Convex Optimization: Lecture 4.

Plan for today: Gradient Descent - see how it performs as a function of different properties of f.

min: $f(x)$
st: $x \in \mathbb{R}^n$ or $\mathcal{X}$, $\quad \mathcal{X} \subseteq \mathbb{R}^n$ convex set.

Basic Algorithm: $\quad x^+ = x - \eta \nabla f(x)$

1. If $\nabla f(x)$ not defined, will replace $\nabla f(x)$ by $g_x \in \partial f(x)$

2. If $\mathcal{X} \subsetneq \mathbb{R}^n$ - in this case x^+ may not even be in $\mathcal{X}$.

Intuition from quadratic example:

$f(x) = ax^2 + \cancel{bx} + c \qquad$ need $a \geq 0$

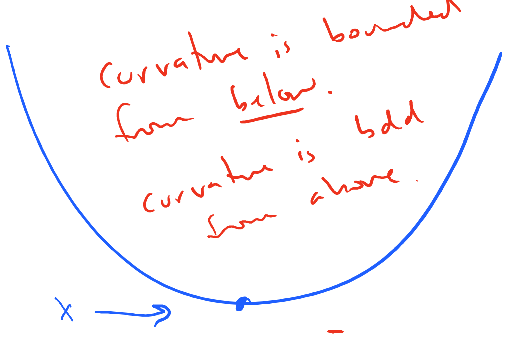
$x^+ = x - \eta \cdot 2ax = (x - 2\eta a)x$

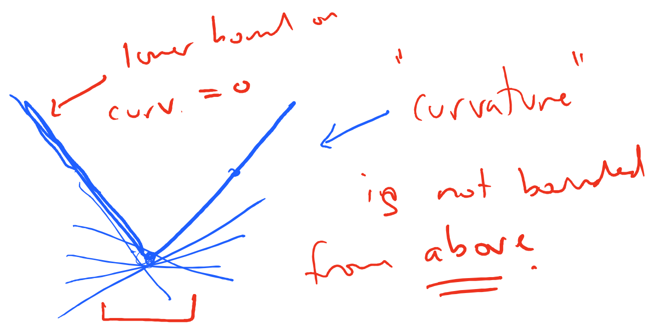
Need: $|1 - 2\eta a| < 1 \implies \eta < 1/a$ (recall $a > 0$)

Graphical intuition: step size less than max curvature.

$f(x) = |x| \qquad$ no curvature $\implies$ step size had to be $\sim \varepsilon$, because $\nabla f(x)$ does not nec.

go to zero as $x \to x^*$.





Def'n: Convexity

f is convex if $\forall g \in \partial f(x)$,

$$f(y) \geq f(x) + \langle g, y - x \rangle \quad \forall x, y$$

If f is diff'ble at x, $\partial f(x) = \{\nabla f(x)\}$

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

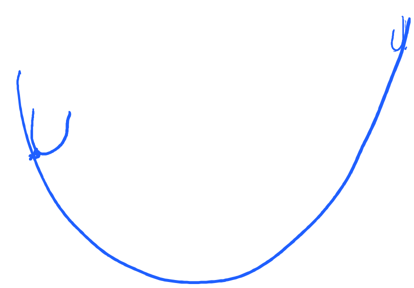
Def'n: ("Curvature Bdd from above")

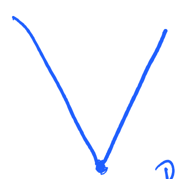
β-Smooth

A convex fn is called β-smooth if its gradient is β-Lipschitz.

Def'n: A function $h: \mathbb{R}^n \to \mathbb{R}$ is called L-Lipschitz if $\forall x, y$, $\left(\|x\|_2 = \left(\sum x_i^2\right)^{1/2}\right)$

$$|h(x) - h(y)| \leq L \|x - y\|_2 .$$





Exercise:

Prove that $\exists$ quadratic going through 0, that lies above $|x|$.

lies above $|x|$.

f is β-smooth $\iff \|\nabla f(x) - \nabla f(y)\|_2 \leq \beta \|x-y\|_2$

Def'n: A convex fn $f : \mathbb{R}^n \to \mathbb{R}$ is called α-strongly convex at x if $\left(f(x) - \frac{\alpha}{2}\|x\|_2^2\right)$ is convex.

f is α-SC, if it is α-SC at x $\forall x$.

