$\boxed{\text{Starting 9/16}}$ (ii) $\Rightarrow$ (iii)

$$f(y) \geq f(x) + \nabla f(x)^T (y-x)$$
$$+ \quad f(x) \geq f(y) + \nabla f(y)^T (x-y)$$

$$0 \geq \nabla f(x)^T (y-x) + \nabla f(y)^T (x-y)$$
$$= -(\nabla f(x) - \nabla f(y))^T (x-y). \quad \checkmark$$

(iii) $\Rightarrow$ (ii) Consider $x, y \in \mathbb{R}^d$ and $\phi(t) = f(x + t(y-x))$

Then $f(y) = \phi(1) = \phi(0) + \int_0^1 \phi'(t)\,dt$

$$= \phi(0) + \phi'(0) + \int_0^1 (\phi'(t) - \phi'(0))\,dt$$

$$= f(x) + \nabla f(x)^T (y-x)$$
$$+ \int_0^1 (\nabla f(x + t(y-x)) - \nabla f(x))^T \, t(y-x)/t \; dt$$

$$\geq f(x) + \nabla f(x)^T (y-x). \quad \checkmark$$

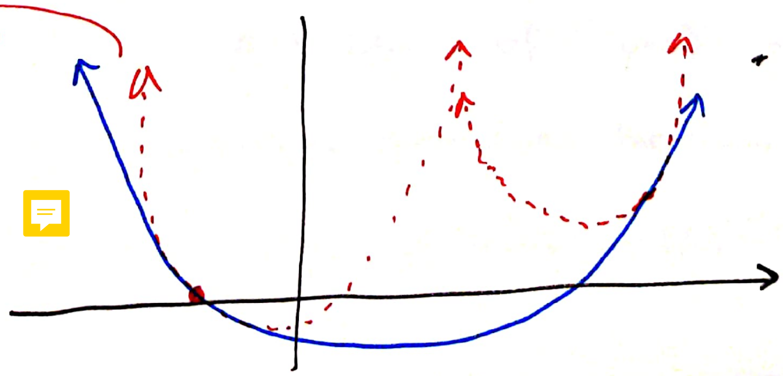
$\square$

Lemma (Second-Characterization of Convexity)

For twice diff f, f is convex if and only if

(iv) $\nabla^2 f(x) \succeq 0 \quad \forall x$

May not be on upper bound. $f(x) = x^4$.



* 2^{nd} order model is flat or curving up.

Proof Sketch in 1D

$$\text{(iv)} \iff s^T \nabla f(x)\, s \geq 0$$

$$\iff \phi''(t) \geq 0$$

$$\iff \phi'(t) \text{ is monotone increasing}$$

(non decreasing)

$$\iff \text{(iii)}$$

$\square$

Lemma (Equivalent Conditions for Smoothness)

Suppose f is convex, then the following are equivalent

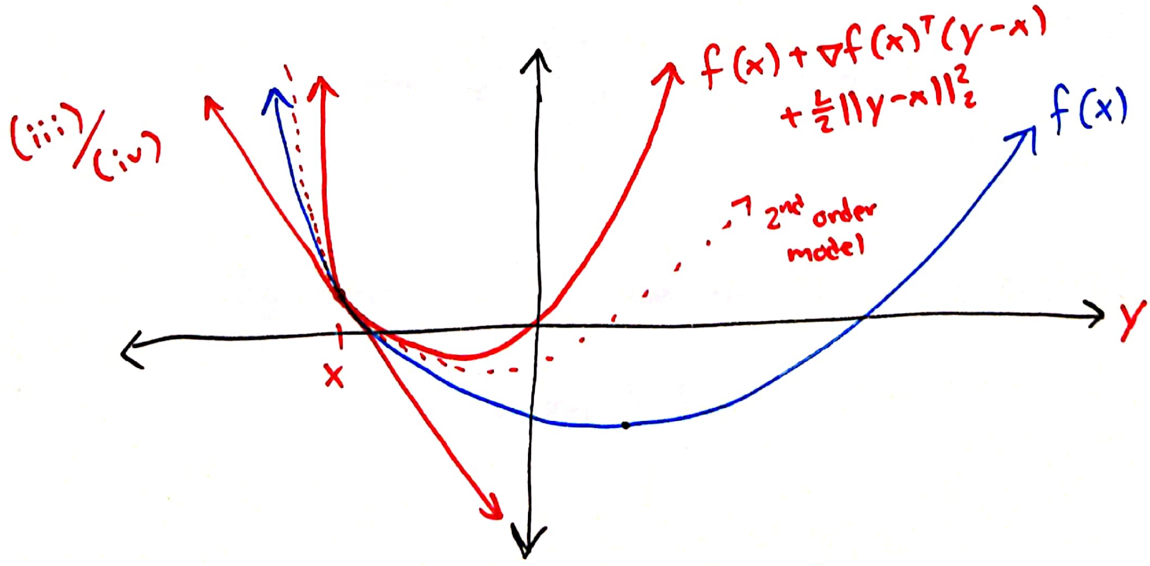
(i) f has L-Lipschitz gradient (L-smooth)

