Lemma The Armijo Condition holds for

$$\alpha \in \left[0, \frac{2(1-\eta)}{L}\right].$$

Starting 9/14

Proof. Our descent lemma ensures

$$f(\underset{\substack{\uparrow \\ x_k - \alpha \nabla f(x_k)}}{\cancel{\;\;\;}}) \le f(x_k) - \left(\alpha - \frac{L\alpha^2}{2}\right) \|\nabla f(x_k)\|_2^2 \quad \forall \alpha$$

We want $\alpha - \frac{L\alpha^2}{2} \ge \eta\alpha \iff$ Armijo Condition

$$\iff (1-\eta)\alpha \ge \frac{L\alpha^2}{2}$$

$$\iff \alpha \in \left[0, \frac{2(1-\eta)}{L}\right]. \qquad \square$$

Then our backtracking will have

$$\alpha_k \ge \min\left(\alpha, \frac{2\tau(1-\eta)}{L}\right).$$

$\Rightarrow$ GD with backtracking linesearch has

$$f(x_{k+1}) \le f(x_k) - \eta\alpha_k \|\nabla f(x_k)\|_2^2$$

$$\le f(x_k) - \min\left(\eta\alpha, \frac{2\tau\eta(1-\eta)}{L}\right) \|\nabla f(x_k)\|_2^2.$$

If $\alpha \ge 1/L$, $\eta = \tau = 1/2$, then

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

3. Nonconvex Smooth Opt Guarantees

Consider solving $\min_{x \in \mathbb{R}^d} f(x)$ with L-Lips ∇f by

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

for $x_0 \in \mathbb{R}^d$.

<u>Theorem</u> Suppose f is cont diff with L-Lips gradient,

Then for $T \geq 0$,

$$\frac{1}{T} \sum_{k=0}^{T-1} \|\nabla f(x_k)\|_2^2 \leq \frac{2L(f(x_0) - \min f(x))}{T}$$

when $\alpha_k = 1/L$ or with exact linesearch, and

$$\frac{1}{T} \sum_{k=0}^{T-1} \|\nabla f(x_k)\|_2^2 \leq \max\left(\frac{1}{\eta \alpha}, \frac{L}{2\tau\eta(1-\eta)}\right) \frac{f(x_0) - \min f(x)}{T}$$

when we use Armijo backtracking.

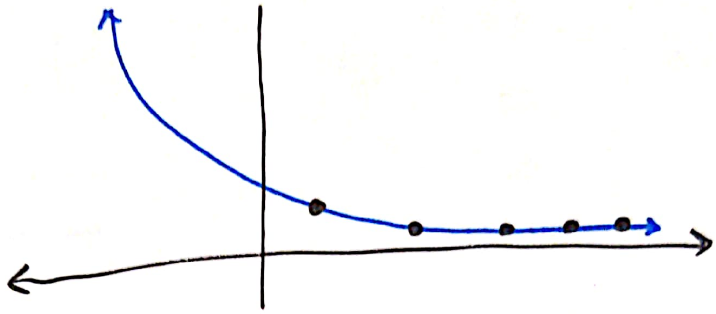
Picking $T = \overset{\text{sublinear}}{O(1/\varepsilon)}$, we have

$$\|\nabla f(x_k)\|_2^2 \leq \varepsilon \quad \text{for some } k \leq T$$

(Nearly first-order optimal).

Note: x_k may not be converging.

For example, $f(x) = e^{-x}$



if x_k does converge, it may not be local min

(our condition is necessary but not suff).

Proof. With either $\alpha_k = 1/L$ or exact linesearch

$$f(x_{k+1}) - \min_x f(x) \leq f(x_k) - \min_x f(x) - \frac{1}{2L} \|\nabla f(x_k)\|_2^2$$

