradients of f at x is called the subdifferential *of f at the point x.*

Subgradients of convex functions play an important role in the theory and numerical methods of Convex Programming – they are quite reasonable surrogates of the gradients. The most elementary properties of the subgradients are summarized in the following statement:

Proposition 3.4.1 *Let f be a convex function and x be a point from* int Dom f. *Then*

(i) *$\partial f(x)$ is a closed convex set which for sure is nonempty and bounded*

(ii) *for any $h \in \mathbf{R}^n$*

$$f'_h(x) = \max\{h^T d \mid d \in \partial f(x)\}.$$

In other words, the directional derivative is nothing but the support function of the set ∂f.

(iii) *If f is differentiable at x, then $\partial f(x)$ is the singleton comprised of the usual gradient of f at x.*

Proof. (i): Closeness and convexity of $\partial f(x)$ are evident – (3.4.7) is an infinite system of nonstrict linear inequalities with respect to d, the inequalities being indexed by $y \in \mathbf{R}^n$. Nonemptiness of $\partial f(x)$ for the case when $x \in \operatorname{int} \operatorname{Dom} f$ – this is the most important fact about the subgradients – is readily given by our preceding results. Indeed, we should prove that if $x \in \operatorname{int} \operatorname{Dom} f$, then there exists an affine minorant of f which is exact at x. Note that the point $(f(x), x)$ belongs to the boundary of epi (f). Hence, in view of Theorem 1.2.1 there is a linear form $(-\alpha, d)$ which properly separates x and epi (f):

$$d^T y - \alpha\tau \le d^T x - \alpha f(x)$$

for any $(\tau, y) \in$ epi (f). Note that we can take

$$|d|^2 + \alpha^2 = 1 \tag{3.4.8}$$

and such that $(-\alpha, d) \in \operatorname{Aff}(\operatorname{epi}(f))$. And since for any $\tau \ge f(x)$ the point (τ, x) belongs to the epigraph of f, we conclude that $\alpha \ge 0$.

Now recall that a convex function is locally Lipschitz continuous on the interior of its domain (Proposition 3.3.1). This means that there exist some $\epsilon > 0$ and $M > 0$ such that the ball of the radius ϵ, centered at x belongs to $\operatorname{Dom} f$ and for any y in this ball

$$f(y) - f(x) \le M|y - x|.$$

Thus, for any y in the ball

$$d^T(y - x) \le \alpha(f(y) - f(x)) \le \alpha M|y - x|.$$

When choosing $y = x + \epsilon d$ we get $|d|^2 \le M\alpha|d|$ and in view of the normalizing equation (3.4.8),

$$\alpha \ge \frac{1}{\sqrt{1 + M^2}}.$$

Now we can choose $h = d/\alpha$ to get

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

Finally, if $h \in \partial f(x)$, $h \ne 0$, then by choosing $y = x + \epsilon h/|h|$ we obtain:

$$\epsilon|h| = h^T(y - x) \le f(y) - f(x) \le M|y - x| = M\epsilon,$$

what implies the boundedness of $\partial f(x)$.

(ii) Note that (as $f_0'(x) = 0$)

$$f_h'(x) - f_0'(x) = f_h'(x) = \lim_{t \downarrow 0} \frac{f(x + ht) - f(x)}{t} \ge h^T d, \tag{3.4.9}$$

for any vector d from $\partial f(x)$. Therefore, the subdifferential of the function $f_h'(x)$ at $h = 0$ exists and

$$\partial f(x) \subset \partial_h f_0'(x). \tag{3.4.10}$$

As $f'_h(x)$ is convex in h (cf Proposition 3.3.2) so that

$$f'_{y-x}(x) = f'_{y-x}(x) - f'_0(x) \geq d^T(y-x)$$

for any $d \in \partial_h f'_0(x)$, and by Proposition 3.3.2 we have for any $y \in \mathrm{Dom}\, f$,

$$f(y) \geq f(x) + f'_{y-x}(x) \geq f(x) + d^T(y-x) \;\text{ for }\; d \in \partial_h f'_0(x).$$

We conclude that $\partial_h f'_0(x) \subset \partial f(x)$ and by (3.4.10), $\partial_h f'_0(x) \equiv \partial f(x)$.

Let now $d_h \in \partial_h f'_h(x)$. Then for any $v \in \mathbf{R}^n$ and $\tau > 0$ we have

$$\tau f'_v(x) = f'_{\tau v}(x) \geq f'_h(x) + d_h^T(\tau v - h),$$

so that when $\tau \to \infty$ we obtain $f'_v(x) \geq d_h^T v$ what means that $d_h \in \partial_h f'_0(x)$.

On the other hand, when $\tau \to 0$ we get $f'_h(x) - d_h^T h \leq 0$ and by (3.4.9) we conclude that $d_h^T h = f'_h(x)$, what implies (ii).

(iii): If $x \in \mathrm{int}\,\mathrm{Dom}\, f$ and f is differentiable at x, then $\nabla f(x) \in \partial f(x)$ by Proposition 3.3.2 (notice that $f'_p(x) = p^T \nabla f(x)$ in this case). To prove that $\nabla f(x)$ is the only subgradient of f at x, note that if $d \in \partial f(x)$, then, by definition,

$$f(y) - f(x) \geq d^T(y-x) \quad \forall y$$

Substituting $y - x = th$, h being a fixed direction and t being > 0, dividing both sides of the resulting inequality by t and passing to limit as $t \to +0$, we get

$$h^T \nabla f(x) \geq h^T d.$$

This inequality should be valid for all $h \in \mathbf{R}^n$, which is possible if and only if $d = \nabla f(x)$. ∎

Proposition 3.4.1 explains why subgradients are good surrogates of gradients: at a point where the gradient exists, it is the only subgradient, but, in contrast to the gradient, a subgradient exists basically everywhere (for sure in the interior of the domain of the function). E.g., let us look at the simple function

$$f(x) = |x|$$

on the axis. It is, of course, convex (as maximum of two linear forms x and $-x$). Whenever $x \neq 0$, f is differentiable at x with the derivative $+1$ for $x > 0$ and -1 for $x < 0$. At the point $x = 0$ f is not differentiable; nevertheless, it must have subgradients at this point (since 0 is an interior point of the domain of the function). And indeed, it is immediately seen (why?) that the subgradients of $|x|$ at $x = 0$ are exactly the reals from the segment $[-1, 1]$. Thus,

$$\partial|x| = \begin{cases} \{-1\}, & x < 0, \\ [-1, 1], & x = 0, \\ \{+1\}, & x > 0. \end{cases}$$

Note also that if x is a boundary point of the domain of a convex function, even a "good" one, the set of subgradients of f at x may be empty, as it is the case with the function

$$f(y) = \begin{cases} -\sqrt{y}, & y \geq 0, \\ +\infty, & y < 0; \end{cases}$$

it is clear that there is no non-vertical supporting line to the epigraph of the function at the point $(0, f(0))$, and, consequently, there is no affine minorant of the function which is exact at $x = 0$.

