Lecture 3

Last time

- The setting
- convex sets
- Separation

Today

- Convex functions
- Continuity
- Gradients

Convex functions

We can bootstrap ourselves from the set definitions.

Def: Given a function $f : E \to \overbrace{\mathbb{R} \cup \{\infty\}}^{\overline{\mathbb{R}}}$, its **epigraph** is given by

$$\operatorname{epi} f := \{ (x, t) \in E \times \overline{\mathbb{R}} \mid f(x) \leq t \}.$$

$\dashv$

Example

Def: A function $f : E \to \overline{\mathbb{R}}$ is convex if epi f is convex.

We can also pass from functions to sets.

Def: For any set C, define its indicator function is

$$\mathcal{I}_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{otherwise.} \end{cases}$$

Exercise: Show that C is convex if, and only if, $\mathcal{I}_C$ is convex.

Def: We say that $f: C \to \mathbb{R}$ is convex if $f + \mathcal{I}_C$ is convex.
Equivalently, C is convex and

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

$$\forall x, y \in S, \lambda \in [0,1].$$

The function f is strictly convex if the inequality above is strict in $(0,1)$.

Examples

- Any norm is convex.
- Any norm squared $\|\cdot\|_*^2$ is strictly convex (why?) $\dashv$

Def: We say that $x^* \in E$ is a (global) minimizer of $f: E \to \overline{\mathbb{R}}$ if
$$f(x^*) \leq f(x) \qquad \forall x \in E.$$
We say x^* is a local minimizer if if $\exists \delta > 0$ st x^* is a minimizer of $f + z_u$ with $u = x^* + \delta B$. $\dashv$

Convex functions are particularly nice for optimization.

Proposition For convex f, local minimizers are minimizers. $\dashv$

Proof: Convexity can be understood through 1-dimensional lens.

Claim(★): Suppose that $g: \mathbb{R}^+ \to \mathbb{R}$

is convex with $g(0)=0$. Then, $t \mapsto \frac{g(t)}{t}$ is nondecreasing.

Proof of the claim:

Take $t_1 \leq t_2$, then

$$g(t_1) = g\left(\left(1 - \frac{t_1}{t_2}\right)\cdot 0 + \frac{t_1}{t_2}\cdot t_2\right)$$

$$\leq (1 - t_1/t_2)\cdot 0 + \frac{t_1}{t_2} g(t_2)$$

$\square$

Define $g(t) = f(\underbrace{\bar{x}}_{\text{local min}} + t(\underbrace{x}_{\text{any } x} - \bar{x})) - f(\bar{x})$, then

$$0 = g(0) \leq g(1) \leq f(x) - f(\bar{x}),$$

which shows the result. $\square$

Furthermore, just like convex sets, convex functions enjoy of neat topological properties.

Define

$$\text{dom } f := \{x \in E \mid f(x) < \infty\}.$$

Theorem (4) Suposse $f: E \to \bar{\mathbb{R}}$ is con vex with $\bar{x} \in \operatorname{dom} f$. Then, f is locally Lipschitz near $\bar{x}$ if, and only if, f is bounded above on a neighborhood of $\bar{x}$.

$\exists\, \delta, L > 0$ s.t.

$$|f(x) - f(y)| \le L\|x-y\| \quad \forall x, y \in \bar{x} + \delta B. \tag{$\star$}$$

Proof: "$\Longrightarrow$" This follows trivially.

"$\Longleftarrow$" WLOG suppose $\bar{x} = 0$ and $f(0) = 0$.

Suppose there exist $a, \varepsilon > 0$ s.t.