Summing this up for $k = 0, \ldots, T-1$

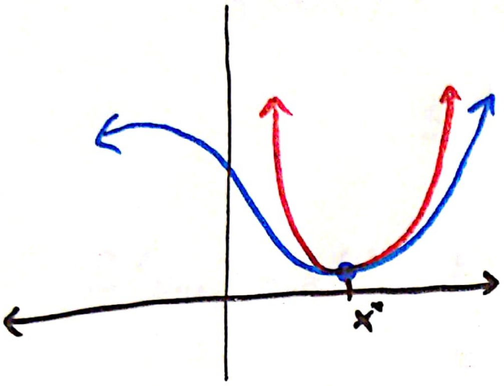
$$\Rightarrow \underbrace{f(x_T) - \min f(x)}_{0 \leq} \leq f(x_0) - \min f(x) - \frac{1}{2L} \sum_{k=0}^{T-1} \|\nabla f(x_k)\|_2^2$$

$$\Rightarrow \sum_{k=0}^{T-1} \|\nabla f(x_k)\|_2^2 \leq 2L\,(f(x_0) - \min f(x)).$$

Similar proof for backtracking. $\square$

Suppose $x_k \to x^*$ satisfying our sufficient 2^{nd}-order conditions

x^* is a strict local min since
$$\nabla f(x^*) = 0$$
$$\nabla^2 f(x^*) \succ 0.$$



$\Rightarrow$ For $\varepsilon > 0$, $\lambda > 0$, all $x \in B(x^*, \varepsilon)$ has
$$\lambda_{\min}(\nabla^2 f(x)) > \lambda$$

(assuming f is C^2 near x^*)

Then $\phi(t) = f(x^* + ts)$ for some $\|s\|_2 \le \varepsilon$
$$\phi'(1) = \phi'(0) + \int_0^1 \phi''(t)\,dt$$

Not correct

$s^T \nabla^2 f(x+ts) s$
$\le \|s\| \, \|\nabla^2 f(x+ts) s\|$
$\le \lambda \|s\| \, \|s\|$

Using that for any symmetric matrix
$s^T \nabla^2 f(x^*+ts) s$
(1) $\ge \lambda_{\min}(\nabla^2 f) \cdot \|s\|_2^2$

$$\Rightarrow \nabla f(x^* + s)^T s \ge 0^T s + \lambda \|s\|_2^2$$
$$\Rightarrow \|\nabla f(x^*+s)\|_2 \|s\|_2 \ge \lambda \|s\|_2^2$$
$$\Rightarrow \|\nabla f(x^*+s)\|_2 \ge \lambda \|s\|_2$$

Taylor Approximation Theorem:
$$\frac{L}{2}\|s\|_2^2 \ge f(x^*+s) - (f(x^*) + 0^T s)$$
$$= f(x^*+s) - f(x^*) \qquad (2)$$

$(1)^2 + \frac{2}{L}(2)$
$$\Rightarrow \|\nabla f(x^*+s)\|_2^2 \ge \frac{2\lambda^2}{L}(f(x^*+s) - f(x^*)).$$

Then for k large enough $x_k \in B(x^*, \varepsilon)$.

$$f(x_{k+1}) \le f(x_k) - \frac{1}{2L} \|\nabla f(x_k)\|_2^2$$
$$\le f(x_k) - \frac{\lambda^2}{L^2} \left(f(x_k) - f(x^*) \right)$$

$$f(x_{k+1}) - f(x^*) \le f(x_k) - f(x^*) - \frac{\lambda^2}{L^2} \left(f(x_k) - f(x^*) \right)$$
$$= \left(1 - \frac{\lambda^2}{L^2}\right) \left(f(x_k) - f(x^*) \right)$$

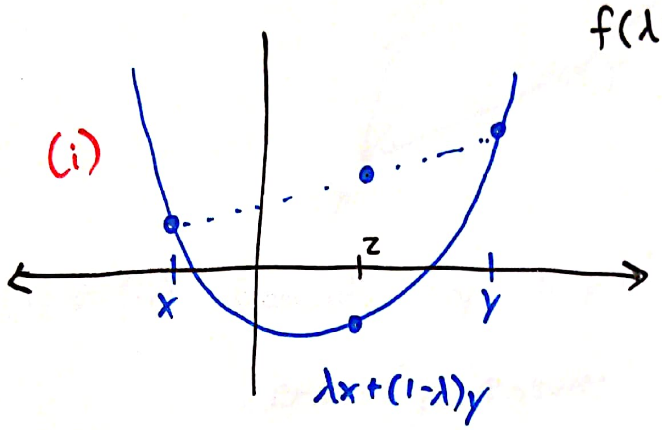
$$\Rightarrow f(x_{k+T}) - f(x^*) \le \left(1 - \frac{\lambda^2}{L^2}\right)^T \left(f(x_k) - f(x^*) \right)$$

$$\Rightarrow T = O\left(\left(\frac{\lambda^2}{L^2} \right)^{-1} \log\left(\frac{f(x_k) - f(x^*)}{\varepsilon} \right) \right)$$

then $f(x_k) - f(x^*) \le \varepsilon$.

