BASICS OF CONVEX ANALYSIS

MARKUS GRASMAIR

1. Main Definitions

We start with providing the central definitions of convex functions and convex sets.

Definition 1. A function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called *convex*, if

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

for all x, $y \in \mathbb{R}^n$ and all $0 < \lambda < 1$.

It is called *strictly convex*, if it is convex and

$$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda) f(y)$$

whenever $x \ne y$, $0 < \lambda < 1$, and $f(x) < \infty$, $f(y) < \infty$.

It is important to note here that we allow the functions we consider to take the value $+\infty$. This leads sometimes to complications in the notation, but will be very helpful later on. Indeed, convex functions taking the value $+\infty$ appear naturally as soon as we start looking at *duality*, which is possibly *the* central topic in convex analysis. Note, however, that the functions may not take the value $-\infty$.

Definition 2. A set $K \subset \mathbb{R}^n$ is convex, if

$$\lambda x + (1-\lambda)y \in K$$

whenever x, $y \in K$, and $0 < \lambda < 1$.

We note that there is a very close connection between the convexity of functions and the convexity of sets. Given a set $K \subset \mathbb{R}^n$, we denote by $\chi_K\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ its *characteristic function*, given by

$$\chi_K(x) = \begin{cases} 0 & \text{if } x \in K, \\ +\infty & \text{if } x \notin K. \end{cases}$$

Lemma 3. *A set K is convex, if and only if its characteristic function χ_K is convex.*

Proof. Assume that K is convex, and let x, $y \in \mathbb{R}^n$, and $0 < \lambda < 1$. If either $x \notin K$ or $y \notin K$, then $\lambda \chi_K(x) + (1-\lambda)\chi_K(y) = \infty$, and therefore (1) (with $f = \chi_K$) is automatically satisfied. On the other hand, if x, $y \in K$, then also $\lambda x + (1-\lambda)y \in K$, and therefore

$$\chi_K(\lambda x + (1-\lambda)y) = 0 = \lambda \chi_K(x) + (1-\lambda)\chi_K(y).$$

Thus χ_K is convex.

Now assume that χ_K is convex, and let x, $y \in K$, and $0 < \lambda < 1$. The convexity of χ_K implies that

$$\chi_K(\lambda x + (1-\lambda)y) \le \lambda \chi_K(x) + (1-\lambda)\chi_K(y) = 0.$$

Since χ_K only takes the values 0 and $+\infty$, this shows that, actually, $\chi_K(\lambda x + (1-\lambda)y) = 0$, and therefore $\lambda x + (1-\lambda)y \in K$. Thus K is convex. $\square$

Date: March 2015.

Conversely, given a function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, define its *epigraph*[1] as

$$\operatorname{epi}(f) := \bigl\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : t \geq f(x)\bigr\}.$$

Lemma 4. *The function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, if and only if its epigraph $\operatorname{epi}(f) \subset \mathbb{R}^{n+1}$ is convex.*

Proof. Assume first that the epigraph of f is convex and let x, $y \in \mathbb{R}^n$ be such that $f(x)$, $f(y) < \infty$, and let $0 < \lambda < 1$. Since $(x, f(x))$, $(x, f(y)) \in \operatorname{epi}(f)$ it follows that

$$\lambda(x, f(x)) + (1-\lambda)(y, f(y)) \in \operatorname{epi}(f),$$

which means that

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

Since this holds for every x, y and $0 < \lambda < 1$, the function f is convex.

Now assume that the function f is convex and let (x,t), $(y,s) \in \operatorname{epi}(f)$, and $0 < \lambda < 1$. Then

$$\lambda t + (1-\lambda) s \geq \lambda f(x) + (1-\lambda) f(y) \geq f(\lambda x + (1-\lambda) y),$$

which implies that $\lambda(x,t) + (1-\lambda)(y,s) \in \operatorname{epi}(f)$. Thus $\operatorname{epi}(f)$ is convex. □

2. Continuity

A convex function need not be continuous—as can be easily seen from the example of the characteristic function of a convex set, which is only ever continuous in the trivial situations $K = \emptyset$ and $K = \mathbb{R}^n$.