It is worth to note that the subgradient property is characteristic of convex functions. We have also the following simple

Lemma 3.4.1 *Let a continuous function f be such that for any $x \in \operatorname{int} \operatorname{Dom} f$ the subdifferential $\partial f(x)$ is not empty. Then f is convex.*

Proof takes a line: let $x, y \in \operatorname{int} \operatorname{Dom} f$, $0 \le \lambda \le 1$ and $z = x + \lambda(y - x)(\in \operatorname{int} \operatorname{Dom} f)$. Let $h \in \partial f(z)$, then

$$\begin{array}{rcl} f(y) & \geq & f(z) + h^T(y - z) = f(z) + (1 - \lambda)h^T(y - x); \\ f(x) & \geq & f(z) + h^T(x - z) = f(z) - \lambda h^T(y - x), \end{array}$$

and

$$\lambda f(y) + (1 - \lambda) f(x) \ge f(z).$$

We conclude now by Proposition 3.2.2. ■

Subgradient calculus

As we already know from Proposition 3.4.1.(ii), the directional derivative $f'_h(x)$ is the support function of the subdifferential $\partial f(x)$. This basic observation is the basis of our future developments. A part of Convex Analysis deals with *subgradient calculus* – rules for computing subgradients of "composite" functions, like sums, superpositions, maxima, etc., given subgradients of the operands. For instance, the list of the operations preserving convexity of Section 3.2.1 can be completed with the corresponding rules of subgradient computations. These rules extend onto nonsmooth convex case the standard Calculus rules and are very nice and instructive. We present here some of them:

- [taking weighted sums] if f, g are convex functions on $\mathbf{R}^n$ and $\lambda, \mu > 0$ then the subgradient of the function $h(x) = \lambda f(x) + \mu g(x)$ satisfies

$$\partial h(x) = \lambda \partial f(x) + \mu \partial g(x) \tag{3.4.11}$$

  for any $x \in \operatorname{int} \operatorname{Dom} h$. [Let $x \in \operatorname{int} \operatorname{Dom} f \cap \operatorname{int} \operatorname{Dom} g$. Then for any $h \in \mathbf{R}^n$ we have by Proposition 3.4.1.(ii):

$$\begin{array}{rcl} f'_h(x) & = & \lambda f'_h(x) + \mu g'_h(x) \\ & = & \max\{\lambda h^T d_1 |\, d_1 \in \partial f(x)\} + \max\{\mu h^T d_2 |\, d_2 \in \partial g(x)\} \\ & = & \max\{h^T(\lambda d_1 + \mu d_2) |\, d_1 \in \partial f(x),\, d_2 \in \partial g(x)\} \\ & = & \max\{h^T d |\, d \in \lambda \partial f(x) + \mu \partial g(x)\} \end{array}$$

  Using Corollary 1.2.1 we obtain (3.4.11).]

- [affine substitution of the argument] let the function $f(y)$ be convex with $\mathrm{Dom}\, f \subset \mathbf{R}^m$. Consider the affine operator

$$\mathcal{A}(x) = Ax + b \mid \mathbf{R}^n \to \mathbf{R}^m$$

and $\phi(x) = f(\mathcal{A}(x))$.

Then for any $x \in \mathrm{Dom}\,\phi = \{x \in \mathbf{R}^n \mid \mathcal{A}(x) \in \mathrm{Dom}\, f\}$

$$\partial\phi(x) = A^T \partial f(\mathcal{A}(x)).$$

[The proof of the subdifferential relation is immediate: if $y = \mathcal{A}(x)$ then for any $h \in \mathbf{R}^n$ we have

$$\phi'_h(x) = f'_{Ah}(y) = \max\{d^T Ah \mid d \in \partial f(y)\} = \max\{\bar{d}^T h \mid \bar{d} \in A^T \partial f(y)\}.$$

Now by Proposition 3.4.1.(ii) and Corollary 1.2.1 we get $\partial\phi(x) = A^T\partial f(\mathcal{A}(x))$.]

- [taking pointwise sup] let f be an upper bound $\sup_i f_i$ of a family of convex functions on $\mathbf{R}^n$. Then its subgradient at any $x \in \mathrm{int\,Dom}\, f$ satisfies

$$\partial f = \mathrm{Conv}\,\{\partial f_i \mid i \in I(x)\}, \qquad (3.4.12)$$

where $I(x) = \{i : \ f_i(x) = f(x)\}$ the set of functions f_i which are "active at x.

[let $x \in \cap_i \mathrm{int\,Dom}\, f_i$, and assume that $I(x) = 1, .., k$. Then for any $h \in \mathbf{R}^n$ we have by Proposition 3.4.1.(ii)

$$f'_h(x) = \max_{1 \le i \le k} f'_{ih}(x) = \max_{1 \le i \le k} \max\{h^T d_i \mid d_i \in \partial f_i(x)\}.$$

Note that for any numbers $a_1, ..., a_k, ...$ we have

$$\max a_i = \max\{\sum_i \lambda_i a_i \mid \lambda \in \Delta_k\},$$

where $\Delta_k = \{\lambda \ge 0,\ \sum_i \lambda_i = 1\}$ is the *standard simplex* in $\mathbf{R}^k$. Thus

$$\begin{aligned} f'_h(x) &= \max_{\lambda \in \Delta_k} \sum_{i=1}^k \lambda_i \max\{h^T d_i \mid d_i \in \partial f_i(x)\} \\ &= \max\left\{ h^T \left(\sum_{i=1}^k \lambda_i d_i\right) \mid d_i \in \partial f_i(x),\ \lambda \in \Delta_k \right\} \\ &= \max\left\{ h^T d \mid d = \sum_{i=1}^k \lambda_i d_i,\ d_i \in \partial f_i(x),\ \lambda \in \Delta_k \right\} \\ &= \max\left\{ h^T d \mid d \in \mathrm{Conv}\,\{\partial f_i(x),\ 1 \le i \le k\right\}.\Big] \end{aligned}$$

Unfortunately, the above rule does not have a "closed form". However, it allows to compute elements of the subgradient.

