MAT2 - INTRO

Setup: $D \subset \mathbb{R}^n$, $f: D \to \mathbb{R}$

Main task: find a min. x^* of f.

Strategy: educated guessing: $x_1, x_2, x_3, \ldots \to$

- $\to$ 0-order methods
- $\to$ 1st order methods (use f and ∇f GD, QN, ...)
- $\to$ 2nd order methods (use f, ∇f and ... Newtons methods)

Def: $x_{k+1} = x_k - \underbrace{\gamma \cdot \nabla f(x_k)}_{\text{STEP}}$

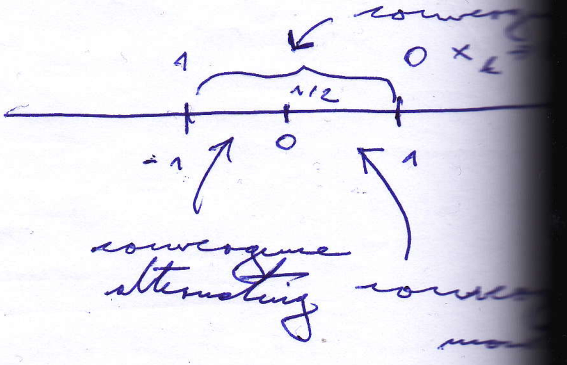
Big Q:

How to choose gamma?

$\boxed{\text{Example: GD on } f(x) = x^2,\ x^* = 0,\ x_1 = 1}$

$$\begin{aligned}
x_{k+1} &= x_k - \gamma f'(x_k) = \\
&= x_k - \gamma\, 2x_k = \\
&= (1 - 2\gamma)\, x_k \\
&= (1 - 2\gamma)^k \cdot x_1 = \\
&= (1 - 2\gamma)^k
\end{aligned}$$

$$\lim_{k \to \infty} a^k = \begin{cases} 0 & |a| < 1 \\ 1 & a = 1 \\ \text{div } \pm 1 & a = -1 \\ \text{div } \to \pm\infty & \ldots \end{cases}$$



FAV. Question on exam:

why is alternative convergence optimal

PROPERTIES OF FUNCTIONS

$D \subset \mathbb{R}^n$, $f : D \to \mathbb{R}$ cont's, differentiable

CONVEX FUNCTION

Def: D is convex if $\forall x, y \in D$, $\forall t \in [0,1]$
$tx + (1-t)y \in D$

Def: f ~~con~~ is convex if D is convex and
$\forall x, y \in D$, $\forall t \in [0,1]$ INDUCTION

$$f(tx + (1-t)y) \le t f(x) + (1-t) f(y)$$

Proposition 1: Γ the graph of f. TFAE:

(1) f is ~~con~~ convex

(2) $\forall x, y \in D$, the line segment $(x, f(x)) \to (y, f(y))$ lies above Γ.

(3) $\forall x \in D$, the tangent plane to Γ at x lies below Γ.

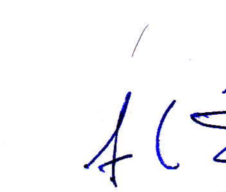
$$f(y) \ge f(x) + \nabla f(x)(y-x)$$

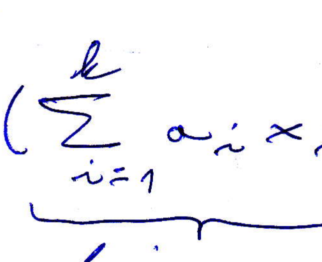
(4) H.W. Each $x_1, x_2, \dots, x_n \in D$, $\forall \alpha_1, \alpha_2, \dots, \alpha_n \in [0,1]$ $\sum_{i=1} \alpha_i = 1$

$$f\Big(\underbrace{\sum_{i=1}^{k} \alpha_i x_i}_{\text{linear combination}}\Big) \le \sum_{i=1}^{k} \alpha_i f(x_i)$$ ← generalization of line, to triangle, ...