An even more worrying situation (at least from an optimization point of view) occurs for instance in the situation of the function $f\colon \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f(x) = \begin{cases} +\infty, & \text{if } x < 0, \\ 1, & \text{if } x = 0, \\ x^2, & \text{if } x > 0. \end{cases}$$

This function can easily seen to be convex, and it is obviously not continuous at 0. The problem is that it is not even lower semi-continuous: If you approach the value 0 from the right, the function values tend to 0, but $f(0) = 1$. Thus this is a convex (and obviously coercive) function, which does not attain its minimum. Fortunately, problems with continuity and lower semi-continuity only occur near points where a convex function takes the value $+\infty$, as the following result shows.

In the following we will always denote by

$$\operatorname{dom}(f) := \bigl\{x \in \mathbb{R}^n : f(x) < \infty\bigr\}$$

the *domain* of the function f. We will say that a function is *proper*, if its domain is non-empty; that is, the function is not identically equal to $+\infty$.

Proposition 5. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and that x is contained in the interior of* $\operatorname{dom}(f)$. *Then f is locally Lipschitz continuous at x.*

In other words, discontinuities of convex functions can only appear at the boundary of their domain.

Proof. See [3, Thm. 10.4]. □

[1] The word *epigraph* is composed of the greek preposition *epi* meaning *above, over* and the word graph. Thus it simply means "above the graph."

3. Differentiation

Assume now that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is any function and that $x \in \operatorname{dom}(f)$. Given $d \in \mathbb{R}^n$, we define the *directional derivative of f at x in direction d* as

$$Df(x;d) := \lim_{\substack{t\to 0\\ t>0}} \frac{f(x+td)-f(x)}{t} \in \mathbb{R} \cup \{\pm\infty\}$$

provided that the limit exists (in the extended real line $\mathbb{R} \cup \{\pm\infty\}$.

For general functions f, the directional derivative need not necessarily exist (take for instance the function $x \mapsto x\sin(1/x)$, which can be shown to have no directional derivatives at 0 in directions $d \neq 0$). For convex functions, however, the situation is different.

Lemma 6. *Assume that f is convex, $x \in \operatorname{dom}(f)$, and $d \in \mathbb{R}^n$. Then $Df(x;d)$ exists. If x is an element of the interior of $\operatorname{dom}(f)$, then $Df(x;d) < \infty$.*

Proof. Let $0 \le t_1 < t_2$. Then

$$x + t_1 d = \frac{t_1}{t_2}(x + t_2 d) + \Big(1 - \frac{t_1}{t_2}\Big)x,$$

and thus the convexity of f implies that

$$f(x+t_1 d) \le \Big(1 - \frac{t_1}{t_2}\Big) f(x) + \frac{t_1}{t_2} f(x + t_2 d),$$

which can be rewritten as

$$\frac{f(x+t_1 d) - f(x)}{t_1} \le \frac{f(x+t_2 d) - f(x)}{t_2}.$$

Thus the function

$$t \mapsto \frac{f(x+td)-f(x)}{t}$$

is increasing. As a consequence, its limit as $t \to 0$ from above—which is exactly the directional derivative $Df(x;d)$—exists. In addition, if x is in the interior of the domain of f, then $f(x+td)$ is finite for t sufficiently small, and therefore $Df(x;d) < \infty$. □

Using the same argumenation as in the proof of the previous statement, one sees that

$$f(x+td) \ge f(x) + tDf(x;d)$$

for all $t \ge 0$. Thus a sufficient (and also necessary) condition for $x \in \operatorname{dom}(f)$ to be a *global* minimizer of f is that

$$Df(x;d) \ge 0 \qquad \text{for all } d \in \mathbb{R}^n.$$

Next, we will introduce another notion of derivatives that is tailored to the needs of convex functions.

Definition 7. Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $x \in \operatorname{dom}(f)$. An element $\xi \in \mathbb{R}^n$ with the property that

$$f(y) \ge f(x) + \xi^T(y-x) \qquad \text{for all } y \in \mathbb{R}^n$$

is called a *subgradient* of f at x.

The set of all subgradients of f at x is called the *subdifferential* of f at x, and is denoted by $\partial f(x)$.