- One more rule, which we state without proof, is as follows:

  Lemma 3.4.2 *Let f be an upper bound $\sup_{\alpha\in\mathcal{A}} f_\alpha$ of an arbitrary family $\mathcal{A}$ of convex and closed functions. Then for any x from $\mathrm{Dom}\, f = \cap_\alpha \mathrm{Dom}\, f_\alpha$*

  $$\partial f(x) \supset \mathrm{Conv}\,\{\partial f_\alpha(x) | \alpha \in \alpha(x)\},$$

  where $\alpha(x) = \{\alpha |\, f_\alpha(x) = f(x)\}$ Furthermore, if the set $\mathcal{A}$ is compact (in some metric) and the function $\alpha \to f_\alpha(x)$ is closed, then

  $$\partial f(x) = \mathrm{Conv}\,\{\partial f_\alpha(x) | \alpha \in \alpha(x)\},$$

Example 3.4.1

- *Consider the function*

  $$f(x) = \sum_{i=1}^m |a_i^T x - b_i|.$$

  Let for $x \in \mathbf{R}^n$,

  $$\begin{aligned} I_-(x) &= \{i : a_i^T x - b_i < 0\},\\ I_+(x) &= \{i : a_i^T x - b_i > 0\},\\ I_0(x) &= \{i : a_i^T x - b_i = 0\}. \end{aligned}$$

  Then

  $$\partial f(x) = -\sum_{i\in I_-(x)} a_i + \sum_{i\in I_+(x)} a_i + \sum_{i\in I_0(x)} [-a_i, a_i].$$

- *Consider the function $f(x) = \max_{1\le i\le n} x^{(i)}$ (here $x^{(i)}$ are the components of x) and denote $I(x) = \{i : x^{(i)} = f(x)\}$. Then $\partial f(x) = \mathrm{Conv}\,\{e_i | \, i \in I(x)\}$ (here e_i are the orths of the canonical basis of $\mathbf{R}^n$). In particular, for $x = 0$ the subdifferential is* the standard simplex – *a convex hull of the origin and canonical orths: $\partial f(x) =$* $\mathrm{Conv}\,\{e_i | \, 1 \le i \le n\}$.

- *For the Euclidean norm $f(x) = |x|$ we have*

  $$\partial f(x) = \begin{cases} \left\{\frac{x}{|x|}\right\} & \text{for } x \ne 0,\\ B_2(0,1) = \{x \in \mathbf{R}^n | \, |x| \le 1\} & \text{for } x = 0. \end{cases}$$

- *For the l_∞-norm $f(x) = |x|_\infty = \max_{1\le i\le n} |x^{(i)}|$ we have*

  $$\partial f(x) = \mathrm{Conv}\,\{\{e_i | \, i \in I_+(x)\} \cup \{-e_j | \, j \in I_-(x)\}\},$$

  where $I_+(x) = \{i : x^{(i)} = |x|_\infty\}$, $I_-(x) = \{i : -x^{(i)} = |x|_\infty\}$. In particular, $\partial f(0) =$ $B_1(0,1) = \{x \in \mathbf{R}^n | \, |x|_1 \le 1\}$.

- *For the l_1-norm $f(x) = |x|_1 = \sum_{i=1}^n |x^{(i)}|$ we have*
$$\partial f(x) = \sum_{i \in I_+} e_i - \sum_{i \in I_-} e_i + \sum_{i \in I_0} [-e_i, e_i].$$
where $I_+(x) = \{i : x^{(i)} > 0\}$, $I_-(x) = \{i : x^{(i)} < 0\}$ and $I_0(x) = \{i : x^{(i)} = 0\}$. In particular, $\partial f(0) = B_\infty(0,1) = \{x \in \mathbf{R}^n |\, |x|_\infty \leq 1\}$.

- *Maximum eigenvalue of a symmetric matrix. $f(x) = \lambda_{\max}(A(x))$, where*
$$A(x) = A_0 + x_1 A_1 + ... + x_n A_n,$$
with $m \times m$ symmetric matrices $A_1, ..., A_n$ $m \times m$ and $x \in \mathbf{R}^n$. Here $\lambda_{\max}(A)$ stands for the maximal eigenvalue of A. We can express f as the pointwise supremum of convex functions, using Rayleigh's variational definition of the maximal eigenvalue of a symmetric matrix A: $\lambda_{\max}(A(x)) = \sup_{|y|=1} y^T A y$. Here the index set $\mathcal{A}$ is the unit sphere: $\mathcal{A} = \{y \in \mathbf{R}^n |\, |y| = 1\}$. Each of the functions $f_y(x) = y^T A(x) y$ is affine in x for fixed y, as can be easily seen from
$$y^T A(x) y = y^T A_0 y + x_1 y^T A_1 y + ... + x_n y^T A_n y,$$
so it is differentiable with gradient $\nabla f_y(x) = (y^T A_1 y, ..., y^T A_n y)$.
The active functions $y^T A(x) y$ are those associated with the eigenvectors y corresponding to the maximum eigenvalue. Hence to find a subgradient, *it suffices to compute compute an eigenvector y with eigenvalue $\lambda_{\max}$, normalized to have unit norm, and take*
$$g = (y^T A_1 y, ..., y^T A_n y).$$
Therefore, by Lemma 3.4.2 we come to
$$\partial f(x) = \mathrm{Conv}\, \{\nabla f_y(x) |\, A(x) y = \lambda_{\max}(A(x)) y,\ |y| = 1\}.$$

3.5 Optimality conditions

As it was already mentioned, optimization problems involving convex functions possess nice theoretical properties. One of the most important of these properties is given by the following

Theorem 3.5.1 ["Unimodality"] *Let f be a convex function on a convex set $M \subset \mathbf{R}^n$, and let $x^* \in M \cap \mathrm{Dom}\, f$ be a local minimizer of f on M:*
$$(\exists r > 0): \quad f(y) \geq f(x^*) \quad \forall y \in M,\ |y - x| < r. \tag{3.5.13}$$
Then x^* is a global minimizer of f on M:
$$f(y) \geq f(x^*) \quad \forall y \in M. \tag{3.5.14}$$
Moreover, the set $\mathrm{Argmin}_M\, f$ of all local ($\equiv$ global) minimizers of f on M is convex.