(hyperplane - convex hull)

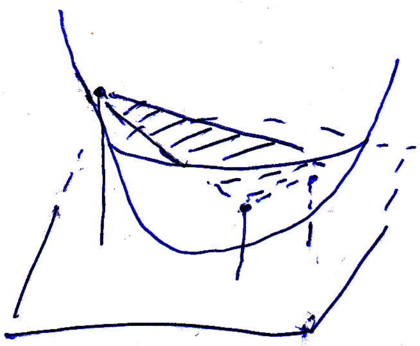
barycentric points





$(\alpha_1, \alpha_2, \alpha_3)$ - barycentric coord.

$\alpha_1 x + \alpha_2 y + \alpha_3 z$ $\alpha_i \in [0,1]$ $\sum \alpha_i = 1$



$\alpha_i \in [0,1]$
$\sum \alpha_i = 1$

extra: If f is twice cont. differentiable, then:

$\boxed{f \text{ is convex} \Leftrightarrow \text{all eigenvalues of } \nabla^2 f \geq 0}$

Proposition 2: Let f be ~~conv~~ convex. Then:

(1) $\forall h \in \mathbb{R}$, $f^{-1}(-\infty, h]$ is convex. collection of all points (1)

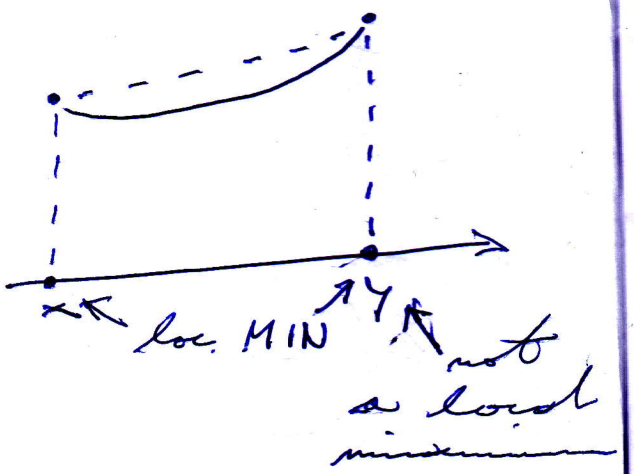
(2) Each local MIN is a global MIN. might not exist

(3) ~~The set~~ {global MINS} is convex.

Proof: (1) $f(x), f(y) \leq h$
$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$
$t h + (1-t) h = h \cdot 1$

[(1) $f^{-1}(A) = \{x;\ f(x) \in A\}$
sublevel set.]

(2) Assume not



(3) use (1) for
$f^{-1}(-\infty, \mathrm{MIN}]$

All partial derivatives of cont. differentiable

Def: f is strictly convex if D is convex and $\forall x, y \in D$, $\forall t \in \cancel{[0,1]}\ (0,1)$:

$$f(tx + (1-t)y) < t \cdot f(x) + (1-t) f(y)$$

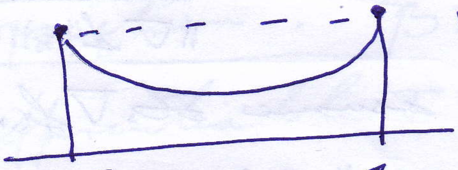
Proposition 1: holds for $<$

Prop 2 holds as is.