(ii) $\frac{L}{2}\|x\|_2^2 - f(x)$ is convex.

(iii) $f(y) \leq f(x) + \nabla f(x)^T(y-x) + \frac{L}{2}\|y-x\|^2 \quad \forall x,y$

(if f is C^2) (iv) $\nabla^2 f(x) \preceq L \cdot I \quad \forall x \quad (L\cdot I - \nabla^2 f(x) \succeq 0)$

(v) $(\nabla f(y) - \nabla f(x))^T(y-x) \geq \frac{1}{L}\|\nabla f(y) - \nabla f(x)\|_2^2 \quad \forall x,y$



$f(x) + \nabla f(x)^T(y-x)$

Proof. $\boxed{(ii) \Leftrightarrow (iii) \Leftrightarrow (iv)}$ by previous convexity results

$\boxed{(v) \Rightarrow (i)}$ By Cauchy-Schwarz,

Dividing through by $\frac{\|\nabla f(y) - \nabla f(x)\|}{L}$

$$\|\nabla f(y) - \nabla f(x)\|_2 \|y-x\|_2 \geq \frac{1}{L}\|\nabla f(y) - \nabla f(x)\|_2^2$$

$$L\|y-x\|_2 \geq \|\nabla f(y) - \nabla f(x)\|_2 . \checkmark$$

$\boxed{(i) \Rightarrow (iii)}$ By Taylor Approx Theorem.

$\boxed{(iii) \Rightarrow (v)}$ Consider any $x, y \in \mathbb{R}^d$, $z = y - \frac{1}{L}(\nabla f(y) - \nabla f(x))$.

$$\begin{aligned}
f(x) - f(y) &= f(x) - f(z) + f(z) - f(y) \\
&\leq \nabla f(x)^T(x-z) + \nabla f(y)^T(z-y) + \frac{L}{2}\|z-y\|^2 \\
&= \nabla f(x)^T(x-y) + (\nabla f(y) - \nabla f(x))^T(z-y) \\
&\qquad + \frac{L}{2}\|z-y\|^2 \\
&= \nabla f(x)^T(x-y) - \frac{1}{L}\|\nabla f(y) - \nabla f(x)\|_2^2 \\
&\qquad + \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2 \\
&= \nabla f(x)^T(x-y) - \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|_2^2 .
\end{aligned}$$

Swapping x, y:

$$f(y) - f(x) \leq \nabla f(y)^T(y-x) - \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_2^2 .$$

Summing those gives the claim:

$$0 \leq (\nabla f(y) - \nabla f(x))^T(y-x) - \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|_2^2$$

$\checkmark$ $\square$

5. Better Guarantees for Smooth Convex OPT

Considering running gradient descent

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

with reasonable stepsize we had

$$f(x_{k+1}) \le f(x_k) - \frac{1}{4L} \|\nabla f(x_k)\|_2^2 .$$

<u>Theorem</u> Let f be convex with L-Lipschitz grad. Then letting $x^* \in \arg\min f$, we have

$$f(x_{k+1}) - f(x^*) \le \frac{2L \|x_0 - x^*\|^2}{k}, \quad \alpha_k = 1/L .$$

After $k = \frac{2L\|x_0 - x^*\|^2}{\varepsilon}$ steps, $f(x_k) - f(x^*) \le \varepsilon$.

