BASIC PROPERTIES OF CONVEX FUNCTIONS

MARKUS GRASMAIR

1. Convex Functions and Sets

Definition 1 (Convex Function)**.** A function $f\colon \mathbb{R}^n \to \mathbb{R}$ is *convex*, if for every x, $y \in \mathbb{R}^n$ and $0 \le \lambda \le 1$ the inequality

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda) f(y)$$

holds.

The function f is *strictly convex*, if for every $x \neq y \in \mathbb{R}^n$ and $0 < \lambda < 1$ the inequality

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

holds. That is, the inequality defining the convexity of a function is strict whenever possible.

More graphically, this means that for each pair of points $(x, f(x))$ and $(y, f(y))$ lying on the graph of f, the connecting line segment remains above (or rather: not below) the graph. It is strictly convex if the connecting line segment stays strictly above the graph. See Figure 1.

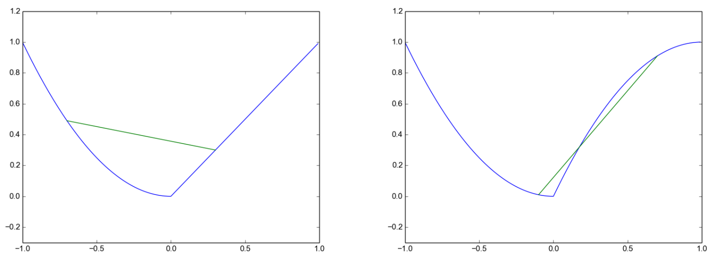

Figure 1. *Left:* Typical example of a convex (but not strictly convex) function. Note that no differentiability is assumed. *Right:* Typical example of a non-convex function. There exist points on the graph such that the connecting line segment does not lie completely above the graph.

Similarly, we can also define convex sets:

Definition 2. A set $C \subset \mathbb{R}^n$ is convex, if for all points x, $y \in C$ and $0 \le \lambda \le 1$ we have

$$\lambda x + (1-\lambda)y \in C.$$

That is, a set is convex, if whenever we are given two points x and y in C the whole line segment connecting these two points is also contained in C.

Date: January 2016.

Remark 3. There is a very close connection between convex sets and convex functions: One can show that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex, if and only if the so-called *epigraph* of f, which is the subset of $\mathbb{R}^n \times \mathbb{R}$ consisting of all points (x, t) with $t \geq f(x)$, is a convex set.

It is easy to show the following properties of convex functions:

- If the functions f, $g\colon \mathbb{R}^n \to \mathbb{R}$ are convex, then so is the function $f + g$.
- If $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex and $\lambda \geq 0$, then also the function λf is convex.
- Every linear (or affine) function is convex.
- If both f and $-f$ are convex, then the function f is affine (that is, $f(x) = a^T x + b$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$).
- If f and g are convex functions, then the function h defined by $h(x) := \max\{f(x), g(x)\}$ is also convex.

2. Differentiable Convex Functions

In the definition of convex functions above, we have not assumed any regularity of f (apart from f only taking finite values). Indeed, one of the main advantages of the (rather extensive) theory of convex functions is that it allows to deal with non-differentiable functions using almost the same methods as we would use for differentiable functions. In particular, it is possible to introduce generalised notions of derivatives that in turn can be used for the characterisation and computation of solutions of optimisation problems. However, we will consider in the following *differentiable* convex functions, and we will study what the convexity of a function implies for its derivative.[1]

Proposition 4. *Assume that the function $f\colon \mathbb{R}^n \to \mathbb{R}$ is differentiable. Then f is convex, if and only if for every x, $y \in \mathbb{R}^n$ the inequality*

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \tag{1}$$

is satisfied.

Proof. Assume first that f is convex and let $x \neq y \in \mathbb{R}^n$. The convexity of f implies that

$$f\big((x + y)/2\big) \leq \frac{1}{2} f(x) + \frac{1}{2} f(y).$$

Denote now $h := y - x$. Then this inequality reads as

$$f(x + h/2) \leq \frac{1}{2} f(x) + \frac{1}{2} f(x + h).$$

Using elementary transformations, this can be rearranged as

$$f(x + h) - f(x) \geq \frac{f(x + h/2) - f(x)}{1/2}.$$

Repeating this line argumentation with x and $x + h/2$ instead of x and $x + h$, we obtain that

$$f(x + h) - f(x) \geq \frac{f(x + h/2) - f(x)}{1/2} \geq \frac{f(x + h/4) - f(x)}{1/4},$$

[1] We will not discuss non-continuous convex functions—the main reason being that they do not exist in the setting used here: It can be shown that every function $f\colon \mathbb{R}^n \to \mathbb{R}$ is not only continuous, but actually locally Lipschitz continuous. Moreover, this implies that convex functions are actually almost everywhere differentiable (in the sense that the set of points where the derivative does not exists has Lebesgue measure zero).

It *is* possible to construct non-continuous convex functions, but only if one either restricts the domain of the function to some (non-open) convex subset C of $\mathbb{R}^n$ (that is, we have a function $f\colon C \to \mathbb{R}$), or one allows the function to take the value $+\infty$.

or, more general,

$$f(x+h) - f(x) \ge \frac{f(x+2^{-k}h) - f(x)}{2^{-k}} \tag{2}$$

for all $k \in \mathbb{N}$. Now recall that the *directional derivative* of the function f at the point x in direction h is defined as

$$Df(x;h) := \lim_{t\to 0} \frac{1}{t}\big(f(x+th) - f(x)\big)$$

and satisfies

$$Df(x;h) = \nabla f(x)^T h.$$

Thus taking the limit $k \to \infty$ in (2) we see that

$$f(x+h) - f(x) \ge \limsup_{k\to\infty} \frac{f(x+2^{-k}h) - f(x)}{2^{-k}} = Df(x;h) = \nabla f(x)^T h.$$

Replacing now h by $y - x$ yields the required inequality.