Bonus: $|\{\text{global MINS}\}| \leq 1$

proof

if strictly convex



2 global MINS? NO

LIPSCHITZ FUNCTIONS

Def: $L > 0$ f is L-lipschitz if

$$|f(x) - f(y)| \leq L \cdot \|x - y\|, \quad \forall x, y \in D$$

contraction is a map where the points are closer toghether

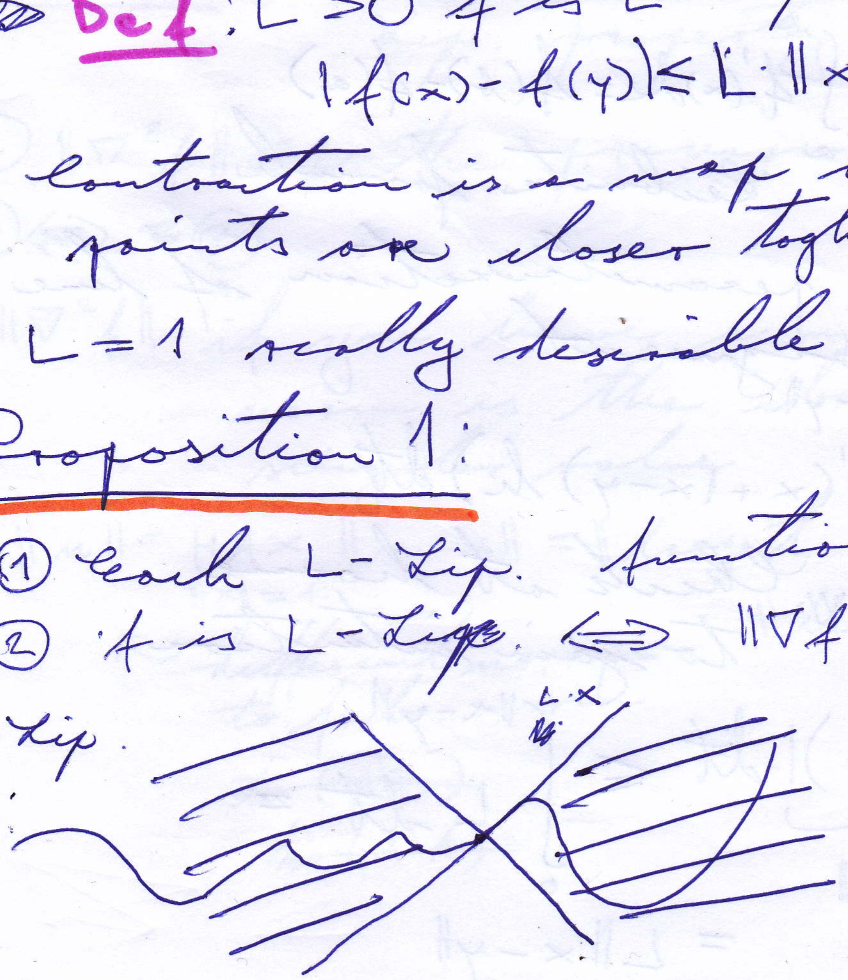
$L = 1$ really desirable

Proposition 1:

(1) Each L-Lip. function is continuous

(2) f is L-Lip. $\iff \|\nabla f\| \leq L$

L-Lip.



Proof: ① $\lim_{x\to y} |f(y)-f(x)|$ should be 0

$$\leq \lim_{x\to y} L\|x-y\| = 0 \checkmark$$

② $|f(x)-f(y)| \leq L\|x-y\|$

$$\Rightarrow \frac{|f(x)-f(y)|}{\|x-y\|} \leq L$$

Direction of gradient: $h = \frac{\nabla f(x)}{\|\nabla f(x)\|}$

assume ~~ts~~ $\nabla f(x) \neq 0$

$$\|\nabla f(x)\| = \lim_{\substack{\cancel{t\to 0} \\ t\searrow 0}} \frac{|f(x+th)-f(x)|}{t} \leq \lim_{t\searrow 0} L\frac{\|th\|}{t}$$

$$|f(x)-f(y)| = \left|\int_0^{\|x-y\|} f'(x+th)\,dt\right| \leq$$

can switch ~~abs~~ because h: unit directional vect.

$$\int_a^b f'(x)dx = f(b)-f(a)$$

Leibnitz formula

parametrization of line segment

$$\int_0^{\|x-y\|} f'(x+(x-y)h)\,dt$$

HW: Check at home, to gain intuition

$$\leq \int_0^{\|x-y\|} \underbrace{|f'(x+th)|}_{\leq\|\nabla f\|}\,dt \leq \int_0^{\|x-y\|} L\,dt =$$

$$= L\|x-y\|$$

DEF: $\beta > 0$ f is β-smooth if $\forall x, y \in D$

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta \|x - y\|$$

gradient is β-lipschitz

<u>Proposition</u>: $\beta > 0$, f twice continuously differentiable TFAE

1. f is β-smooth
2. $\|\nabla^2 f\| \leq \beta$
3. All eigenvalues of $\nabla^2 f$ lie on $[0, \beta]$

lower bound bcs of convexity — upper bound bc. of norm def.

