* Our standard optimization problem:

$$\min_{\boxed{x} \in \boxed{\mathcal{X}}} \boxed{f(x)} \qquad (P)$$

vector   feasible set   objective function

$$\max_{x \in \mathcal{X}} f(x) \iff \min_{x \in \mathcal{X}} -f(x)$$

* the value of (P) : $\mathrm{val}(P) = \inf_{x \in \mathcal{X}} f(x)$.
* To give (P) a meaning, we need to specify:
  - vector space, feasible set, objective function
  - what it means to "solve" (P)

Q. Can we even hope to solve an arbitrary opt. problem?

Ex. Can you come up with an example of positive integers x, y, z s.t.

$$x^2 + y^2 = z^2 \Leftarrow \text{(Pythagorean triples)}$$

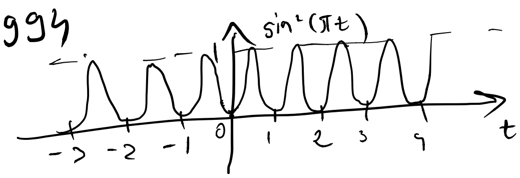
$(3,4,5)$; $(5,12,13)$; $(8,15,17)$ ....

How about $x^3 + y^3 = z^3$? $\Leftarrow$

* Fermat's conjecture [Fermat's Last Theorem] (1637)

For any $n \geq 3$, $x^n + y^n = z^n$ has no solutions over positive integers.

→ Proved by Andrew Wiles in 1994

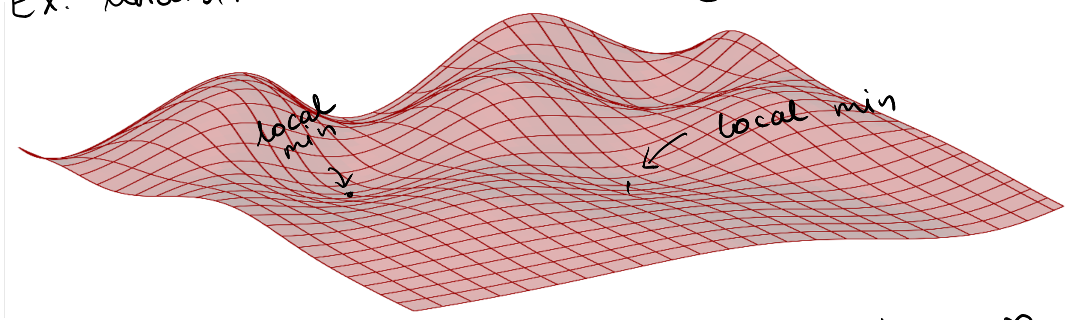


Consider:

$$(P_F) \begin{cases} \min_{x,y,z,n} (x^n + y^n - z^n)^2 \\ \text{s.t. } x \geq 1,\ y \geq 1,\ z \geq 1,\ n \geq 3 \\ \sin^2(\pi n) + \sin^2(\pi x) + \sin^2(\pi y) + \sin^2(\pi z) = 0 \end{cases}$$

If you could certify whether $\mathrm{val}(P_F) \neq 0$, you would have found a proof for Fermat's conjecture.

Ex. Unconstrained optimization, many local minima:



" Arbitrary optimization problems are hopeless, we always need some structure "

* Specifying the optimization problem:

$$\min_{x \in \mathcal{X}} f(x) \qquad (P)$$

① Vector space (where the optimization variables and the feasible set "live")

$$(\underbrace{\mathbb{R}^d}_{\text{tells us that } x \text{ is a vector in } \mathbb{R}^d}, \underbrace{\|\cdot\|}_{\text{tells us how to measure distances in } \mathbb{R}^d}) \text{ : normed vector space; "primal space"}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

* Most often, we will take $\|x\| = \|x\|_2 = \left(\sum_{i=1}^{d} x_i^2 \right)^{1/2}$ (Euclidean norm)

* We might sometimes also consider ℓ_p norms:

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

$$\boxed{\|x\|_1 = \sum_i |x_i|}, \quad \boxed{\|x\|_\infty = \max_{1 \leq i \leq d} |x_i|}.$$

* We will use $\langle \cdot, \cdot \rangle$ to denote inner products.

standard inner product:

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{d} x_i y_i .$$

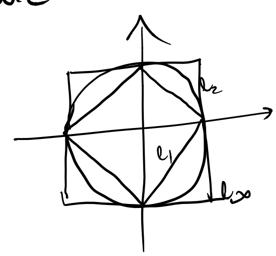
[! When we work with $(\mathbb{R}^d, \|\cdot\|_p)$, view $\langle y, x \rangle$ as the value of a linear function y at x. So, if we are measuring $\overset{\text{the length}}{\checkmark}$ x using the $\|\cdot\|_p$, we should measure the length of y using $\|\cdot\|_{p^*}$, where $\frac{1}{p} + \frac{1}{p^*} = 1$.]

dual norm: $\|z\|_* = \sup\limits_{\|x\| \le 1} \langle z, x \rangle$

$\Rightarrow \forall z, x$: $\boxed{\langle z, x \rangle \le \|z\|_* \cdot \|x\|}$.

pf: Fix any two vectors x, z. Assume $x \neq 0, z \neq 0$, o.w. trivial. Define $\hat{x} = \frac{x}{\|x\|}$.

$$\|z\|_* \geq \langle z, \hat{x} \rangle = \frac{\langle z, x \rangle}{\|x\|} .$$



② Feasible set:

- specifies what solution points we are allowed to output $\mathcal{X} \subseteq \mathbb{R}^d$. If $\mathcal{X} = \mathbb{R}^d$, we say that (P) is unconstrained. O.w., we say that (P) is constrained.

$\mathcal{X}$ can be specified:

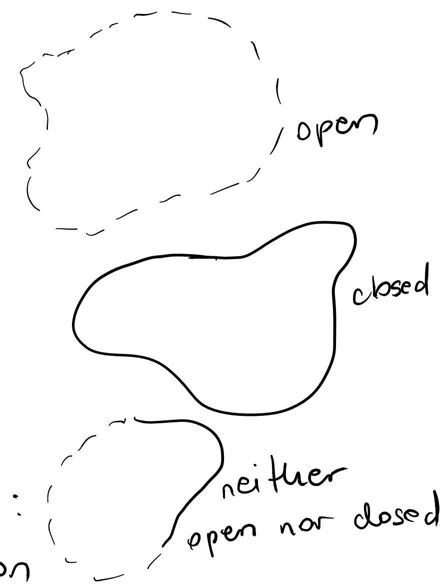
- as an abstract geometric body (a ball, a box, a polyhedron)
- via functional constraints: $g_i(x) \le 0$, $i = 1, 2, \dots, m$, $h_i(x) = 0$, $i = 1, \dots, p$.

$$f_i(x) \geq C \qquad g_i(x) = C - f_i(x)$$

E.g., $\mathcal{X} = \underbrace{\mathcal{B}_2(0,1)}_{\text{unit Euclidean ball}}$

$$\mathcal{X} = \{ x \in \mathbb{R}^d : \|x\|_2 \le 1 \}$$

* In this class, we will always assume that $\mathcal{X}$ is closed and convex.



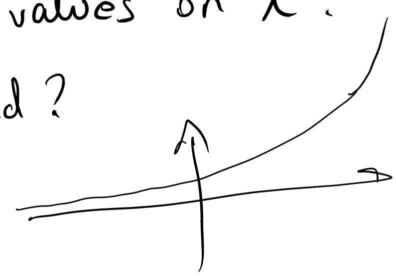
* Heine-Borel Thm: If $\mathcal{X}$ is closed and bounded, then it is compact [if $\mathcal{X} \subseteq \bigcup_{\alpha \in A} U_\alpha$ for some family of open sets $\{U_\alpha\}$, then $\exists$ a finite subfamily $\{U_{\alpha_i}\}_{i=1}^{n}$ s.t. $\mathcal{X} \subseteq \bigcup_{1 \le i \le n} U_{\alpha_i}$]

* Weierstrass Extreme Value Theorem: If $\mathcal{X}$ is compact and f is a function that is defined and continuous on $\mathcal{X}$, then f attains its extreme values on $\mathcal{X}$.

* What if $\mathcal{X}$ is not bounded?

Consider $f(x) = e^x$

$$\inf_{x \in \mathbb{R}} f(x) = 0.$$



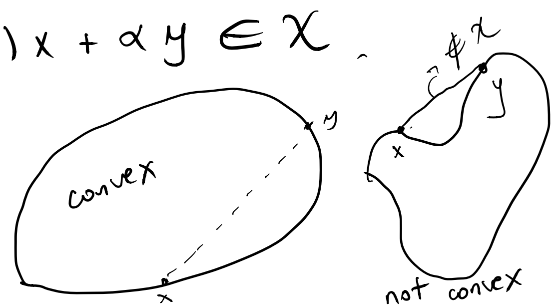
* When we work with unconstrained problems, we will normally assume that f is bounded below.