Assume now that the inequality (1) holds for all x, $y \in \mathbb{R}^n$. Let moreover w, $z \in \mathbb{R}^n$ and $0 \le \lambda \le 1$. Denote moreover

$$x := \lambda w + (1-\lambda)z.$$

Then the inequality (1) implies that

$$\begin{aligned} f(w) &\ge f(x) + \nabla f(x)^T(w-x), \\ f(z) &\ge f(x) + \nabla f(x)^T(z-x). \end{aligned} \tag{3}$$

Note moreover that

$$w - x = (1-\lambda)(w-z) \qquad \text{and} \qquad z - x = \lambda(z-w).$$

Thus, if we multiply the first line in (3) with λ, the second line with $1-\lambda$, and then add the two inequalities, we obtain

$$\begin{aligned} \lambda f(w) + (1-\lambda)f(z) \ge f(x) + \lambda\nabla f(x)^T(1-\lambda)(w-z) + (1-\lambda)\nabla f(x)^T\lambda(z-w) \\ = f\big(\lambda w + (1-\lambda)z\big). \end{aligned}$$

Since w and z were arbitrary, this proves the convexity of f. □

Remark 5. Following basically the same proof as above and strategically replacing inequalities by strict inequalities, one can show that a differentiable function f is strictly convex, if and only if

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

whenever $x \ne y \in \mathbb{R}^n$.

As an immediate consequence of Proposition 4 one obtains the result that the first order necessary condition for a minimiser is, in the case of convex functions, also a sufficient condition. More precisely, the following holds:

Corollary 6. *Assume that $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable. Then x^* is a global minimiser of f, if and only if $\nabla f(x^*) = 0$.*

Proof. First recall that the condition $\nabla f(x^*) = 0$ is, independent of the convexity of f, a necessary condition for x^* to be a global (and indeed already local) minimiser. Thus we only need to show that this condition actually implies that x^* *is* a global minimiser. Assume therefore that $\nabla f(x^*) = 0$ and let $y \in \mathbb{R}^n$. Then Proposition 4 implies that

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

Thus x^* is a global minimiser. □

3. Hessians of Convex Functions

Proposition 7. *A twice differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex, if and only if the Hessian $\nabla^2 f(x)$ is positive semi-definite for all $x \in \mathbb{R}^n$.*

Proof. Assume first that f is convex and let $x \in \mathbb{R}^n$. Define moreover the function $g\colon \mathbb{R}^n \to \mathbb{R}$ setting

$$g(y) := f(y) - \nabla f(x)^T(y - x).$$

Since the mapping $y \mapsto -\nabla f(x)^T(y - x)$ is affine, it follows that g is convex. Moreover

$$\nabla g(y) = \nabla f(y) - \nabla f(x)$$

and

$$\nabla^2 g(y) = \nabla^2 f(y)$$

for all $y \in \mathbb{R}^n$. In particular, $\nabla g(x) = 0$. Thus Corollary 6 implies that x is a global minimiser of g. Now the second order necessary condition for a minimiser implies that $\nabla^2 g(x)$ is positive semi-definite. Since $\nabla^2 g(x) = \nabla^2 f(x)$ and x was arbitrary, this proves that the Hessian of f is positive semi-definite for all $x \in \mathbb{R}^n$.

Now assume that the Hessian $\nabla^2 f(x)$ of f is positive semi-definite for all $x \in \mathbb{R}^n$. Let moreover $x, y \in \mathbb{R}^n$. Then Taylor's theorem implies that

$$f(y) = f(x) + \nabla f(x)^T(y - x) + \frac{1}{2}(y - x)^T \nabla^2 f(x + t(y - x))(y - x)$$

for some $0 \le t \le 1$. Since $\nabla^2 f$ is everywhere positive semi-definite, the quadratic term in this equation is always non-negative. Thus we can estimate

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

Proposition 4 proves now the convexity of f. □

Remark 8. There is *some* relation between the strict convexity of a function f and the positive definiteness of its Hessian. However, this relation is not completely straight-forward. It is possible to show (and actually pretty simple to show) that a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is strictly convex, if its Hessian $\nabla^2 f(x)$ is positive definite for all x. However, the converse direction does not hold: The strict convexity of a function f does not imply that its Hessian is everywhere positive definite. As an example consider the function $f\colon \mathbb{R} \to \mathbb{R}$, $f(x) = x^4$. This function is strictly convex, but $f''(0) = 0$. Still, it is possible to characterise the strict convexity of a univariate function $f\colon \mathbb{R} \to \mathbb{R}$ by the condition that the set of points $x \in \mathbb{R}$ with $f''(x) > 0$ is dense. Thus a twice differentiable function $f\colon \mathbb{R} \to \mathbb{R}$ is strictly convex, if and only if the set $\{x \in \mathbb{R} : f''(x) > 0\}$ is dense in $\mathbb{R}$. To the best of knowledge, there exists no (simple) generalisation of this characterisation to multivariate functions.

4. Summary

From the viewpoint of optimisation, the main results concerning convex functions (that we will need/refer to during this class) are:

- Convexity of a differentiable function can either characterised by the fact that all secants lie above the graph (Definition 1) or that all tangents lie below the graph (Proposition 4).
- If a function $f\colon \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable, then the first order necessary condition for a minimum is actually sufficient. That is, the minimisation of f is equivalent to the solution of the equation

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

- A function f is convex, if its Hessian is everywhere positive semi-definite. This allows us to test whether a given function is convex.
- If the Hessian of a function is everywhere positive definite, then the function is strictly convex. The converse does not hold.

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

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