If f is strictly convex (i.e., the convexity inequality $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$ is strict whenever $x \neq y$ and $\lambda \in (0,1)$), then the above set is either empty or is a singleton.

Proof. 1) Let x^* be a local minimizer of f on M and $y \in M$, $y \neq x^*$; we should prove that $f(y) \geq f(x^*)$. There is nothing to prove if $f(y) = +\infty$, so that we may assume that $y \in \mathrm{Dom}\, f$. Note that also $x^* \in \mathrm{Dom}\, f$ for sure – by definition of a local minimizer.

By the convexity of f, for all $\tau \in (0,1)$ we have for $x_\tau = \tau y + (1-\tau)x^*$:

$$f(x_\tau) - f(x^*) \leq \tau(f(y) - f(x^*)).$$

Since x^* is a local minimizer of f, the left hand side in this inequality is nonnegative for all small enough values of $\tau > 0$. We conclude that the right hand side is nonnegative, i.e., $f(y) \geq f(x^*)$. ■

2) To prove convexity of $\mathrm{Argmin}_Q f$, note that $\mathrm{Argmin}_M f$ is nothing but the level set $\mathrm{lev}_\alpha(f)$ of f associated with the minimal value $\min_M f$ of f on M; as a level set of a convex function, this set is convex (Proposition 3.1.4).

3) To prove that the set $\mathrm{Argmin}_Q f$ associated with a strictly convex f is, if nonempty, a singleton, note that if there were two distinct minimizers x', x'', then, from strict convexity, we would have

$$f(\frac{1}{2}x' + \frac{1}{2}x'') < \frac{1}{2}[f(x') + f(x'')] = \min_M f,$$

which clearly is impossible - the argument in the left hand side is a point from M! ■

Another pleasant fact is the following

Theorem 3.5.2 [Necessary and sufficient condition of optimality for convex functions] *A point $x^* \in \mathrm{Dom}\, f$ is the minimizer of $f(x)$ if and only if $0 \in \partial f(x^*)$.*

Proof takes two lines: indeed when $0 \in \partial f(x^*)$, by the definition of the subgradient,

$$f(x) \geq f(x^*) + 0^T(x - x^*) = f(x^*)$$

for any $x \in \mathrm{Dom}\, f$. On the other hand, if $f(x) \geq f(x^*)$ for any $x \in \mathrm{Dom}\, f$, then by the definition of the subgradient, $0 \in \partial f(x^*)$.■

The simple optimality condition given in Theorem 3.5.2 becomes the celebrated Fermat rule in the case when the function is differentiable at $x^* \in \mathrm{int}\,\mathrm{Dom}\, f$. Indeed, by Proposition 3.4.1.(iii) in this case $\partial f(x^*) = \{\nabla f(x^*)\}$ and the rule becomes simply

$$\nabla f(x^*) = 0.$$

A natural question is what happens

if x^* in the above statement is not necessarily an interior point of $M = \mathrm{Dom}\, f$?

In order to provide a (partial) answer to this question let us consider the case of convex function f on $M \subset \mathbf{R}^n$, which is differentiable at $x^* \in M$. Under these assumptions when x^* is a minimizer of $f(x)$ on M?

To continue we need to define a new object: let M be a (nonempty) convex set, and let $x^* \in M$. *The tangent cone of M at x^** is the cone

$$T_M(x^*) = \{h \in \mathbf{R}^n \mid x^* + th \in M \quad \forall \text{ small enough } t > 0\}.$$

Geometrically, this is the set of all directions leading from x^* inside M, so that a small enough positive step from x^* along the direction keeps the point in M. From the convexity of M it immediately follows that the tangent cone indeed is a (convex) cone (not necessary closed). E.g., when x^* is an interior point of M, then the tangent cone to M at x^* clearly is the entire $\mathbf{R}^n$. A more interesting example is the tangent cone to a polyhedral set

$$M = \{x \mid Ax \leq b\} = \{x \mid a_i^T x \leq b_i,\ i = 1, ..., m\}; \tag{3.5.15}$$

for $x^* \in M$ the corresponding tangent cone clearly is the polyhedral cone

$$\{h \mid a_i^T h \leq 0 \quad \forall i : a_i^T x^* = b_i\} \tag{3.5.16}$$

corresponding to the *active* at x^* (i.e., satisfied at the point as equalities rather than as strict inequalities) constraints $a_i^T x \leq b_i$ from the description of M.

Now, for the functions in question (i.e., convex on M and differentiable at x^*) the necessary and sufficient condition of Theorem 3.5.2 for x^* to be a minimizer of f on M reads as follows:

(*) *the derivative of f taken at x^* along every direction from $T_M(x^*)$ should be nonnegative:*

$$h^T \nabla f(x^*) \geq 0 \quad \forall h \in T_M(x^*). \tag{3.5.17}$$

Note that the necessity of (3.5.17) is an evident fact which has nothing in common with convexity: assuming that x^* is a local minimizer of f on M, we note that if there were $h \in T_M(x^*)$ with $h^T \nabla f(x^*) < 0$, then we would have

$$f(x^* + th) < f(x^*)$$

for all small enough positive t. On the other hand, $x^* + th \in M$ for all small enough positive t due to $h \in T_M(x^*)$. Combining these observations, we conclude that in every neighborhood of x^* there are points from M with strictly better than the one at x^* values of f; this contradicts the assumption that x^* is a local minimizer of f on M.