$f_1 + f_2$

$\nabla f_1 + \nabla f_2$

Suppose: $\underbrace{\left(f(x) - \frac{\alpha}{2}\|x\|_2^2\right)}_{h(x)}$ is convex.

$$h(y) \geq h(x) + \langle g_x^h, y-x \rangle \quad \forall x, y$$

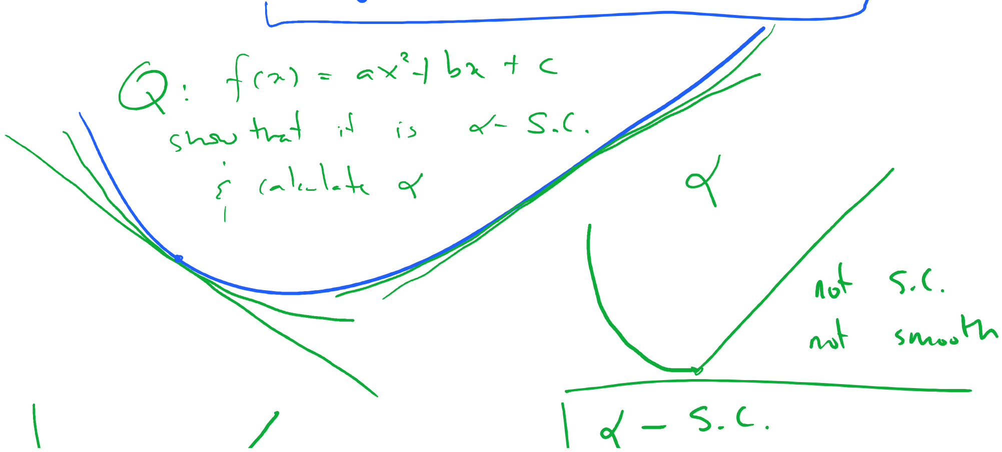
plugging in: $f(y) - \frac{\alpha}{2}\|y\|_2^2 \geq f(x) - \frac{\alpha}{2}\|x\|_2^2 + \langle g_x^f - \alpha x, y-x \rangle$

$$f(y) \geq f(x) + \langle g_x^f, y-x \rangle + \left(\frac{\alpha}{2}\|y\|_2^2 + \frac{\alpha}{2}\|x\|_2^2 - \alpha\langle x, y\rangle\right)$$

$$\underbrace{f(y) \geq f(x) + \langle g_x^f, y-x \rangle}_{\text{convexity}} + \frac{\alpha}{2}\|x-y\|_2^2$$

Q: $f(x) = ax^2 + bx + c$

show that it is α-S.C. & calculate α



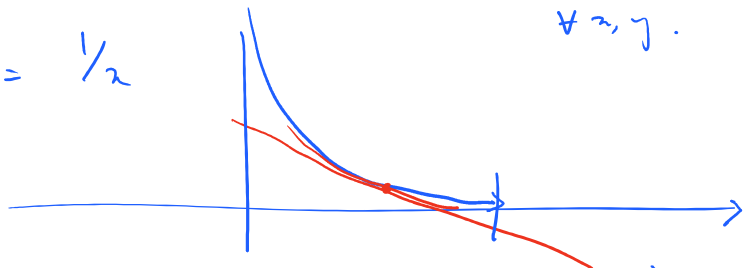
α – s.c.

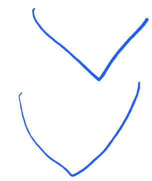
β – smooth

$f : \mathbb{R}^n \to \mathbb{R}$ is called

Def'n: Strictly convex if: $f(y) > f(x) + \langle g_x, y - x \rangle \quad \forall x, y.$

Ex: $f(x) = 1/x$





Back to β-smoothness $\left(\|\nabla f(x) - \nabla f(y)\|_2 \le \beta \|x - y\|_2 \right)$

Lemma: If f is β-smooth, then $\forall x, y \in \mathbb{R}^n$

$$\left| f(x) - f(y) - \langle \nabla f(y), x - y \rangle \right| \le \frac{\beta}{2} \|x - y\|_2^2$$

$$\left| f(x) - f(y) - \langle \nabla f(y), x - y \rangle \right| = \left| \int_0^1 \langle \nabla f(y - t(x-y)), x - y \rangle \, dt - \langle \nabla f(y), x - y \rangle \right|$$

