Last time:

GD: $\boxed{x^{(t+1)} = x^{(t)} - \mu \nabla f(x^{(t)})}$

1. We gave an interpretation of the above GD update step whereby

$x^{(t+1)}$ minimizes the fⁿ

$$g(z) = f(x^{(t)}) + \nabla f(x^{(t)})^T (z - x^{(t)}) + \frac{1}{2\mu} \|z - x^{(t)}\|^2$$

2. We talked about the method of steepest descent:

GD belongs to a class of algorithms where

$$x^{(t+1)} = x^{(t)} - \mu p^{(t)}$$

chosen iteratively

- Specifically GD selects $p^{(t)}$ to minimize

$$\min_p \underline{\underline{\frac{\nabla f(x^{(t)})^T p}{\|p\|_2}}}$$

- Other algorithms can be derived by replacing $\|p\|_2$ by other norms

e.g. We obtain coordinate descent when we pick p to minimize

$$\min_p \underline{\underline{\frac{\nabla f(x^{(t)})^T p}{\|p\|_1}}}$$

Here we end up with

j^* = index of largest $|\ |$ entry of $\nabla f(x^{(t)})$

$$p^* = -\text{sign}\left(\nabla_{j^*} f(x^{(t)})\right) \|\nabla f(x^{(t)})\|_\infty \vec{e}_{j^*}$$

vector of all zeros except at j^* entry where it is 1

$$= -\frac{\partial f(x^{(t)})}{\partial x_{j^*}} \vec{e}_{j^*} \qquad \left(= -\nabla_{j^*} f(x^{(t)}) \vec{e}_{j^*}\right)$$

Gradient Descent under Constraints:

We have been solving

$$\boxed{\min_{x \in \mathbb{R}^n} f(x)}$$

by running the iterations

$$\boxed{x^{(t+1)} = x^{(t)} - \mu \nabla f(x^{(t)})}$$

But what changes do we need to make to GD if we now want to solve:

$$\boxed{\min f(x) \quad \text{subject to} \quad x \in \Omega}$$

$\Omega \subset \mathbb{R}^n$

The issue is that even if $x^{(0)}$ or any $x^{(t)} \in \Omega$, there is no guarantee that $x^{(t+1)}$ as given by GD is also in Ω.

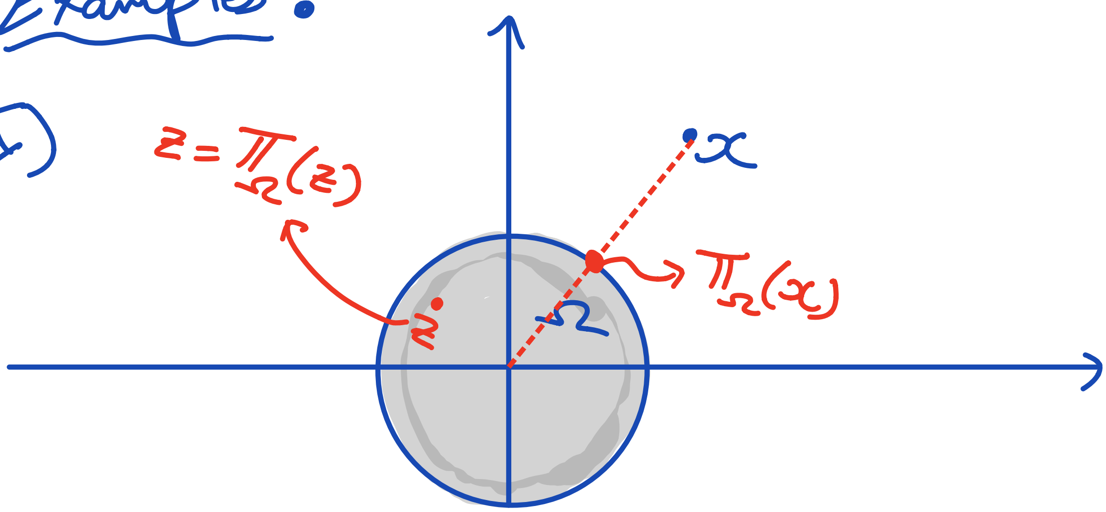
We'll need the following def'n

The projection of a point x onto a set Ω is defined as the closest pt in Ω, to x. That is, it is defined as

$$\Pi_\Omega(x) = \underset{y \in \Omega}{\operatorname{argmin}} \|x - y\|$$

Examples:

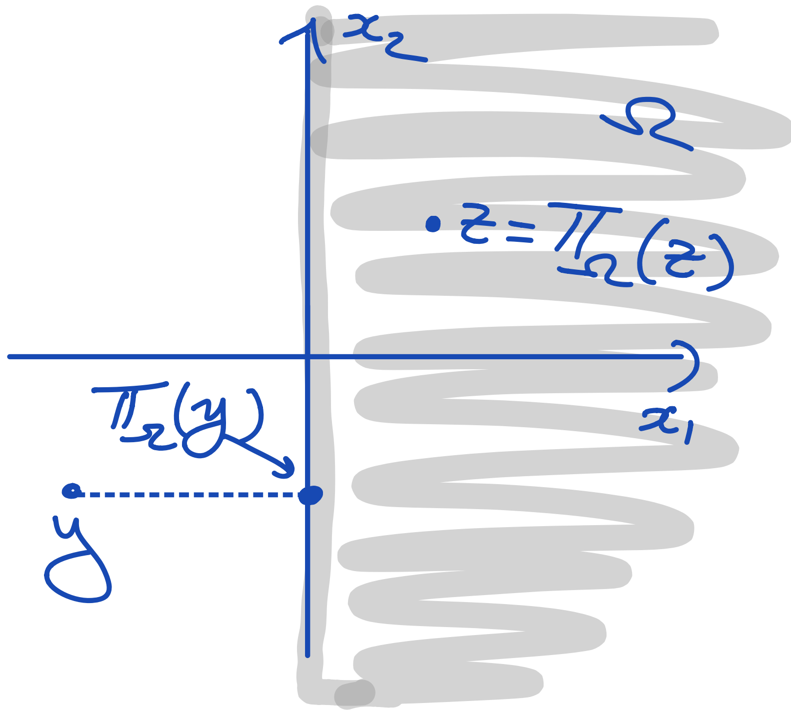
(I)



If $\Omega = B_2^n := \{x : \|x\| \leq 1\}$

then $\Pi_\Omega(x) = \begin{cases} x, & \|x\| \leq 1 \\ \dfrac{x}{\|x\|}, & \|x\| > 1 \end{cases}$

(ii) If $\Omega = \{x : x_1 \geq 0\}$



then $\Pi_\Omega(x) = \begin{cases} x & \text{if } x_1 \geq 0 \\ (0, x_2, x_3, \ldots, x_n) & \text{if } x_1 < 0 \end{cases}$

Projected GD:

In order to solve where Ω convex

$$\min f(x) \quad \text{subject to} \quad x \in \Omega$$

we run the iteration

$$x^{(t+1)} = \Pi_{\Omega}\Big(\underbrace{x^{(t)} - \mu^{(t)} \nabla f(x^{(t)})}_{\text{standard GD step}}\Big)$$

projected onto Ω

$$\Updownarrow$$

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

$$x^{(t+1)} = \arg\min_{x \in \Omega} \| y^{(t+1)} - x \|$$

In general, solving

$$x^{(t+1)} = \operatorname*{argmin}_{x \in \Omega} \| y^{(t+1)} - x \|$$

is itself a constrained opt. problem that may not have a nice closed form solution.

That said, there are interesting cases, where Ω is nice (like the examples above) and where $\Pi_\Omega(x)$ is easy to compute.

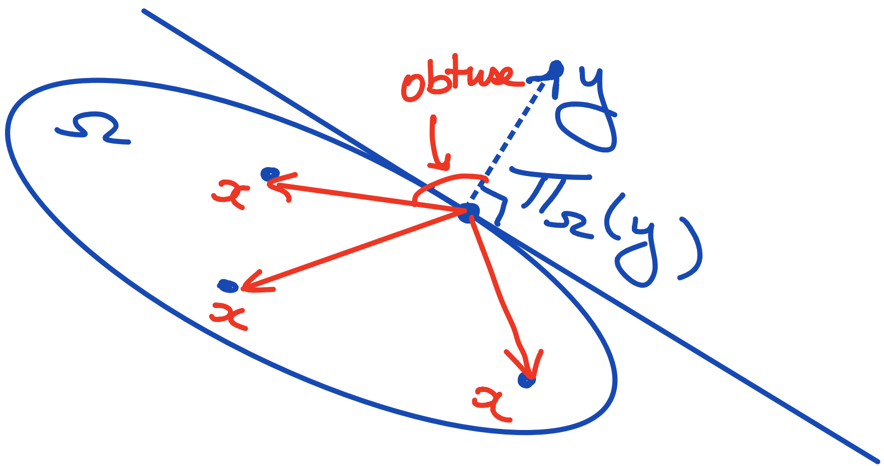
Some facts about projections:

Lemma: If $\Omega \subset \mathbb{R}^n$ is convex and closed (and not empty) and if $x \in \Omega$, $y \in \mathbb{R}^n$

then (1) $\boxed{\Pi_\Omega(y) \text{ is unique.}}$

(2) $\boxed{\Pi_\Omega(\Pi_\Omega(y)) = \Pi_\Omega(y)}$

$$(3) \quad \boxed{\langle \Pi_\Omega(y) - x, \Pi_\Omega(y) - y \rangle \leq 0}$$