An immediate consequence of the definition of the subdifferential is the following result:

Lemma 8. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and proper. Then $x \in \mathbb{R}^n$ is a global minimizer of f, if and only if*

$$0 \in \partial f(x).$$

Proof. Exercise. □

The subdifferential is connected to the usual derivative and the directional derivative in the following way:

Proposition 9. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and that $x \in \operatorname{dom}(f)$.*

- *If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.*
- *If $\partial f(x)$ contains a single element, e.g. $\partial f(x) = \{\xi\}$, then f is differentiable at x and $\xi = \nabla f(x)$.*

Proof. See [3, Thm. 25.1]. □

Proposition 10. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and that $x \in \operatorname{dom}(f)$. Then $\xi \in \partial f(x)$, if and only if*

$$\xi^T d \le Df(x;d) \qquad \text{for all } d \in \mathbb{R}^n.$$

Proof. See [3, Thm. 23.2]. □

Next we will compute the subdifferential in some particular (somehow important) cases:

- We consider first the function $f\colon \mathbb{R}^n \to \mathbb{R}$,

  $$f(x) = \|x\|_2 = \Big(\sum_{i=1}^n x_i^2\Big)^{1/2}.$$

  If $x \neq 0$, then f is differentiable at x with deriviative $\nabla f(x) = x/\|x\|_2$, and therefore $\partial f(x) = \{x/\|x\|_2\}$ for $x \neq 0$.

  Now consider the case $x = 0$. We have $\xi \in \partial f(0)$, if and only if

  $$\xi^T y \le f(y) - f(0) = \|y\|_2$$

  for all $y \in \mathbb{R}^n$, which is equivalent to the condition that $\|\xi\|_2 \le 1$.

  Thus we have

  $$\partial f(x) = \begin{cases} \Big\{\dfrac{x}{\|x\|_2}\Big\} & \text{if } x \neq 0, \\ \big\{\xi \in \mathbb{R}^n : \|\xi\|_2 \le 1\big\} & \text{if } x = 0. \end{cases}$$

- Consider now the function $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$,

  $$f(x) = \begin{cases} 0, & \text{if } \|x\|_2 \le 1, \\ +\infty, & \text{if } \|x\|_2 > 1. \end{cases}$$

  If $x \in \mathbb{R}^n$ satisfies $\|x\| < 1$, then f is differentiable at x with gradient equal to 0, and therefore $\partial f(x) = \{0\}$. If $\|x\| > 1$, then $f(x) = +\infty$, and therefore $\partial f(x) = \emptyset$. Finally, if $\|x\| = 1$, then $\xi \in \partial f(x)$, if and only if

  $$\xi^T(y - x) \le f(y) - f(x) \qquad \text{for all } y \in \mathbb{R}^n,$$

  which is equivalent to the condition

  $$\xi^T(y - x) \le 0 \qquad \text{for all } y \in \mathbb{R}^n \text{ with } \|y\|_2 \le 1.$$

  The latter condition can in turn be shown to be equivalent to stating that

  $$\xi = \lambda x \qquad \text{for some } \lambda \ge 0.$$

Thus we obtain that

$$\partial f(x) = \begin{cases} \emptyset, & \text{if } \|x\| > 1, \\ \{\lambda x : \lambda \geq 0\} & \text{if } \|x\| = 1, \\ \{0\}, & \text{if } \|x\| < 1. \end{cases}$$

Now assume that f is convex and that $x,\ y \in \mathbb{R}^n$, and that $\xi \in \partial f(x)$ and $\eta \in \partial f(y)$. Then the definition of the subdifferential implies that

$$f(y) - f(x) - \xi^T(y-x) \geq 0,$$
$$f(x) - f(y) - \eta^T(x-y) \geq 0.$$

Summing up these inequalities, we see that

$$(\eta - \xi)^T(y-x) \geq 0$$

whenever $\xi \in \partial f(x)$ and $\eta \in \partial f(y)$.