?

$$\left[\int_0^1 (y - t(x-y)), x - y \rangle \, dt \right.$$
$$\left. \text{(Exercise: FTC)} \right]$$

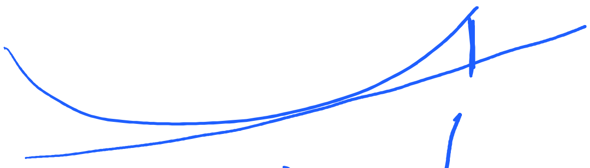
Re-writing:

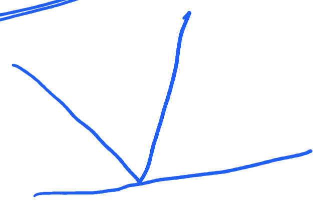
$$\left| f(x) - f(y) - \langle \nabla f(y), x - y \rangle \right| = \left| \int_0^1 \langle \nabla f(y + t(x-y)), (x-y) \rangle \, dt - \langle \nabla f(y), x - y \rangle \right|$$

$$\le \int_0^1 \left| \qquad \right|$$

$$|\langle v, w \rangle| \le \|v\| \cdot \|w\| \quad \text{Cauchy-Schwarz}$$

$$\le \int_0^1 \|\nabla f(y + t(x-y)) - \nabla f(y)\| \cdot \|x - y\| \, dt$$





$$\le \int_0^1 \|\nabla f(y + t(x-y)) - \nabla f(y)\| \cdot \|x-y\| \, dt$$

$$\le \int_0^1 \beta \|t(x-y)\| \cdot \|x-y\| \, dt$$

$$= \beta \|x-y\|_2^2 \int_0^1 t \, dt = \underline{\frac{\beta}{2} \|x-y\|_2^2}$$

If f is convex and β-smooth

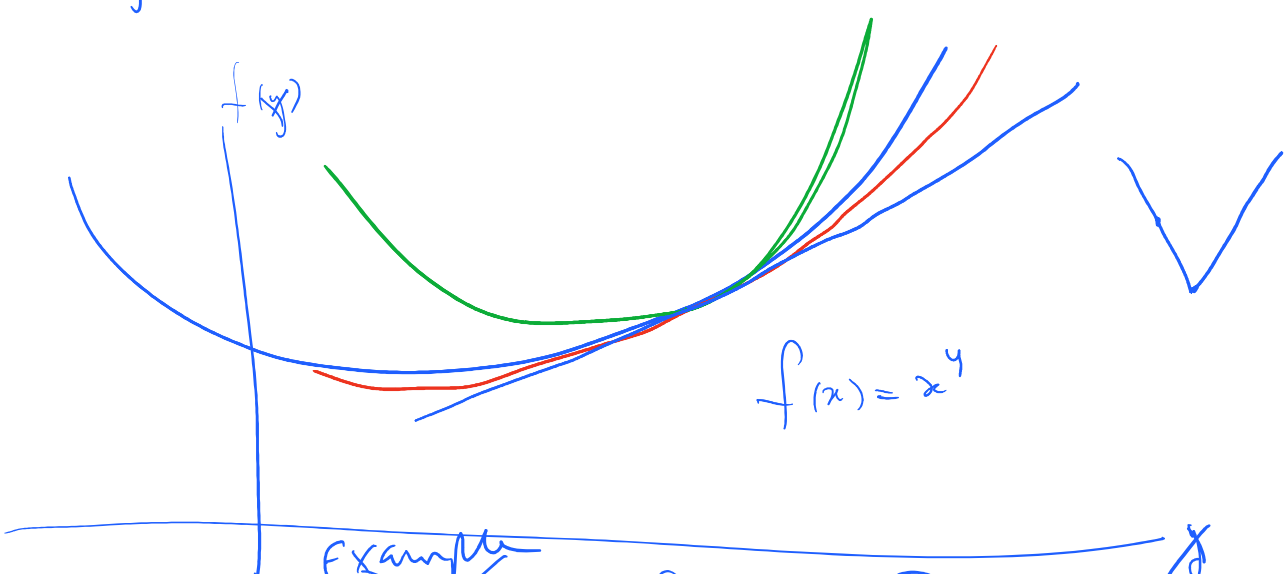
$$0 \le \underbrace{f(y) - \underbrace{f(x) - \langle \nabla f(x), y-x \rangle}_{} \le \frac{\beta}{2} \|x-y\|_2^2}_{}$$

