1 11th of October 2018 — A. Frangioni

Last lecture we left with the task of understanding how to check if a function is convex.

As we already stated for convex sets, we can prove that a function is convex deriving from convex functions, though “convex friendly” operations.

Note

There is a softare called CVX, designed to model convex object. A pretty easy way to check if an object is convex is to try to write it in CVX. If such an operation is possible, then the object is convex.

The following functions are convex:

1. $f(x) = wx$: linear functions are both convex and concave;
2. $f(x) = \frac{1}{2}xQx + qx$ if convex iff $Q \succeq 0$;
3. $f(x) = e^{ax}$ for any $a \in \mathbb{R}$ and $x \in \mathbb{R}$
4. $f(x) = -\log(x)$ for $x > 0$
5. $f(x) = x^a$ for $a \geq 1$ or $a \leq 0$ on $x \geq 0$;
6. $f(x) = \|x\|_p$ for $p \geq 1$;
7. $f(x) = \max\{x_1, \ldots, x_n\}$;
8. for any convex set C, its indicator function

$$\imath_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{if } x \notin C \end{cases} \quad (\text{l.s.c.} \iff C \text{ closed})$$

9. $A \in \mathbb{R}^{n\times n}$ symmetric, eigenvalues customarily ordered $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_n$: $f_m(A) = \sum_{i=1}^{m} \lambda_i$ (sum of m largest eigenvalues)

Fact 1.1. *The following operations preserve convexity:*

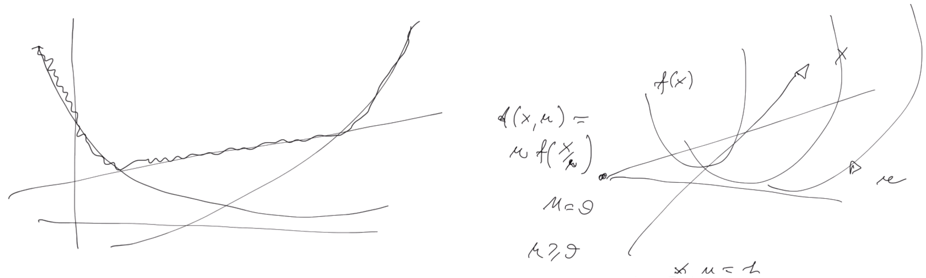
1. *f, g convex, α, $\beta \in \mathbb{R}_+ \implies \alpha f + \beta g$ convex (non-negative combination);*
2. *$\{f_i\}_{i\in I}$ (infinitely many) convex functions $\implies f(x) = \sup_{i\in I} f_i(x)$ convex, see Figure 1.1a;*
3. *Pre-composition with linear function is convex, formally f convex $\implies f(Ax+b)$ convex;*
4. *Post-composition with increasing convex function is convex. Formally, $f : \mathbb{R}^n \to \mathbb{R}$ convex, $g : \mathbb{R} \to \mathbb{R}$ convex increasing $\implies g \circ f = g(f(x))$ is convex;*

5. *f_1, f_2 convex $\implies$ $f(x) = \inf\{ f_1(x_1) + f_2(x_2) \ : \ x_1 + x_2 = x\}$ convex (infimal convolution);*

6. *g convex $\implies f(x) = \inf\{g(y) \ : \ Ay = x\}$ convex (image under a linear mapping, aka value function of convex constrained problem);*

7. *$g(x, y) : \mathbb{R}^{n+m} \to \mathbb{R}$ convex $\implies$ $f(x) = \inf\{g(x, y) \ : \ y \in \mathbb{R}^m\}$ convex (partial minimization);*

8. *$f(x)$ convex $\implies$ $\tilde{f}(x, u) = uf(x/u)$ when $u > 0$, $\tilde{f}(x, u) = \infty$ otherwise, convex (perspective or dilation function of f), see Figure 1.1b.*



(A) Suprema — second property (B) Perspective — eighth property

Figure 1.1: Graphic hints.