4. Shape of Smooth, Convex Functions

Recall f is convex if



$$f(\lambda x+(1-\lambda)y) \le \lambda f(x)+(1-\lambda)f(y) \quad \forall x,y \in \mathbb{R}^d,\ \lambda\in[0,1].$$

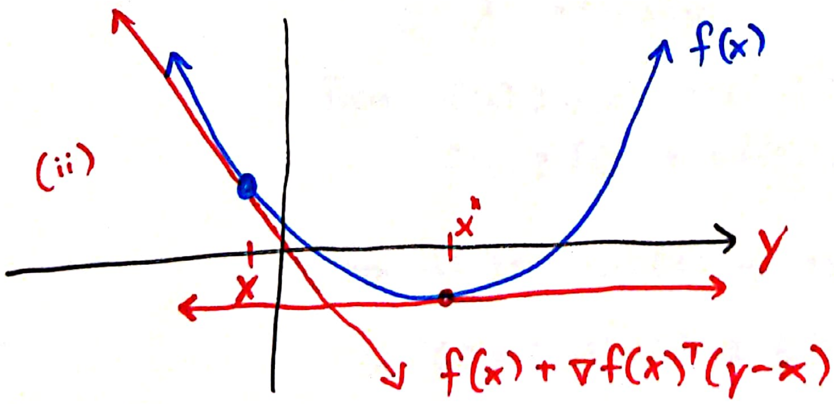
Lemma (First-Order Characterization)

For cont diff f, the following are equivalent

(i) f is convex.

(ii) $f(y) \ge f(x) + \nabla f(x)^T(y-x) \quad \forall x,y$

monotone $\nabla f(x)$ → (iii) $\langle \nabla f(x) - \nabla f(y) \rangle^T (x-y) \ge 0 \quad \forall x,y$

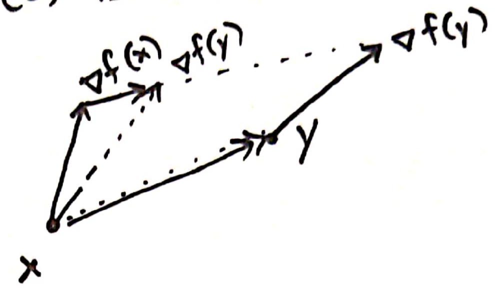


(iii) In 1D, $(\phi'(0) - \phi'(1))(0-1) \ge 0$

$(\phi'(t) - \phi'(t'))(t-t') \ge 0$

If t increases, $\phi'(t)$ increases
(doesn't decrease)

$\phi'(t)$ is monotone increasing.



Proof. (i) ⇒ (ii) Consider any $x, y \in \mathbb{R}^d$ and $\lambda \in (0,1]$

Convexity ensures

$$f(x+\lambda(y-x)) \le (1-\lambda)f(x) + \lambda f(y)$$

$$\Leftrightarrow f(y) - f(x) \ge \frac{f(x+\lambda(y-x)) - f(x)}{\lambda}$$

As $\lambda \to 0$

$$f(y) - f(x) \ge \nabla f(x)^T(y-x) . \Leftrightarrow (ii)$$

(ii) ⇒ (i) Consider any $x, y \in \mathbb{R}$. $\lambda \in [0,1]$

$z = \lambda x + (1-\lambda)y$.

Then $f(x) \ge f(z) + \nabla f(z)^T(x-z)$ (1)

$f(y) \ge f(z) + \nabla f(z)^T(y-z)$ (2)

Then $\lambda(1) + (1-\lambda)(2)$ gives

$$\lambda f(x) + (1-\lambda)f(y) \ge f(z) + \nabla f(z)^T(\underbrace{\lambda x + (1-\lambda)y - z}_{=0})$$

$= f(z)$. $\Leftrightarrow$ (i)