In case f is differentiable, this inequality becomes

$$(\nabla f(y) - \nabla f(x))^T(y-x) \geq 0 \qquad \text{for all } x,\ y \in \mathbb{R}^n.$$

Now consider the case of a one-dimensional function. Then this further simplifies to

$$(f'(y) - f'(x))(y-x) \geq 0 \qquad \text{for all } x,\ y \in \mathbb{R},$$

which can be restated as

$$f'(y) \geq f'(x) \qquad \text{whenever} \qquad y \geq x.$$

In other word, the derivative of a one-dimensional function is monotoneously increasing. Even more, the converse is also true:

Proposition 11. *Assume that $f\colon \mathbb{R} \to \mathbb{R}$ is convex and differentiable. Then f' is monotoneously increasing. Conversely, if $g\colon \mathbb{R} \to \mathbb{R}$ is monotoneously increasing and continuous, then there exists a convex function f such that $f' = g$.*

Proof. The necessity of the monotonicity of f' has already been shown above. Conversely, if g is monotoneously increasing, we can define

$$f(x) := \int_0^x g(y)\,dy.$$

Then, we have for $y > x$ and $0 < \lambda < 1$ that

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \lambda f(x) + (1-\lambda) f(y) \\ &\qquad + \lambda \int_x^{\lambda x + (1-\lambda)y} g(z)\,dz - (1-\lambda) \int_{\lambda x + (1-\lambda)y}^{y} g(z)\,dz \\ &\leq \lambda f(x) + (1-\lambda) f(y) + \lambda(1-\lambda)(y-x) g'(\lambda x + (1-\lambda)y) \\ &\qquad - (1-\lambda)\lambda(y-x) g'(\lambda x + (1-\lambda)y) \\ &= \lambda f(x) + (1-\lambda) f(y), \end{aligned}$$

showing that f is convex. □

In higher dimensions, the situation is slightly more complicated. Still, one can show the following result:

Proposition 12. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is convex, if and only if ∇f is increasing in the sense that*

$$(\nabla f(y) - \nabla f(x))^T(y-x) \geq 0 \qquad \textit{for all } x,\ y \in \mathbb{R}^n.$$

Proof. Apply Proposition 11 to the one-dimensional functions $t \mapsto f(x + t(y - x))$. □

Note the difference between the two different statements: In the first proposition (in one dimension), the monotonicity of g implies that it is the derivative of some convex function. In the second proposition (in higher dimensions), we already know that the object we are dealing with (∇f) is the derivative of some function; monotonocity of the gradient implies that f is convex, but we do not have to care about the existence of f.

Finally, it is important to note that there is a close connection between properties of the Hessian of a function and its convexity:

Proposition 13. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is twice differentiable on its domain, and that $\operatorname{dom}(f)$ is convex. Then f is convex, if and only if $\nabla^2 f(x)$ is positive semi-definite for all $x \in \operatorname{dom}(f)$.*

In addition, if $\nabla^2 f(x)$ is positive definite for all but a countable number of points $x \in \operatorname{dom}(f)$, then f is strictly convex.

4. Transformations of convex functions

Next we will discuss, which operations preserve the convexity of functions, and how these operations affect the subdifferential.

The first thing to note is that convex functions behave well under multiplication with positive scalars: If $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and $\lambda > 0$, then also λf is convex and

$$\partial(\lambda f)(x) = \lambda \partial f(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

Note that this result does not hold for negative scalars: If f is convex and $\lambda < 0$, then usually λf will not convex. Also, for $\lambda = 0$ some problems occur if f takes the value $+\infty$. First, it is not clear how to define the product $0 \cdot (+\infty)$ in a meaningful way, and, second, the equality $\partial(0 \cdot f)(x) = 0 \cdot \partial f(x)$ need not necessarily be satisfied for all x.

Next we consider a function that is defined as a supremum of convex functions:

Lemma 14. *Assume that I is any index set (not necessarily finite or even countable) and that $f_i\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, $i \in I$, are convex. Define*

$$g(x) := \sup_{i \in I} f_i(x).$$

Then g is convex.

In addition, if $x \in \mathbb{R}^n$ and $g(x) = f_j(x)$ for some $j \in I$, then

$$\partial f_j(x) \subset \partial g(x).$$

Proof. If x, $y \in \mathbb{R}^n$ and $0 < \lambda < 1$, then

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= \sup_{i \in I} f_i(\lambda x + (1-\lambda)y) \\ &\le \sup_{i \in I} \Big(\lambda f_i(x) + (1-\lambda) f_i(y) \Big) \\ &\le \lambda \sup_{i \in I} f_i(x) + (1-\lambda) \sup_{i \in I} f_i(y) \\ &= \lambda g(x) + (1-\lambda) g(y). \end{aligned}$$

Thus g is convex.