Fact 1.2. *Let $f : \mathbb{R}^n \to \mathbb{R} \cup \infty$ a convex function. If $\exists \bar{x} \in dom(f)$ such that $f(\bar{x}) = -\infty$, then $f \equiv -\infty$.*

From now on we will solve the issue of functions with non convex domain, sayting that in those points where the function is not defined, we value it $+\infty$.

Fact 1.3. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then f is Lipschitz continuous $\forall$ bounded convex set $S \subseteq$ int(dom(f)) but on $\partial dom(f)$ (the border of the domain) anything can happen.*

Moreover, a function f, which is continuous but not Lipschitz continuous is not convex.

A couple of examples of Proposition 1.3 can be found in Figure 1.2.

Fact 1.4. *Let convex $f : \mathbb{R}^n \to \mathbb{R}$. Then it is Lipschitz continuous on any bounded set and continuous everywhere.*

It happens often that the set of points in which a function is non differentiable have measure 0.

Theorem 1.5 (Convexity characterization)**.** *Let $f \in C^1$. It is convex on C convex iff*

$$f(y) \geq f(x) + \nabla f(x)(y - x)\forall x, y \in C$$

In other words, given a point x, we compute the derivative and the value fo the function is above the derivative in that point.

(A) Take a compact set in the interior of the domain (far from the boundaries) there the function is Lipschitz continuous.

(B) If a function is not Lipschitz on a compact subset it is not convex.

FIGURE 1.2

Proof. $\Rightarrow$) $\alpha(f(y) - f(x)) \geq f(\alpha(y-x)+x) - f(x)$, send $\alpha \to 0$

$\Leftarrow$) TODO

□

Theorem 1.6. *Let $f \in C^1$ convex. x is a stationary point iff x is a global minimum.*

Fact 1.7. *Let f be twice differentiable (aka has Hessian). f is convex iff the Hessian is positive semidefinite.*

Formally, let $f \in C^2$. f is convex on the open set S iff $\nabla^2 f(x) \succeq 0 \; \forall x \in S$.

This proposition gives us an algorithm to check if a function is convex or not: we only need to compute the eigenvalues of the Hessian and check if they are positive.

There are some functions which do not have differentiability property.

A way to work with functions which are not defined on all $\mathbb{R}$ is to solve the following problem:

$$(P) \equiv \inf\{f_X(x) = f(x) + \text{ß}_X(x) \; : \; x \in \mathbb{R}^n\}$$

thanks to

Theorem 1.8 (Essential objective)**.** *x_* optimal for $(P) \iff x_*$ local minimum of f_X.*

1.0.1 Subgradients and subdifferentials

Definition 1.1 (Subgradient)**.** *For each $s \in \mathbb{R}$ we term **s-subgradient** of f at x as*

$$f(y) \geq f(x) + s(y-x) \qquad \forall\, y \in \mathbb{R}^n$$

Let us assume that the minimum of the non differentiable function resides in one of its kiky points, then for $s = 0$ we have a subgradient which is flat and this is a sufficient condition for minimality, for a pictorial example see Figure 1.4.

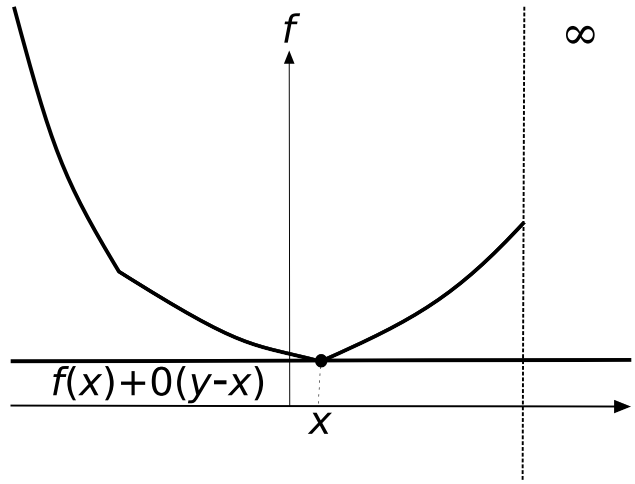


FIGURE 1.3: Pictorial example of subgradients of a non diffrentaible function.

The issue here is that it is unfeasible to check if the subgradient with $s = 0$ is a subgradient for f.