f convex: $f(y) \ge f(x) + \langle \nabla f(x), y-x \rangle$

$+$

β-smooth: $f(y) \le \underline{f(x) + \langle \nabla f(x), y-x \rangle + \frac{\beta}{2} \|x-y\|_2^2}$

α-SC: $f(y) \ge f(x) + \langle \nabla f(x), y-x \rangle + \frac{\alpha}{2} \|x-y\|_2^2$



Example ~~Exercise~~: $f : \mathbb{R} \to \mathbb{R}$

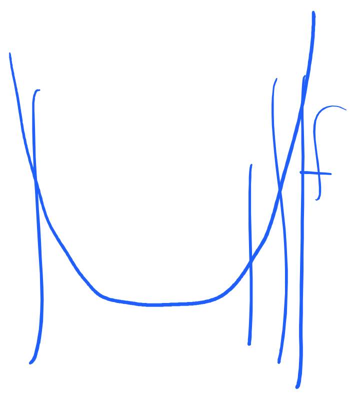
$f(x) = x^4$ is neither smooth nor strong convex i.e. $\exists\ \beta < \infty$ or $\alpha > 0$ s.t. rel. def'ns are satisfied.

$f : [a,b] \longrightarrow \mathbb{R}$, $0 \in [a,b]$

$x \longmapsto x^4$

β-smooth

not s.c.



$f : [a,b] \longrightarrow \mathbb{R}$

$x \longmapsto x^4$ $\qquad 0 \notin [a,b]$

s.c. and smooth.

If f has 1^{st} & 2^{nd} derivatives,

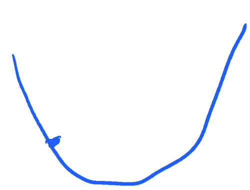
or in m.v. case, $\nabla^2 f$ defined,

what are ~~is~~ s.c. & smoothness parameters at

at pt x. ?

<u>Answer</u>: max & min e-values of $\nabla^2 f(x)$.

Back to $f(x) = x^4$. $\qquad f''(x) = 12x^2$

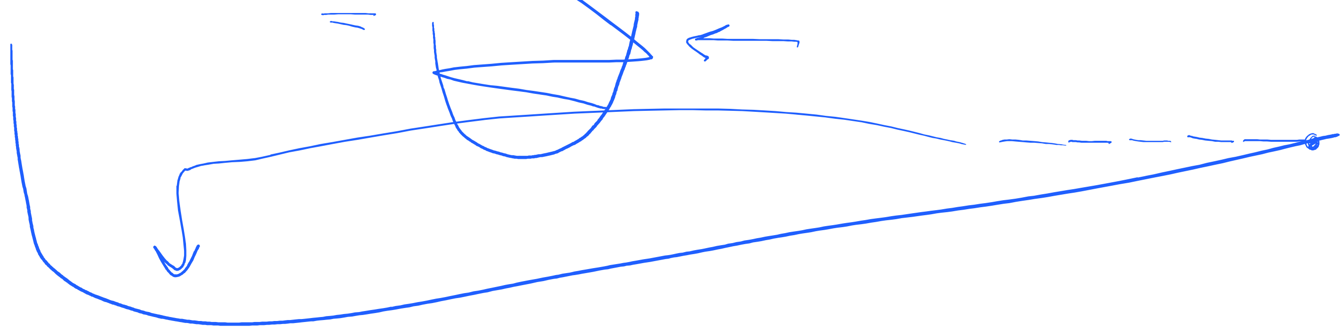


$\beta \triangleq \sup \lambda_{max}(x)$

$\alpha \triangleq \inf \lambda_{min}(x)$

Def'n: Condition #

$\beta/\alpha \triangleq$ condition number (κ)



Next time: Rates of Conv. for GD dep. on α, β.