In addition, if $g(x) = f_j(x)$ and $\xi \in \partial f_j(x)$, then

$$g(y) \ge f_j(y) \ge f_j(x) + \xi^T(y-x) = g(x) + \xi^T(y-x),$$

showing that $\xi \in \partial f_j(x)$. $\square$

There are two important things to note here: First, the result states that the supremum of *any* number of convex functions is convex; there is no need to restrict oneself to only finitely many functions. Second, the *minimum* even of two convex functions is usually not convex any more.

Next we discuss sums of convex functions:

Proposition 15. *Assume that f, $g\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ are convex. Then also the function $f+g$ is convex and*

$$\partial(f+g)(x) \subset \partial f(x) + \partial g(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

If in addition f and g are lower semi-continuous and there exists some $y \in \operatorname{dom}(f) \cap \operatorname{dom}(g)$ such that either one of the functions f and g is continuous at y, then

$$\partial(f+g)(x) = \partial f(x) + \partial g(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

Proof. The first part of the proposition is straightforward:

$$\begin{aligned}(f+g)(\lambda x + (1-\lambda)y) &= f(\lambda x + (1-\lambda)y) + g(\lambda x + (1-\lambda)y)\\ &\le \lambda f(x) + (1-\lambda)f(y) + \lambda g(x) + (1-\lambda)g(y)\\ &= \lambda(f+g)(x) + (1-\lambda)(f+g)(y).\end{aligned}$$

Also, if $\xi \in \partial f(x)$ and $\eta \in \partial g(x)$, then

$$(f+g)(y) \ge f(x) + \xi^T(y-x) + g(x) + \eta^T(y-x) = (f+g)(x) + (\xi+\eta)^T(y-x),$$

showing that $\partial(f+g)(x) \subset \partial f(x) + \partial g(x)$.

The second part is tricky. A proof can for instance be found in [2, Chap. I, Prop. 5.6]. □

In the above proposition, the sum of subdifferentials is defined as

$$\partial f(x) + \partial g(x) = \big\{\xi + \eta \in \mathbb{R}^n : \xi \in \partial f(x) \text{ and } \eta \in \partial g(x)\big\}.$$

Example 16. Define the functions f, $g\colon \mathbb{R}^2 \to \mathbb{R} \cup \{+\infty\}$,

$$f(x) = \begin{cases} 0, & \text{if } \|x\|_2 \le 1,\\ +\infty, & \text{if } \|x\|_2 > 1,\end{cases} \qquad \text{and} \qquad g(x) = \begin{cases} 0, & \text{if } \|x-(2,0)\|_2 \le 1,\\ +\infty, & \text{if } \|x-(2,0)\|_2 > 1,\end{cases}$$

Both of these functions are convex. Moreover

$$(f+g)(x) = \begin{cases} 0, & \text{if } x = (1,0),\\ +\infty & \text{else.}\end{cases}$$

Now consider the subdifferential of the different functions at $(1,0)$. We have

$$\partial f(1,0) = \big\{\lambda(1,0) : \lambda \ge 0\big\}$$

and

$$\partial g(1,0) = \big\{\mu(-1,0) : \mu \ge 0\big\}.$$

Thus

$$\partial f(1,0) + \partial g(1,0) = \mathbb{R}(1,0).$$

On the other hand,

$$\partial(f+g)(1,0) = \mathbb{R}^2.$$

Thus in the case we have

$$\partial f(1,0) + \partial g(1,0) \subsetneq \partial(f+g)(1,0).$$

Finally, we study the composition of convex functions and linear transformations:

Proposition 17. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and that $A \in \mathbb{R}^{n\times m}$. Then the function $g\colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ defined by*

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

is convex.

If in addition there exists $y \in \operatorname{Ran} A$ such that f is continuous at y, then

$$\partial g(x) = A^T \partial f(Ax)$$

for all $x \in \mathbb{R}^n$.

Proof. Proving the convexity of g is straightforward. The proof of the representation of the subdifferential can for instance be found in [2, Chap. I, Prop. 5.7]. □

Note that the second part of the previous result is just a kind of chain rule for the composition of convex and linear functions. In the particular case where f is differentiable, the result reads as

$$\nabla(f \circ A)(x) = A^T \nabla f(Ax),$$

which is exactly the usual chain rule. Differentiability, however, is not required in the convex setting.