The sufficiency is immediately given by the (sub-)gradient inequality: if $x \in M$, then $h = x - x^* \in T_M(x^*)$ and

$$f(x) \geq f(x^*) + (x - x^*)\nabla f(x^*) \geq f(x^*). \ \blacksquare$$

Condition (*) says that whenever f is convex on M and differentiable at $x^* \in M$, the necessary and sufficient condition for x^* to be a minimizer of f on M is that the linear form given by the gradient $\nabla f(x^*)$ of f at x^* should be nonnegative at all directions from the tangent cone $T_M(x^*)$. *The linear forms nonnegative at all directions from the tangent cone also form a cone; it is called the* cone normal to M at x^* *and is denoted $N_M(x^*)$.* Thus, (*) says that the necessary and sufficient condition for x^* to minimize f on M is the inclusion $\nabla f(x^*) \in N_M(x^*)$. What does this condition actually mean, it depends on what is the normal cone: whenever we have an explicit description of it, we have an explicit form of the optimality condition.

E.g., when $T_M(x^*) = \mathbf{R}^n$ (it is the same as to say that x^* is an interior point of M), then the normal cone is comprised of the linear forms nonnegative at the entire space, i.e., it is the trivial cone $\{0\}$; consequently, for the case in question the optimality condition becomes the Fermat rule $\nabla f(x^*) = 0$, as we already know.

When M is the polyhedral set (3.5.15), the tangent cone is the polyhedral cone (3.5.16); it is comprised of all directions which have nonpositive inner products with all a_i coming from the active, in the aforementioned sense, constraints. The normal cone is comprised of all vectors which have nonnegative inner products with all these directions, i.e., of

all vectors a such that the inequality $h^T a \geq 0$ is a consequence of the inequalities $h^T a_i \leq 0$, $i \in I(x^*) \equiv \{i \mid a_i^T x^* = b_i\}$.

From the Homogeneous Farkas Lemma (Lemma 2.5.1) we conclude that the normal cone $N_M(x^*)$ is simply the conic hull of the vectors $-a_i$, $i \in I(x^*)$:

$$N_M(x^*) = \{z \in \mathbf{R}^n \mid z = -\sum_{i \in I(x^*)} \lambda_i a_i, \ \lambda_i \geq 0\}.$$

Thus, in the case in question (*) reads:

$x^* \in M$ is a minimizer of f on M if and only if there exist nonnegative reals λ_i^* associated with "active" (those from $I(x^*)$) values of i such that

$$\nabla f(x^*) + \sum_{i \in I(x^*)} \lambda_i^* a_i = 0.$$

The next result is of main importance in optimization – it is the basis of the cutting plane scheme we consider in the sequel.

Theorem 3.5.3 [Support vectors for level sets] *For any $x \in \operatorname{Dom} f$, all vectors $d \in \partial f(x)$ satisfy*

$$d^T(x - y) \geq 0, \quad \textit{for any } y \in \operatorname{lev}_{f(x)}(f),$$

where $\operatorname{lev}_{f(x)}(f)$ is the level set of f:

$$\operatorname{lev}_{f(x)}(f) = \{y \in \operatorname{Dom} f \mid f(y) \leq f(x)\}$$

(we say that such vectors d are supporting *to the set $\operatorname{lev}_{f(x)}(f)$ at x).*

Proof: indeed, if $f(y) \leq f(x)$ and $d \in \partial f(x)$, then $f(x) + d^T(y - x) \leq f(y) \leq f(x)$. ■

We have the following trivial though useful

Corollary 3.5.1 *Let $M \subset \operatorname{Dom} f$ be a convex closed set, $x \in M$ and*

$$x^* = \operatorname*{argmin}_{x \in M} f(x).$$

Then for any $d \in \partial f(x)$ we have: $d^T(x - x^*) \geq 0$.

We conclude for now our study of minima of a convex function with the following optimality condition for constrained optimization:

Theorem 3.5.4 [Karush-Kuhn-Tucker] *Let f, g_i $i = 1, ..., m$ be differentiable convex functions. Suppose that there exists a point $\bar{x}$ such that*

$$\text{(Slater condition)} \quad g_i(\bar{x}) < 0, \;\; i = 1, ..., m.$$

A point x^* is a solution to the optimization problem

$$\min f(x) \;\; s.t. \;\; g_i(x) \le 0, \;\; i = 1, ..., m \tag{3.5.18}$$

if and only if there exist nonnegative real λ_i, i=1,...,m such that

$$\nabla f'(x^*) + \sum_{i=1}^{m} \lambda_i g_i'(x^*) = 0$$

and

$$\text{(Complementary slackness)} \quad \lambda_i g_i(x^*) = 0, \;\; i = 1, ..., m.$$

Proof: we reduce the constrained problem (3.5.18) to an unconstrained though not smooth optimization problem. To this end look at the following construction: consider the *parametric maxtype function*

$$f(t; x) = \max\{f(x) - t;\; g_i,\; i = 1, ..., m\}$$

and the function $h(t) = \min_x f(t; x)$.

Lemma 3.5.1 *Let f^* be the optimal value of the optimization problem (3.5.18). Then*

$$h(t) \le 0 \;\; \text{for all} \;\; t \ge f^*;$$
$$h(t) > 0 \;\; \text{for all} \;\; t < f^*.$$

Proof of the lemma: let x^* be an optimal solution to (3.5.18). If $t \ge f^* = f(x^*)$ then

$$h(t) \le f(t; x^*) = \max\{f(x^*) - t;\, g_i(x^*)\} \le \max\{f^* - t; g_i(x^*)\} \le 0.$$

On the other hand, if $t < f^*$ and $h(t) \le 0$ then there exists $x \in \mathbf{R}^n$ such that

$$f(x) \le t < f^*, \;\; \text{and} \;\; g_i(x) \le 0, \;\; i = 1, ..., m.$$

Thus, f^* cannot be the optimal value of (3.5.18). ∎

Now we are done: indeed, in view of the lemma, x^* is an optimal solution of the problem (3.5.18) if and only if it is a global minimizer of the function

$$\phi(x) = \max\{f(x) - f^*;\; g_i,\; i = 1, ..., m\}$$

We have from Theorem 3.5.2 that this is the case if and only if $0 \in \partial\phi(x^*)$. As we know this is the case (cf. the corresponding rule (3.4.12)) if 0 belongs to the convex envelope of those among the vectors $f'(x^*), f_1'(x^*), ..., f_m'(x^*)$ which are active at x^*, i.e.

