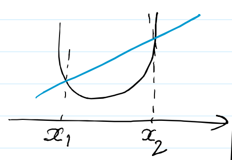
* Exam 2 discussion
* Grading scheme discussion
* HW5 to be released soon
* Next topics: Convex optimization, Online Learning, Complexity

Convex Functions

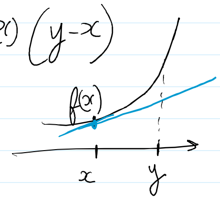
1-dim $f: \mathbb{R} \to \mathbb{R}$ E.g: $f(x) = (ax - b)^2$

(a) 0-th order:
For any $\lambda \in [0,1]$



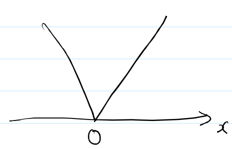
$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$

(b) 1st order:
$$f(y) \ge f(x) + f'(x)(y-x)$$



(c) 2nd order defn: $f''(x) \ge 0$

E.g. $f(x) = |x|$



n-dim $f: \mathbb{R}^n \to \mathbb{R}$ E.g., $f(x) = \|Ax - b\|_2^2$

(a) 0-th order: $\forall x_1, x_2 \in \mathbb{R}^n$
$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2)$$

(b) 1st-order defn: $\forall x, y \in \mathbb{R}^n$
$$f(y) \ge f(x) + \langle \nabla f(x), y - x \rangle$$
$\left(\ldots, \frac{\partial f}{\partial x_i}, \ldots\right)$

(c) 2nd order:
$$\underbrace{\nabla^2 f(x)}_{\begin{array}{c|c} & j \\ \hline i & \begin{bmatrix} \cdots \frac{\partial^2 f}{\partial x_i \partial x_j} \end{bmatrix} \end{array}} \succeq 0$$

Hessian

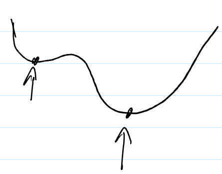
positive semidefinite (PSD)
(i) symmetric
(ii) non-neg eigenvalues

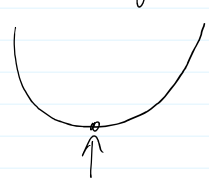
Convex Optimization

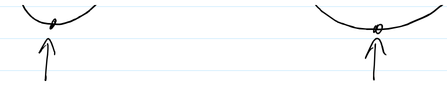
Given a convex fn f and convex body K, find
$$\min_{x \in K} f(x)$$

Some oracle or explicit form

Lemma: For convex optim, $\overbrace{\text{local opt}}^{\text{All nbrs can only be higher: } \nabla f(x) = 0} \Rightarrow$ global opt.







Pf: Sps $x \in K$ is a local opt but $f(x) > f(x^*)$

For $\lambda > 0$, $f(\underbrace{\lambda x + (1-\lambda) x^*}_{\in K}) \le \lambda f(x) + (1-\lambda) \underbrace{f(x^*)}_{< f(x)} < f(x)$

Now for $\lambda \to 1$, we get a contradiction that x is local opt. ∎

Gradient Descent (assume no constraints K)

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

Since we want to find a local opt,

take a step in $-\nabla f(x)$ ← direction of steepest descent

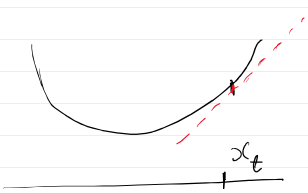
Grad Descent Alg:

1) Start with x_1

2) In t-th step:

$$x_{t+1} = x_t - \underbrace{\eta_t}_{\text{Step size}} \nabla f(x_t)$$

> Exercise: $\lim_{\epsilon \to 0} \max_{y : \|y\| = \epsilon} \frac{f(x+y) - f(x)}{\epsilon} = \|\nabla f(x)\|_2$

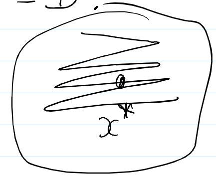


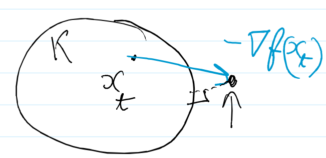
→ We will show that after $\text{poly}(1/\epsilon)$ steps, G.D. finds ϵ-optimal soln.

Properties:
(1) G.D. is fast when ϵ is not very small.
(2) Only requires gradient oracle
(3) Well-defined (and works) even for non-convex fns to give local-opt.
(4) Similar results even hold with convex constraints K, but then we need to project onto K

Thm: Assume $\|\nabla f(x)\|_2 \le G$ and $\|x_1 - x^*\| \le D$.

After T steps of G.D with $\eta_t = \frac{D}{G\sqrt{T}}$,





$f\left(\sum x_t \ldots\right) \quad f(x^*) \le GD \ldots \quad T \ge G^2 D^2$

$G\sqrt{T}$,

$$f\left(\frac{\sum_{t=1}^{T} x_t}{T}\right) - f(x^*) \leq \frac{GD}{\sqrt{T}} \implies \text{For } \epsilon \text{ error, } T \geq \frac{G^2 D^2}{\epsilon^2}.$$

Obs: Suffices to prove $\sum_{t=1}^{T}\left[\frac{f(x_t) - f(x^*)}{T}\right] \leq \frac{GD}{\sqrt{T}}$

since $f\left(\frac{\sum_t x_t}{T}\right) \leq \sum_t \frac{f(x_t)}{T}$

Pf: Define a potential function $\phi_t = \|x_t - x^*\|^2 \cdot \frac{1}{2\eta}$

We will show that

$$f(x_t) - f(x^*) \leq \left(\phi_t - \phi_{t+1}\right) + \frac{\eta G^2}{2} \qquad \text{(1)}$$

This suffices since $\sum_t \left(f(x_t) - f(x^*)\right) \leq \phi_1 - \phi_{T+1} + \frac{\eta G^2}{2} T$

$$\leq \frac{D^2}{2\eta} + \frac{\eta G^2 T}{2} = DG\sqrt{T}.$$

Finally, to prove (1),

$$2\eta \cdot \phi_{t+1} = \|x_{t+1} - x^*\|^2$$

$$= \|x_t - \eta \nabla f(x_t) - x^*\|^2$$

$$= \|x_t - x^*\|^2 + \eta^2 \underbrace{\|\nabla f(x_t)\|^2}_{\leq G^2} + 2\eta \underbrace{\langle \nabla f(x_t), x^* - x_t \rangle}_{\leq f(x^*) - f(x_t)}$$

$$\implies f(x_t) - f(x^*) \leq \phi_t - \phi_{t+1} + \frac{\eta G^2}{2}.$$

∎