Recall previously $\frac{1}{T}\sum_{i=0}^{T-1} \|\nabla f(x_i)\|_2^2 \le \frac{L(f(x_0) - f(x^*))}{T}$

$$\Rightarrow \frac{1}{T}\sum_{i=k}^{T+k-1} \|\nabla f(x_i)\|_2^2 \le \frac{L\cdot(f(x_k) - f(x^*))}{T}$$
$$\le \frac{L\cdot\varepsilon}{T}$$

By step $T = 1/\varepsilon$, we have some $\|\nabla f(x_i)\|_2^2 \le L\varepsilon^2$.

(square better than before).

Proof. First prove distance to optimal does not grow:

$$\|x_{k+1}-x^*\|_2^2 = \|x_k - \tfrac{1}{L}\nabla f(x_k) - x^*\|_2^2$$
$$= \|x_k - x^*\|_2^2 + \tfrac{1}{L^2}\|\nabla f(x_k)\|_2^2 - \tfrac{2}{L}\nabla f(x_k)^T(x_k - x^*)$$

using (v) from smoothness at x_k, x^*

$$\le \|x_k - x^*\|_2^2 + \tfrac{1}{L^2}\|\nabla f(x_k)\|_2^2 - \tfrac{2}{L^2}\|\nabla f(x_k)\|_2^2$$
$$= \|x_k - x^*\|_2^2 - \tfrac{1}{L^2}\|\nabla f(x_k)\|_2^2$$
$$\le \|x_k - x^*\|_2^2.$$

Let $\delta_k = f(x_k) - f(x^*)$ be our k^{th} objective gap.

From our descent lemma,

$$\delta_{k+1} \le \delta_k - \tfrac{1}{2L}\|\nabla f(x_k)\|_2^2$$

By convexity

$$\delta_k \le \nabla f(x_k)^T(x_k - x^*) \le \|\nabla f(x_k)\|\|x_k - x^*\|$$
$$\le \|\nabla f(x_k)\|\,\|x_0 - x^*\|$$

$$\Rightarrow \|\nabla f(x_k)\| \ge \delta_k / \|x_0 - x^*\|.$$

$$\Rightarrow \delta_{k+1} \le \delta_k - \frac{\delta_k^2}{2L\|x_0-x^*\|^2} \le \delta_k - \frac{\delta_k\delta_{k+1}}{2L\|x_0-x^*\|^2}$$

Divide by $\delta_k\delta_{k+1}$

$$\Rightarrow \frac{1}{\delta_k} \le \frac{1}{\delta_{k+1}} - \frac{1}{2L\|x_0-x^*\|^2}$$

$$\Rightarrow \frac{1}{\delta_{k+1}} \ge \frac{1}{\delta_k} + \frac{1}{2L\|x_0-x^*\|^2} \ge \frac{1}{\delta_0} + \frac{k}{2L\|x_0-x^*\|^2}$$
$$\ge \frac{k}{2L\|x_0-x^*\|^2}. \quad \square$$

We expect a speed up under "nice curvature condition".

Definition We say f is μ-strongly convex if

$$f(x) - \frac{\mu}{2}\|x\|_2^2 \text{ is convex.}$$

Lemma For cont diff f, the following are equivalent

(i) f is μ-strongly convex

(ii) $f(y) \geq f(x) + \nabla f(x)^T(y-x) + \frac{\mu}{2}\|y-x\|_2^2 \quad \forall x, y$

(if f is C^2) (iii) $\nabla^2 f(x) \succeq \mu I \quad \forall x$

(iv) $(\nabla f(y) - \nabla f(x))^T(y-x) \geq \mu\|y-x\|_2^2 \quad \forall x, y$

Proof. These are exactly the conditions for $f(x) - \frac{\mu}{2}\|x\|_2^2$ to be convex.

$\square$