So $\langle y - \Pi_\Omega(y), \Pi_\Omega(y) \rangle \geq \langle y - \Pi_\Omega(y), x \rangle$

$\forall x \in \Omega$

$$(4) \quad \boxed{\begin{aligned} &\|\Pi_\Omega(y) - x\|^2 + \|y - \Pi_\Omega(y)\|^2 \\ &\leq \|y - x\|^2 \end{aligned}}$$

$$(5) \quad \boxed{\|\Pi_\Omega(y) - x\| \leq \|y - x\|}$$

Ex: $\max_{\|x\|_\infty \leq 1} \langle y, x \rangle$

$$\leq \max_{\|x\|_\infty \leq 1} \|y\|_1 \cancelto{1}{\|x\|_\infty}$$

Find x s.t. $\langle y, x \rangle = \|y\|_1$

$$\|y\|_1 = \sum_i |y_i|$$

$$\langle y, x \rangle = \sum_i y_i x_i$$

$\Rightarrow$ if $x_i = \text{sign}(y_i)$, then

$$\langle y, x \rangle = \sum_i |y_i| \quad \checkmark$$

Application: Fund opt.

r: vector of expected returns for assets

x: fraction invested in each asset

μ: desired expected return on fund

σ^2: desired expected variance on fund

$$\mu = r^T x$$

asset covariance

$$\sigma^2 = x^T \Sigma x$$

Problem: $\min \sigma^2$

s.t. μ fixed

fraction bought ≥ 0

$\Downarrow$

$$\min_x \quad x^T \Sigma x$$

$$\text{s.t.} \quad r^T x = \mu$$

$$\mathbf{1}^T x = 1$$

$$x \in \{x : x_i \geq 0 \ \forall i\}$$

OR

$$\max_x \quad r^T x$$

$$\text{s.t.} \quad x^T \Sigma x = \sigma^2$$

$$\mathbf{1}^T x = 1$$

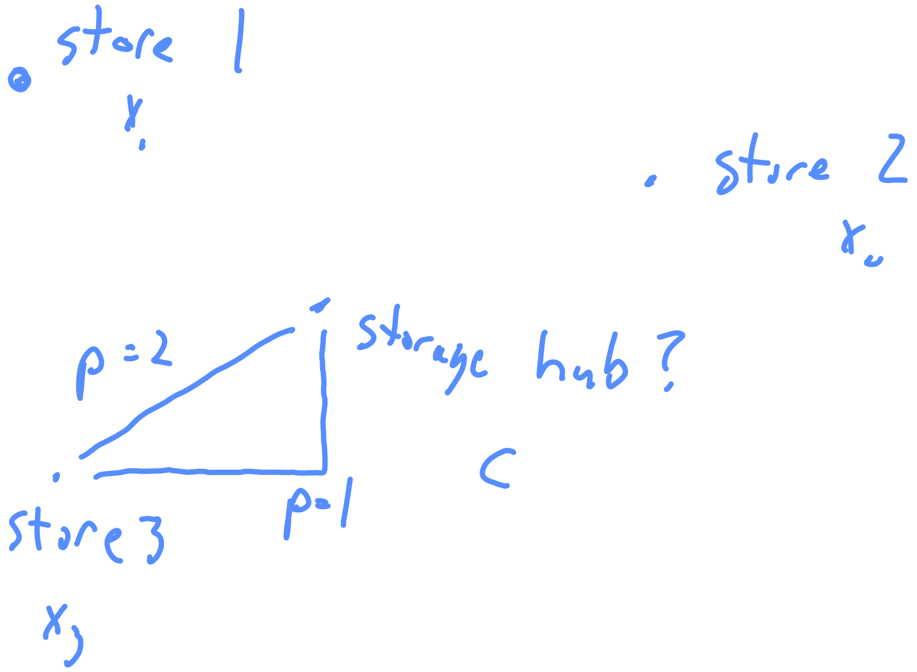
$$x_i \geq 0 \quad \forall i$$

Transaction cost: Return $r^T x - c_- x_- - c_+ x_+$

$$\mathbf{1}^T x + c_-^T x_- + c_+^T x_+ = 1$$

$$x_i = \bar{x}_i + x_{+i} - x_{-i}$$

Application: Facility location



$$\min_{c} \sum_{i=1}^{n} \| x_i - c \|_p$$

If $p=2$, & took cost $\|x_i - c\|_p^2$, then

$$c^* = \frac{1}{n} \sum_i x_i$$

$p=1 \Rightarrow$ roads are on grid