$$\mu_0 f'(x^*) + \sum_{i\in I^*} \mu_i g_i'(x^*) = 0, \;\; \mu \geq 0, \;\; \mu_0 + \sum_{i\in I^*} \mu_i = 1,$$

where $I^* = \{1 \leq i \leq m \,|\, g_i(x^*) = 0\}$. Further, if we had $\mu_0 = 0$ then $\sum_{i=1}^m \mu_i g_i'(x^*) = 0$ and

$$\sum_{i\in I^*} \mu_i g_i(\bar{x}) \geq \sum_{i\in I^*} \mu_i [g_i(x^*) + g_i'(x^*)^T(\bar{x} - x^*)] = 0,$$

what is a contradiction. Therefore, $\mu_0 \neq 0$ and we can set $\lambda_i = \mu_i/\mu_0$ for $i \in I$.■

The above results demonstrate that the fact that a point $x^* \in \mathrm{Dom}\, f$ is a global minimizer of a convex function f depends only on the local behavior of f at x^*. This is not the case with maximizers of a convex function. In fact, such a maximizer, if exists, in all nontrivial cases should belong to the boundary of the domain of the function:

Theorem 3.5.5 *Let f be convex, and let M be the domain of f. Assume that f attains its maximum on M at a point x^* from the relative interior of M. Then f is constant on M.*

Proof. Let $y \in M$; we should prove that $f(y) = f(x^*)$. There is nothing to prove if $y = x^*$, so that we may assume that $y \neq x^*$. Since, by assumption, $x^* \in \mathrm{ri}\, M$, we can extend the segment $[x^*, y]$ through the endpoint x^*, keeping the left endpoint of the segment in M; in other words, there exists a point $y' \in M$ such that x^* is an interior point of the segment $[y', y]$:

$$x^* = \lambda y' + (1 - \lambda) y$$

for certain $\lambda \in (0, 1)$. From the definition of convexity

$$f(x^*) \leq \lambda f(y') + (1 - \lambda) f(y).$$

Since both $f(y')$ and $f(y)$ do not exceed $f(x^*)$ (x^* is a maximizer of f on M!) and both the weights λ and $1 - \lambda$ are strictly positive, the indicated inequality can be valid only if $f(y') = f(y) = f(x^*)$. ■

The next two theorems give further information on maxima of convex functions:

Theorem 3.5.6 *Let f be a convex function on $\mathbf{R}^n$ and E be a subset of $\mathbf{R}^n$. Then*

$$\sup_{\mathrm{Conv}\, E} f = \sup_E f. \tag{3.5.19}$$

In particular, if $S \subset \mathbf{R}^n$ is convex and compact set, then the upper bound of f on S is equal to the upper bound of f on the set of extreme points of S:

$$\sup_S f = \sup_{\mathrm{Ext}(S)} f \tag{3.5.20}$$

Proof. To prove (3.5.19), let $x \in \mathrm{Conv}\, E$, so that x is a convex combination of points from E (Theorem 1.1.3 on the structure of convex hull):

$$x = \sum_i \lambda_i x_i \quad [x_i \in E,\, \lambda_i \geq 0,\, \sum_i \lambda_i = 1].$$

Applying Jensen's inequality (Proposition 3.1.3), we get

$$f(x) \leq \sum_i \lambda_i f(x_i) \leq \sum_i \lambda_i \sup_E f = \sup_E f,$$

so that the left hand side in (3.5.19) is $\leq$ the right hand one; the inverse inequality is evident, since $\mathrm{Conv}\, E \supset E$. ■

To derive (3.5.20) from (3.5.19), it suffices to note that from the Krein-Milman Theorem (Theorem 1.3.1) for a convex compact set S one has $S = \mathrm{Conv}\,\mathrm{Ext}(S)$. ■

The last theorem on maxima of convex functions is as follows:

Theorem 3.5.7 *Let f be a convex function such that the domain M of f is closed and does not contain lines. Then*

(i) *If the set*

$$\mathop{\mathrm{Argmax}}_M f \equiv \{x \in M \mid f(x) \geq f(y)\, \forall y \in M\}$$

of global maximizers of f is nonempty, then it intersects the set $\mathrm{Ext}(M)$ of the extreme points of M, so that at least one of the maximizers of f is an extreme point of M.

Proof. Let us start with (i). We will prove this statement by induction on the dimension of M. The base $\dim M = 0$, i.e., the case of a singleton M, is trivial, since here $M = \mathrm{Ext} M = \mathrm{Argmax}_M f$. Now assume that the statement is valid for the case of $\dim M \leq p$, and let us prove that it is valid also for the case of $\dim M = p + 1$. Let us first verify that the set $\mathrm{Argmax}_M f$ intersects with the (relative) boundary of M. Indeed, let $x \in \mathrm{Argmax}_M f$. There is nothing to prove if x itself is a relative boundary point of M; and if x is not a boundary point, then, by Theorem 3.5.5, f is constant on M, so that $\mathrm{Argmax}_M f = M$; and since M is closed, any relative boundary point of M (such a point does exist, since M does not contain lines and is of positive dimension) is a maximizer of f on M, so that here again $\mathrm{Argmax}_M f$ intersects $\partial_{\mathrm{ri}} M$.

Thus, among the maximizers of f there exists at least one, let it be x, which belongs to the relative boundary of M. Let H be the hyperplane which properly supports M at x (see Section 1.2.4), and let $M' = M \cap H$. The set M' is closed and convex (since M and H are), nonempty (it contains x) and does not contain lines (since M does not). We have $\max_M f = f(x) = \max_{M'} f$ (note that $M' \subset M$), whence

$$\emptyset \neq \mathop{\mathrm{Argmax}}_{M'} f \subset \mathop{\mathrm{Argmax}}_M f.$$

Same as in the proof of the Krein-Milman Theorem (Theorem 1.3.1), we have $\dim M' < \dim M$. In view of this inequality we can apply to f and M' our inductive hypothesis to get

$$\mathrm{Ext}(M') \cap \mathop{\mathrm{Argmax}}_{M'} f \neq \emptyset.$$

Since $\mathrm{Ext}(M') \subset \mathrm{Ext}(M)$ by Lemma 1.3.2 and, as we just have seen, $\mathrm{Argmax}_{M'} f \subset \mathrm{Argmax}_M f$, we conclude that the set $\mathrm{Ext}(M) \cap \mathrm{Argmax}_M f$ is not smaller than $\mathrm{Ext}(M') \cap \mathrm{Argmax}_{M'} f$ and is therefore nonempty, as required. ■