$$f(x) \le a \qquad \forall x \in \varepsilon B.$$

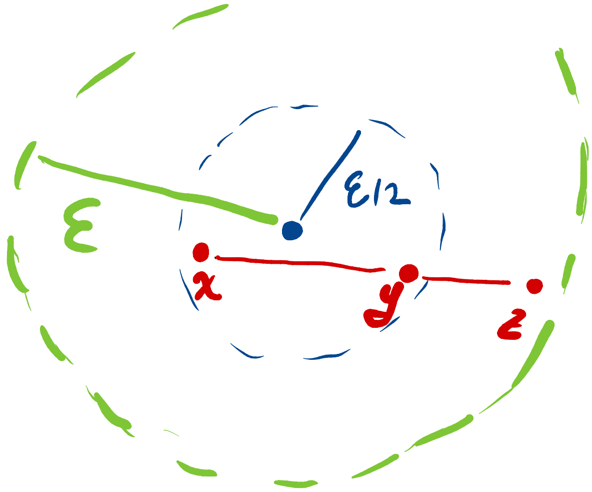
We will prove that $(\star)$ holds with $L = \frac{2a}{\varepsilon}$ and $\delta = \varepsilon/2$.

We can write $0 = \frac{1}{2} x + \frac{1}{2}(-x)$.

So, $0 = f(\bar{x}) \le \frac{1}{2} f(x) + \frac{1}{2} f(-x)$

$$\Rightarrow f(x) \ge -f(-x) \ge -a \quad \forall x \in \varepsilon B. \tag{$\heartsuit$}$$

Consider $x, y \in \frac{\varepsilon}{2} B$ and let $z = y + \varepsilon \frac{(y-x)}{\|y-x\|} \in \varepsilon B$.



Thus, y belongs to the segment $[x, z]$. Indeed we can write

$$y = \frac{\varepsilon}{\varepsilon + \|y-x\|} x + \frac{\|y-x\|}{\varepsilon + \|y-x\|} z .$$

By convexity

$$f(y) - f(x) \le \frac{\varepsilon}{\varepsilon + \|y-x\|} f(x) + \frac{\|y-x\|}{\varepsilon + \|y-x\|} f(z) - f(x)$$

$$= \frac{\|y-x\|}{\varepsilon + \|y-x\|} \left(f(x) - f(z) \right)$$

$$\le \frac{2a}{\varepsilon} \|y-x\| .$$

An analogous argument yields $f(x) - f(y) \le \frac{2a}{\varepsilon} \|x-y\|$, proving the result.

$\square$

We can extend this result to constrained functions.

Corollary: Let $C \subseteq E$ and $f: C \to \mathbb{R}$ convex. Then, f is continuous on $\operatorname{int} C$.

Proof: Exercise. $\square$

Gradients

Let's introduce standard definitions of smoothness.

Def: Let $\mathcal{U} \subseteq E$ be open, $f: \mathcal{U} \to \mathbb{R}$ and $\bar{x} \in \mathcal{U}$. We say that f is **differentiable** at $\bar{x}$ if $\exists g \in E$ s.t.

$$\lim_{h \to 0} \frac{f(\bar{x} + ty) - (f(\bar{x}) + \langle g, h \rangle)}{\|h\|} = 0.$$

We write $\nabla f(x) = g$. If ∇f is continuous we write $f \in C^1$.

Given that we don't necessarily have coordinates, we can write second derivatives in terms of 1D slices.

Def: We say that f is twice continuously differentiable ($f \in C^2$) if

$h_y(x) = \langle y, \nabla f(x) \rangle \in C^1 \quad \forall y \in E.$

Further we write

$$\nabla^2 f(x)[y,y] = \langle y, \nabla h_y(x) \rangle.$$

In $\mathbb{R}^d$ this matches $y^T \nabla^2 f(x) y$.

$\dashv$

Theorem (Taylor Approximation) Suppose $f \in C^2$ and let $x, y \in E$. If we let $p(t) := f(x + ty)$, we have

$p(t) = f(x) + t \langle \nabla f(x), y \rangle + \frac{t^2}{2} \nabla^2 f(x)[y,y] + o(t^2)$

with $p \in C^2$ and $p''(0) = \nabla^2 f(x)[y,y]$.

$\dashv$

Lemma (Smooth convex functions) Suppose that $f : E \to \mathbb{R}$ is differentiable. Then, the following are equivalent:

(1) f is convex.

(2) $\forall x, y \in E \quad f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle.$

(3) $\forall x, y \in E \quad \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$

Moreover if $f \in C^2$, these are equivalent to

(4) $\forall x, y \in E \quad D^2 f(x)[y, y] \geq 0$.

Both of these were proved in Nonlinear 1.