5. Duality

Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semi-continuous and consider any linear function $x \mapsto \xi^T x$ (for some $\xi \in \mathbb{R}^n$). Then it is easy to convince oneself that there exists some $\beta \in \mathbb{R}$ such that

$$f(x) \geq \xi^T x - \beta \qquad \text{for all } x \in \mathbb{R}^n$$

(this follows basically from the definition and the local Lipschitz continuity of convex functions). In other words,

$$\beta \geq \xi^T x - f(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

Now consider the smallest such number β^*, that is,

$$\beta^* := \inf\bigl\{\beta \in \mathbb{R} : \beta \geq \xi^T x - f(x) \text{ for all } x \in \mathbb{R}^n\bigr\}.$$

It is easy to see that $\beta^* > -\infty$ unless the function f is everywhere equal to $+\infty$. Thus one can equivalently write

$$\beta^* = \sup_{x\in\mathbb{R}^n} \xi^T x - f(x).$$

The mapping that maps ξ to β^* plays a central role in covnex analysis.

Definition 18. Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper. Then its *conjugate* $f^*\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is defined as

$$f^*(\xi) := \sup_{x\in\mathbb{R}^n} \xi^T x - f(x).$$

Lemma 19. *Let $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper. Then f^* is convex and lower semi-continuous.*

Proof. The function f^* is the supremum of the collection of all affine (and thus in particular convex and lower semi-continuous) functions $h_x(\xi) := \xi^T x - f(x)$. Thus also f^* is convex and lower semi-continuous. □

Example 20. Let

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

with $1 < p < +\infty$. In order to compute $f^*(\xi)$, we have to maximize the function

$$x \mapsto h_\xi(x) := \xi - \nabla f(x).$$

Since h_ξ is concave (that is, $-h_\xi$ is convex) and differentiable, maximization is equivalent to solving the equation $\nabla h_\xi(x) = 0$. Now note that $\nabla h_\xi(x) = \xi - \nabla f(x)$, and

$$\nabla f(x) = \begin{pmatrix} x_1|x_1|^{p-2} \\ \vdots \\ x_n|x_n|^{p-2} \end{pmatrix}.$$

Thus the maximum x^* satisfies

$$x_i^*|x_i^*|^{p-2} = \xi_i$$

for all i. Solving these equations for x_i^*, we obtain

$$x_i^* = \xi_i|\xi_i|^{\frac{1}{p-1}-1}.$$

Thus

$$f^*(\xi) = h_\xi(x^*) = \sum_{i=1}^n |\xi_i|^{\frac{1}{p-1}+1} - \frac{1}{p}\sum_{i=1}^n |\xi_i|^{\frac{p}{p-1}}.$$

Defining

$$p_* := \frac{p}{p-1},$$

this can be written as

$$f^*(\xi) = \frac{1}{p_*}\sum_{i=1}^n |\xi_i|^{p_*} = \frac{1}{p_*}\|\xi\|_{p_*}^{p_*}.$$

Example 21. Let

$$f(x) = \|x\|_1 = \sum_{i=1}^n |x_i|.$$

Then

$$f^*(\xi) = \sup_{x\in\mathbb{R}^n}\Big(\xi^T x - \sum_{i=1}^n |x_i|\Big).$$

Now assume that $|\xi_j| > 1$ for some $1 \le j \le n$. Now consider only vectors x in the supremum above for which the jth entry is $\lambda \operatorname{sgn}\xi_j$ and all the other entries of which are zero. Then we see that

$$f^*(\xi) \ge \sup_{\lambda>0}\big(\xi_j\lambda\operatorname{sgn}\xi_j - \lambda\big) = \sum_{\lambda>0}\lambda(|\xi_j| - 1) = +\infty.$$

On the other hand, if $|\xi_j| \le 1$ for all j (or in other words $\|\xi\|_\infty \le 1$, then

$$f^*(\xi) = \sup_{x\in\mathbb{R}^n}\Big(\xi^T x - \sum_{i=1}^n |x_i|\Big) = \sup_{x\in\mathbb{R}^n}\sum_{i=1}^n\big(\xi_i x_i - |x_i|\big) = 0,$$

since all the entries in the sum are non-positive. Thus we see that

$$f^*(\xi) = \begin{cases} 0, & \text{if } \|\xi\|_\infty \le 1, \\ +\infty, & \text{if } \|\xi\|_\infty > 1. \end{cases}$$