3.6 Legendre transformation

This section is not obligatory

Let f be a convex function. [4] We know that f is "basically" the upper bound of all its affine minorants; this is exactly the case when f is closed, otherwise the corresponding equality takes place everywhere except, perhaps, some points from the relative boundary of $\mathrm{Dom}\, f$. Now, when an affine function $d^T x - a$ is an affine minorant of f? It is the case if and only if

$$f(x) \geq d^T x - a$$

for all x or, which is the same, if and only if

$$a \geq d^T x - f(x)$$

for all x. We see that if the slope d of an affine function $d^T x - a$ is fixed, then in order for the function to be a minorant of f we should have

$$a \geq \sup_{x \in \mathbf{R}^n} [d^T x - f(x)].$$

The supremum in the right hand side of the latter relation is certain function of d; this function is called the *Legendre transformation* of f and is denoted f^*:

$$f^*(d) = \sup_{x \in \mathbf{R}^n} [d^T x - f(x)].$$

Geometrically, the Legendre transformation answers the following question: given a slope d of an affine function, i.e., given the hyperplane $t = d^T x$ in $\mathbf{R}^{n+1}$, what is the minimal "shift down" of the hyperplane which places it below the graph of f?

From the definition of the Legendre transformation it follows that this is a closed function. Indeed, we loose nothing when replacing $\sup_{x \in \mathbf{R}^n} [d^T x - f(x)]$ by $\sup_{x \in \mathrm{Dom}\, f} [d^T x - f(x)]$, so that the Legendre transformation is the upper bound of a family of affine functions. Since this bound is finite at least at one point (namely, at any d coming form affine minorant of f; we know that such a minorant exists), it is a convex closed function (indeed, defined as intersection of closed half-spaces, its epigraph is itself a convex and closed set).

The most elementary (and the most fundamental) fact about the Legendre transformation is its symmetry:

Proposition 3.6.1 *Let f be a convex function. Then twice taken Legendre transformation of f is the closure $\mathrm{cl}\, f$ of f:*

$$(f^*)^* = \mathrm{cl}\, f.$$

In particular, if f is closed, then it is the Legendre transformation of its Legendre transformation (which also is closed).

---

[4] From now till the end of the Section, if the opposite is not explicitly stated, "a function" means "a function defined on the entire $\mathbf{R}^n$ and taking values in $\mathbf{R} \cup \{+\infty\}$".

Proof is immediate. The Legendre transformation of f^* at the point x is, by definition,

$$\sup_{d\in\mathbf{R}^n}[x^T d - f^*(d)] = \sup_{d\in\mathbf{R}^n, a\geq f^*(d)}[d^T x - a];$$

the second sup here is exactly the supremum of all affine minorants of f (this is the origin of the Legendre transformation: $a \geq f^*(d)$ if and only if the affine form $d^T x - a$ is a minorant of f). And we already know that the upper bound of all affine minorants of f is the closure of f. ■

The Legendre transformation is a very powerful tool – this is a "global" transformation, so that *local* properties of f^* correspond to *global* properties of f. E.g.,

- $d = 0$ belongs to the domain of f^* if and only if f is below bounded, and if it is the case, then $f^*(0) = -\inf f$;
- if f is closed, then the subgradient of f^* at $d = 0$ are exactly the minimizers of f on $\mathbf{R}^n$;
- Dom f^* is the entire $\mathbf{R}^n$ if and only if $f(x)$ grows, as $|x| \to \infty$, faster than $|x|$: there exists a function $r(t) \to \infty$, as $t \to \infty$ such that

$$f(x) \geq r(|x|) \quad \forall x,$$

etc. Thus, whenever we can compute explicitly the Legendre transformation of f, we get a lot of "global" information on f. Unfortunately, the more detailed investigation of the properties of Legendre transformation is beyond the scope of our course; I simply shall list several simple facts and examples:

- From the definition of Legendre transformation,

$$f(x) + f^*(d) \geq x^T d \quad \forall x, d.$$

Specifying here f and f^*, we get certain inequality, e.g., the following one:
[Young's Inequality] *if p and q are positive reals such that $\frac{1}{p}+\frac{1}{q} = 1$, then*

$$\frac{|x|^p}{p} + \frac{|d|^q}{q} \geq xd \quad \forall x, d \in \mathbf{R}$$

(indeed, as it is immediately seen, the Legendre transformation of the function $|x|^p/p$ is $|d|^q/q$)

Consequences. Very simple-looking Young's inequality gives rise to a very nice and useful *Hölder inequality:*

Let $1 \leq p \leq \infty$ and let q be such $\frac{1}{p}+\frac{1}{q} = 1$ ($p = 1 \Rightarrow q = \infty$, $p = \infty \Rightarrow q = 1$). For every two vectors $x, y \in \mathbf{R}^n$ one has

$$\sum_{i=1}^n |x_i y_i| \leq |x|_p |y|_q \tag{3.6.21}$$

Indeed, there is nothing to prove if p or q is ∞ – if it is the case, the inequality becomes the evident relation

$$\sum_i |x_i y_i| \le (\max_i |x_i|)(\sum_i |y_i|).$$

Now let $1 < p < \infty$, so that also $1 < q < \infty$. In this case we should prove that

$$\sum_i |x_i y_i| \le (\sum_i |x_i|^p)^{1/p}(\sum_i |y_i|^q)^{1/q}.$$

There is nothing to prove if one of the factors in the right hand side vanishes; thus, we can assume that $x \neq 0$ and $y \neq 0$. Now, both sides of the inequality are of homogeneity degree 1 with respect to x (when we multiply x by t, both sides are multiplied by $|t|$), and similarly with respect to y. Multiplying x and y by appropriate reals, we can make both factors in the right hand side equal to 1: $|x|_p = |y|_p = 1$. Now we should prove that under this normalization the left hand side in the inequality is ≤ 1, which is immediately given by the Young inequality:

$$\sum_i |x_i y_i| \le \sum_i [|x_i|^p/p + |y_i|^q/q] = 1/p + 1/q = 1.$$

Note that the Hölder inequality says that

$$|x^T y| \le |x|_p |y|_q; \tag{3.6.22}$$

when $p = q = 2$, we get the Cauchy inequality. Now, inequality (3.6.22) is *exact* in the sense that for every x there exists y with $|y|_q = 1$ such that

$$x^T y = |x|_p \quad [= |x|_p |y|_q];$$

it suffices to take

$$y_i = |x|_p^{1-p} |x_i|^{p-1} \mathrm{sign}(x_i)$$

(here $x \neq 0$; the case of $x = 0$ is trivial – here y can be an arbitrary vector with $|y|_q = 1$).