<u>Proof</u>: ① A multivariate version of Lip. property

② $\|\nabla^2 f\| \leq \beta$ the previous proof suffices

③ $\Longleftrightarrow$ ②

$\|\nabla^2 f\|$ if you have a mtx than the norm is the abs. value of the singular value

$$\|M\| = \max_{\|x\|=1} \|M \cdot x\| = \|\text{largest sing. value}\| \overset{M.\ sym.}{=} \|\text{largest eigenvalue}\|^*$$

* different norm to frob. norm

~~$tr(A^T B^T)$~~

~~$tr(A^T A)$~~

Proposition: $\beta > 0$, f convex, β-smooth. Then

error of lin. approx.

$\forall x, y \in D$

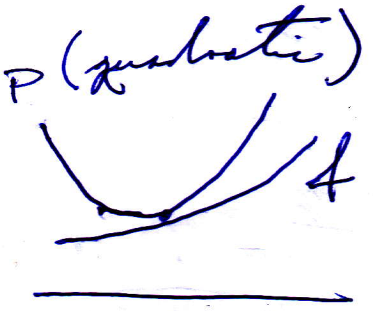
① $\left| f(x) - f(y) - \nabla f(x)^T (y-x) \right| \leq \frac{\beta}{2} \|x-y\|^2$

② $f(x) - f(y) \leq \nabla f(x)^T (x-y) - \frac{1}{2\beta} \|\nabla f(x) - \nabla f(y)\|^2$

Proof:

~~HW~~ ① HW

② TODO



Def: $\alpha > 0$, f is α-strongly convex if $f(x) - \frac{\alpha}{2}\|x\|^2$ is convex

If you have a point, approx. function as that point with hessian

$\Rightarrow$ f has to be bellow

~~Proporties~~

Proposition: $\alpha > 0$

① f is α-strongly convex iff $\forall x, y \in D$:

$$f(y) \geq f(x) + \nabla f(x)^T (y-x) + \frac{\alpha}{2}\|x-y\|^2$$

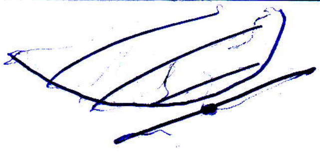
② A twice cont. differentiable function f is α-strongly convex iff eigvalues of $\nabla^2 f$ are $\geq \alpha$.

proof of ②: $\nabla^2\left(f(x) - \frac{\alpha}{2}\|x\|^2\right) = \nabla^2 f - \alpha \cdot I$

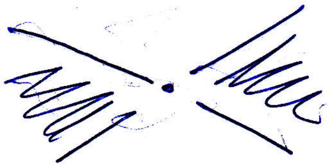
$f - \frac{\alpha}{2}\|x\|^2$ is convex $\Leftrightarrow$ eigenvalues of $\nabla^2 f - \alpha I$ are ≥ 0

$\Leftrightarrow$ eigvalues of $\nabla^2 f$ are $\geq \alpha$

SUMMARY

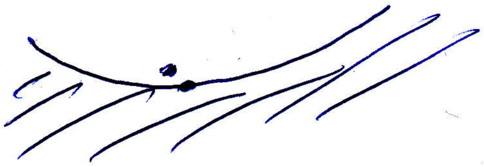


- convex eigenvalues of $\nabla^2 f \geq 0$



- L-lipsch $\|\nabla f\| \leq L$
- β-smooth eigenvalues of $\nabla^2 f \leq \beta$
- α strongly convex eigenvalues of $\nabla^2 f \geq \alpha$

combination
that can't
happen simultaneously



α-strongly convex
opposite