Definition 1.2 (Subdifferential)**.** *We term* ***subdifferential*** *the set of all possible subgradients.*

Formally,

$$\partial f(x) := \{s \in \mathbb{R}^n \ : \ s \text{ is a subgradient at } x\}$$

Theorem 1.9. *x global minimum $\iff 0 \in \partial f(x)$.*

Notice that in general, when we are not in proximity of a border (where f is unbounded above) we get that the subdifferential is a compact interval.

Formally,$\partial f(x)$ closed and convex, compact $\forall\, x \in int\, dom(f)$.

Moreover, we can prove the following

Fact 1.10. *$\partial f(x) = \{\nabla f(x)\} \iff f$ differentiable at x.*

An attentive reader may have noticed that in the case of non differentiable functions it is not possible to derive the directional derivative from the gradient ($\frac{\partial f}{\partial d}(x) =< \nabla f(x), d >$), but we can prove the following

Fact 1.11. *$\frac{\partial f}{\partial d}(x) = \sup\{< s, d > : \ s \in \partial f(x)\} \implies d$ is a descent direction $\iff < s, d >< 0 \ \forall \ s \in \partial f(x)$.*

As in the differentiable case, we are interested in moving in the steepest descent direction, formally $s_* = -\text{argmin}\{\|s\| \ : \ s \in \partial f(x)\}$.

Example 1.1. *Le us assume we are in x and we want to move towards x^* knowing only the subdifferentials. $(-)\partial f(x)$ is convex and compact and All $(-)g \in \partial f(x)$ "point towards x_*": $< g, x - x_* >< 0$.*

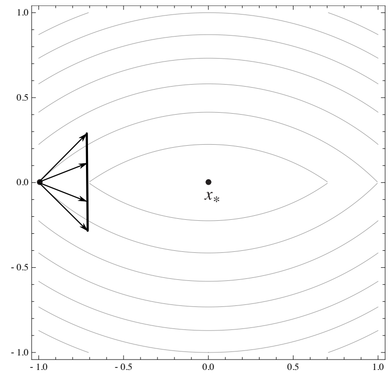


FIGURE 1.4: There are many different subgradients in x. We pick the one with minimum norm among the ones which have a negative scalar product with $x - x^*$.

Not all of them are descent directions, but the (opposite to the) minimum-norm one is a descent direction.

Notice that in $\mathbb{R}^2$ if we take a function and compute the subdifferential, we can scale both the function and the subdifferential by any positive constant (negative contants would lead to concave functions). Formally,

Fact 1.12. *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ and take $\alpha, \beta \in \mathbb{R}_+$, then $\partial[\alpha f + \beta g](x) = \alpha \partial f(x) + \beta \partial g(x)$.*

Fact 1.13 (Chain rule)**.**
- *Let $f : \mathbb{R}^n \to \mathbb{R}$ and let $A \in M(n, \mathbb{R})$ and $b \in R^n$ then $\partial[f(Ax + b)] = A^T[\partial f](Ax + b)$;*
- *Let $g : \mathbb{R} \to \mathbb{R}$ increasing, then $\partial[g(f(x))] = [\partial g](f(x))[\partial f](x)$.*

Definition 1.3 (ε-subgradient)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. s is **ε-subgradient** at x if*

$$f(y) \geq f(x) + s(y - x) - \varepsilon \ \forall y \in \mathbb{R}^n$$

support hyperplane passing ε below epi(f)

Fact 1.14. *Given a point x, the value of the finction in x cannot be further from ε the minimum value fo the function. Formally, $0 \in \partial_\varepsilon f(x) \iff x$ is ε-optimal.*

We are now allowed to compute $s_* = \operatorname{argmin}\{\|s\| \ : \ s \in \partial_\varepsilon f(x)\}$.
If $s_* = 0$ then x is ε optimal. Otherwise, $\exists \alpha > 0$ s.t. $f(x - \alpha s_*) \leq f(x) - \varepsilon$ ($-s_*$ is of ε-descent).
The ε-subgradient is very powerful, but the issue is that is even more expensive to compute than the subgradient.