Combining our observations, we come to an extremely important, although simple, fact:

$$|x|_p = \max\{y^T x \mid |y|_q \le 1\} \quad [\frac{1}{p} + \frac{1}{q} = 1]. \tag{3.6.23}$$

It follows, in particular, that $|x|_p$ is convex (as an upper bound of a family of linear forms), whence

$$|x' + x''|_p = 2|\frac{1}{2}x' + \frac{1}{2}x''|_p \le 2(|x'|_p/2 + |x''|_p/2) = |x'|_p + |x''|_p;$$

this is nothing but the triangle inequality. Thus, $|x|_p$ satisfies the triangle inequality; it clearly possesses two other characteristic properties of a norm – positivity and homogeneity. Consequently, $\| \cdot \|_p$ is a norm – the fact that we announced twice and have finally proved now.

- The Legendre transformation of the function

$$f(x) \equiv -a$$

is the function which is equal to a at the origin and is $+\infty$ outside the origin; similarly, the Legendre transformation of an affine function $\bar{d}^T x - a$ is equal to a at $d = \bar{d}$ and is $+\infty$ when $d \neq \bar{d}$;

- The Legendre transformation of the strictly convex quadratic form

$$f(x) = \frac{1}{2} x^T A x$$

(A is positive definite symmetric matrix) is the quadratic form

$$f^*(d) = \frac{1}{2} d^T A^{-1} d$$

- The Legendre transformation of the Euclidean norm

$$f(x) = |x|$$

is the function which is equal to 0 in the closed unit ball centered at the origin and is $+\infty$ outside the ball.

The latter example is a particular case of the following statement:

Let $\|x\|$ be a norm on $\mathbf{R}^n$, and let

$$\|d\|_* = \sup\{d^T x \mid \|x\| \leq 1\}$$

be the conjugate to $\|\cdot\|$ norm (it can be proved that $\|\cdot\|_*$ indeed is a norm, and that the norm conjugate to $\|\cdot\|_*$ is the original norm $\|\cdot\|$). The Legendre transformation of $\|x\|$ is the characteristic function of the unit ball of the conjugate norm, i.e., is the function of d equal to 0 when $\|d\|_* \leq 1$ and is $+\infty$ otherwise.

E.g., (3.6.23) says that the norm conjugate to $|\cdot|_p$, $1 \leq p \leq \infty$, is $|\cdot|_q$, $1/p+1/q = 1$; consequently, the Legendre transformation of p-norm is the characteristic function of the unit $|\cdot|_q$-ball.

3.7 Exercises: convex functions

Exercise 3.7.1 *Mark by "c" those of the following functions which are convex on the indicated domains:*

- $f(x) \equiv 1$ *on* $\mathbf{R}$
- $f(x) = x$ *on* $\mathbf{R}$
- $f(x) = |x|$ *on* $\mathbf{R}$
- $f(x) = -|x|$ *on* $\mathbf{R}$
- $f(x) = -|x|$ *on* $\mathbf{R}_+ = \{x \geq 0\}$
- $\exp\{x\}$ *on* $\mathbf{R}$
- $\exp\{x^2\}$ *on* $\mathbf{R}$
- $\exp\{-x^2\}$ *on* $\mathbf{R}$
- $\exp\{-x^2\}$ *on* $\{x \mid x \geq 100\}$

Exercise 3.7.2

- *Show that the function*

  $$f(x,t) = -\ln(t^2 - x^2)$$

  is convex on the "angle"

  $$\{(t,x) \in \mathbf{R}^2 \mid t > |x|\}$$

  of $\mathbf{R}^2$.

- *For which* τ, ξ *the function*

  $$-\ln(t^2 - x^2) + \tau t - \xi x$$

  attains its minimum on the "angle"

  $$\{(t,x) \in \mathbf{R}^2 \mid t > |x|\}$$

  of $\mathbf{R}^2$*? What is the value of this minimum?*

Exercise 3.7.3 *Let* $0 < p < 1$. *Show that the function*

$$-x^p y^{1-p}$$

is convex on the domain $\{(x,y) \mid x, y > 0\}$.

Let n *be a positive integer and let* $0 < \alpha \leq 1/n$. *Prove that the function*

$$f(x) = -(x_1 x_2 ... x_n)^\alpha$$

is convex on $\{x \in \mathbf{R}^n \mid x_1, ..., x_n \geq 0\}$.

Hint: please, do not differentiate! Try to use the conic transformation first.

Exercise 3.7.4 *Prove that the following functions are convex on the indicated domains:*

- $\frac{x^2}{y}$ *on* $\{(x,y) \in \mathbf{R}^2 \mid y > 0\}$
- $\ln(\exp\{x\} + \exp\{y\})$ *on the 2D plane.*

Exercise 3.7.5 *A function f defined on a convex set M is called log-convex on M, if it takes real positive values on M and the function $\ln f$ is convex on M. Prove that*

- *a log-convex on M function is convex on M*
- *the sum (more generally, linear combination with positive coefficients) of two log-convex functions on M also is log-convex on the set.*

Hint: use the result of the previous Exercise + your knowledge on operations preserving convexity.

Exercise 3.7.6 *Consider a Linear Programming program*

$$c^T x \to \min \mid Ax \le b$$

with $m \times n$ matrix A, and let x^* be an optimal solution to the problem. It means that x^* is a minimizer of differentiable convex function $f(x) = c^T x$ with convex differentiable constraint $Ax \le b$, so that the necessary and sufficient condition of optimality given by Theorem 3.5.4 should be satisfied at x^*. What does this condition mean in terms of the data A, b, c?

Exercise 3.7.7 *Find the minimizer of a linear function*

$$f(x) = c^T x$$

on the set

$$V_p = \{x \in \mathbf{R}^n \mid \sum_{i=1}^{n} |x_i|^p \le 1\};$$

here p, $1 < p < \infty$, is a parameter.