* Convex sets:

* Def. A set $\mathcal{X} \subseteq \mathbb{R}^d$ is convex if

$$(\forall x, y \in \mathcal{X}) (\forall \alpha \in (0,1)): (1-\alpha)x + \alpha y \in \mathcal{X}.$$

$$x + \alpha(y - x).$$



③ Objective function:

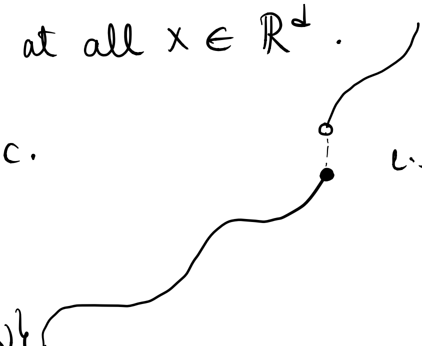
- "cost", "loss"
- Extended real valued functions: $f : \underset{\subseteq \mathbb{R}^d \text{ (domain)}}{\mathcal{D}} \to \mathbb{R} \cup \{-\infty, +\infty\} \equiv \bar{\mathbb{R}}$.
- We will define f on all of $\mathbb{R}^d$ by assigning it value $+\infty$ at each point $x \in \mathbb{R}^d \setminus \mathcal{D}$.
- Effective domain: $\mathrm{dom}(f) = \{ x \in \mathbb{R}^d : f(x) < \infty \}$.
- "nonlinear opt" $\overset{?}{=}$ "continuous opt"

* Lower semicontinuous functions:

* **Def.** A function $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ is said to be lower semicontinuous (l.s.c.) at $x \in \mathbb{R}^d$ if

$$\liminf_{y \to x} f(y) \geq f(x).$$

f is l.s.c. on $\mathbb{R}^d$ if it is l.s.c. at all $x \in \mathbb{R}^d$.



l.s.c.

Ex. indicator of a closed set is l.s.c.

$$I_{\mathcal{X}}(x) = \begin{cases} 0, & x \in \mathcal{X} \\ \infty, & x \notin \mathcal{X} \end{cases}.$$

$$\min_{x \in \mathcal{X}} f(x) \equiv \min_{x \in \mathbb{R}^d} \{ f(x) + I_{\mathcal{X}}(x) \}$$

* Unless we are abstracting away constraints, the least we will assume about f is that it is continuous.

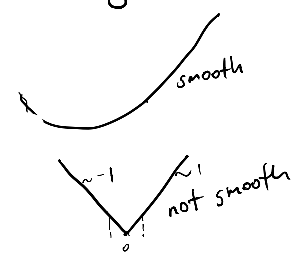
* **Def.** $f : \mathbb{R}^d \to \bar{\mathbb{R}}$ is said to be:

1) Lipschitz-cont. on $\mathcal{X} \subseteq \mathbb{R}^d$ if $\exists M < \infty$

$$\forall x, y \in \mathcal{X} : \quad |f(x) - f(y)| \leq M \|x - y\|.$$

2) Smooth on $\mathcal{X} \subseteq \mathbb{R}^d$ if f's gradients are Lipschitz-cont.'s, i.e., $\exists L < \infty$ s.t. $\forall x, y \in \mathcal{X}$:

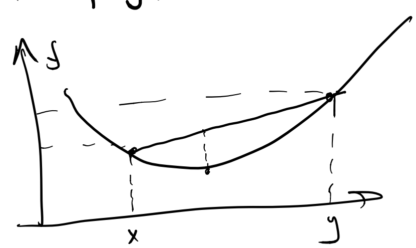
$$\|\nabla f(x) - \nabla f(y)\|_* \leq L \|x - y\| .$$



$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{bmatrix}$$

* **Def.** $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is convex if $\forall x, y \in \mathbb{R}^d$, $\forall \alpha \in (0,1)$:

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



* Ex. Function that is differentiable on its domain but not smooth:

$$f(x) = \frac{1}{x}$$

$$\mathrm{dom}(f) = \mathbb{R}_{++}$$