A trivial consequence of the definition of the conjugate of a function is the so called *Fenchel inequality*:

Lemma 22. *If $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper, then*

$$f(x) + f^*(\xi) \geq \xi^T x \qquad \text{for all } x,\ \xi \in \mathbb{R}^n.$$

Proof. This follows immediately from the definition of f^*. □

Proposition 23. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper, convex and lower semi-continuous and that $x,\ \xi \in \mathbb{R}^n$. Then the following are equivalent:*

$$\begin{aligned} \xi^T x &= f(x) + f^*(\xi), \\ \xi &\in \partial f(x), \\ x &\in \partial f^*(\xi). \end{aligned}$$

Proof. See [3, Thm. 23.5]. □

As a next step, it is also possible to compute the conjugate of the

$$f^{**}(x) := (f^*)^*(x) = \sup_{\xi \in \mathbb{R}^n} \big(\xi^T x - f^*(\xi)\big).$$

Theorem 24. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is proper, convex, and lower semi-continuous. Then $f^{**} = f$.*

Sketch of Proof. First recall that f^{**} is by construction convex and lower semi-continuous. Moreover we have that

$$\begin{aligned} f^{**}(x) = \sup_{\xi \in \mathbb{R}^n} \big(\xi^T x - f^*(\xi)\big) &= \sup_{\xi \in \mathbb{R}^n} \Big(\xi^T x - \sup_{y \in \mathbb{R}^n} \big(\xi^T y - f(y)\big)\Big) \\ &= \sup_{\xi \in \mathbb{R}^n} \inf_{y \in \mathbb{R}^n} \big(\xi^T (x - y) + f(y)\big) \leq f(x) \end{aligned}$$

for every $x \in \mathbb{R}^n$ (for the last inequality, we simply choose $y = x$). Thus f^{**} is a convex and lower semicontinuous function that is smaller or equal to f.

Now assume that there exists some $x \in \mathbb{R}^n$ such that $f^{**}(x) < f(x)$. One can show[2] that, because of the convexity and lower semicontinuity of f, in this case there exist $\delta > 0$ and $\xi \in \mathbb{R}^n$ such that

$$f(y) \geq f^{**}(x) + \xi^T (y - x) + \delta \qquad \text{for all } y \in \mathbb{R}^n.$$

In other words, the function f and the point $f^{**}(x)$ can be separated strictly by the (shifted) hyperplane defined by ξ. Therefore

$$\begin{aligned} f^{**}(x) = \sup_{\zeta \in \mathbb{R}^n} \inf_{y \in \mathbb{R}^n} \big(\zeta^T (x - y) + f(y)\big) \\ \geq \inf_{y \in \mathbb{R}^n} \big(\zeta^T (x - y) + f(y)\big) \geq f^{**}(x) + \delta, \end{aligned}$$

which is obviously a contradiction. Thus $f^{**}(x) \geq f(x)$ for every $x \in \mathbb{R}^n$. Since we have already shown that $f^{**}(x) \leq f(x)$ for all x, this implies that, actually $f^{**} = f$. □

Remark 25. Note that we have actually shown in the proof that for all proper functions f (not necessarily convex or lower semicontinuous), the function f^{**} is the largest convex and lower semicontinuous function below f. That is, the function f^{**} is the *convex and lower semicontinuous hull* of f.

[2]Basically, the proof of this statement relies on so called *separation theorems*, which state that it is possible to separate a closed convex set from a point outside said set by means of a hyperplane. Applying this result to the epigraph of f yields the claimed result. A large collection of such separation theorems can be found in [3, Section 11].

The following list taken from [1, p. 50] gives a short overview of some important functions and their conjugates. Note that all these functions are lower semicontinuous and convex, and thus we have $f^{**} = f$; that is, the functions on the left hand side are also the conjugates of the functions on the right hand side:

$$
\begin{aligned}
&f(x) = 0, && \iff && f^*(\xi) = \begin{cases} 0, & \text{if } \xi = 0, \\ +\infty & \text{else,} \end{cases} \\
&f(x) = \begin{cases} 0 & \text{if } x \geq 0, \\ +\infty & \text{else,} \end{cases} && \iff && f^*(\xi) = \begin{cases} 0 & \text{if } \xi \leq 0, \\ +\infty & \text{else,} \end{cases} \\
&f(x) = \begin{cases} 0 & \text{if } |x| \leq 1, \\ +\infty & \text{else,} \end{cases} && \iff && f^*(\xi) = |\xi|, \\
&f(x) = |x|^p/p, && \iff && f^*(\xi) = |\xi|^{p_*}/p_*, \\
&f(x) = \sqrt{1+x^2}, && \iff && f^*(\xi) = \begin{cases} -\sqrt{1-\xi^2}, & \text{if } |y| \leq 1, \\ +\infty, & \text{else,} \end{cases} \\
&f(x) = e^x, && \iff && f^*(\xi) = \begin{cases} \xi \log \xi - \xi, & \text{if } \xi \geq 0, \\ +\infty & \text{else.} \end{cases}
\end{aligned}
$$

In addition, the next list provides a few basic rules for the computation of conjugates:

$$
\begin{aligned}
&f(x) = h(tx), && f^*(\xi) = h^*(\xi/t) && t \neq 0, \\
&f(x) = h(x+y), && f^*(\xi) = h^*(\xi) - \xi^T y, && y \in \mathbb{R}^n, \\
&f(x) = \lambda h(x), && f^*(\xi) = \lambda h^*(\xi/\lambda), && \lambda > 0.
\end{aligned}
$$

6. Primal and dual problem

We now consider the special (but important) case of an optimization problem on $\mathbb{R}^n$ of the form

$$f(x) + g(Ax) \to \min,$$

where $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, $A \in \mathbb{R}^{m \times n}$, and $g\colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is convex. Its *dual problem* is the optimization problem on $\mathbb{R}^m$ given by

$$f^*(A^T \xi) + g^*(-\xi) \to \min .$$

One of the central theorems in convex analysis is the following one, which shows that, under certain non-restrictive assumptions, these two problems are equivalent in some sense.

Theorem 26. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is convex, $A \in \mathbb{R}^{m \times n}$, and $g\colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is convex. Assume moreover that there exists $x \in \operatorname{dom}(f)$ such that g is continuous at Ax. Assume moreover that $x^* \in \mathbb{R}^n$ and $\xi^* \in \mathbb{R}^m$. Then the following are equivalent:*

(1) *The vector x^* minimizes the function $x \mapsto f(x) + g(Ax)$ and ξ^* minimizes the function $\xi \mapsto f^*(A^T \xi) + g^*(-\xi)$.*
(2) *We have*
$$f(x^*) + g(Ax^*) + f^*(A^T \xi^*) + g^*(-\xi^*) = 0.$$
(3) *We have*
$$x^* \in \partial f^*(A^T \xi^*) \qquad \text{and} \qquad Ax^* \in \partial g^*(-\xi^*).$$
(4) *We have*
$$A^T \xi^* \in \partial f(x^*) \qquad \text{and} \qquad -\xi^* \in \partial g(Ax^*).$$

Proof. See [3, Thms. 31.2, 31.3]. □

Basically, this result allows one to switch freely between the primal (original) problem and the dual problem and to try to solve the one that appears to be easier. For instance, if $n \gg m$, the primal problem may be much more complicated than the dual problem, because the number of variables is much larger. Thus, instead of solving the n-dimensional primal problem, one can try the solve the m-dimensional dual problem. If in addition the function f^* is differentiable (which is the case if f is strictly convex), then x^* can be easily recovered from the condition $x^* \in \partial f^*(A^T \xi^*)$, since the subdifferential of f^* contains only a single element.

References

[1] Jonathan M. Borwein and Adrian S. Lewis. *Convex analysis and nonlinear optimization.* CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3. Springer-Verlag, New York, 2000. Theory and examples.

[2] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems.* North-Holland, Amsterdam, 1976.

[3] R. T. Rockafellar. *Convex Analysis*, volume 28 of *Princeton Mathematical Series.* Princeton University Press, Princeton, 1970.

Department of Mathematics, Norwegian University of Science and Technology, 7491 Trondheim, Norway

E-mail address: markus.grasmair@math.ntnu.no