Lecture-14

Ref: Chapter 3 of [Wright-Recht]

Problem: $\min_{x \in \mathbb{R}^d} f(x)$ → smooth

Model: $x \longrightarrow \boxed{\text{Exact Gradient Oracle}} \longrightarrow \nabla f$ → no noise

If $f(x+td) < f(x)$ for sufficiently small $t > 0$, then "d" is a descent direction.

Claim: If f is continuously diff'ble in a neighborhood of x, then any d satisfying $d^T \nabla f(x) < 0$ is a descent direction

Why? f is smooth ⇒ from Taylor's theorem,

$$f(x+td) = f(x) + t\nabla f(x+\beta t d)^T d, \quad \beta \in (0,1) \quad (*)$$

$\nabla f(x)^T d < 0$ & ∇f continuous

why? ⇒ $\nabla f(x+td)^T d < 0$ for small t.

Using this fact in $(*)$ ⇒

$$f(x+td) < f(x)$$

∎

Remark: $$\min_{\|d\|=1} d^T \nabla f(x) = \frac{\nabla f(x)^T \nabla f(x)}{\|\nabla f(x)\|} = \nabla f(x)$$

steepest descent

GD / steepest descent:

$$\boxed{x_{k+1} = x_k - \alpha \nabla f(x_k),}$$

f is L-smooth $\Rightarrow$

$$f(x+\alpha d) \le f(x) + \alpha \nabla f(x)^T d + \alpha^2 \frac{L}{2}\|d\|^2 \quad -(*)$$

Using $(*)$ & setting $\alpha = 1/L$, $\rightarrow$ Smoothness parameter

$$f(x_{k+1}) = f\left(x_k - \frac{1}{L}\nabla f(x_k)\right)$$

$$\overset{\text{why? }(*)}{\le} f(x_k) - \frac{1}{L}\|\nabla f(x_k)\|^2 + \frac{1}{L^2} \times \frac{L}{2}\|\nabla f(x_k)\|^2$$

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

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

GD- Bound for Smooth f (not necessarily convex)

Assume: $\underbrace{f(x) \ge \bar{f}}_{\hookrightarrow \text{ f is lower-bdd.}} \quad \forall x$

Using 1,

$$f(x_n) \le f(x_{n-1}) - \frac{1}{2L}\|\nabla f(x_{n-1})\|^2$$

$$\leq f(x_{n-2}) - \frac{1}{2L}\|\nabla f(x_{n-1})\|^2 - \frac{1}{2L}\|\nabla f(x_{n-2})\|^2$$

$$\vdots$$

$$f(x_n) \leq f(x_0) - \frac{1}{2L}\sum_{k=0}^{n-1}\|\nabla f(x_k)\|^2$$

$$\frac{1}{2L}\sum_{k=0}^{n-1}\|\nabla f(x_k)\|^2 \leq f(x_0) - f(x_n) \leq f(x_0) - \bar{f}$$

$$\Rightarrow \sum_{k=0}^{n-1}\|\nabla f(x_k)\|^2 \leq 2L\left(f(x_0) - \bar{f}\right) \quad \text{— } (**)$$

$$\Rightarrow \lim_{n\to\infty}\sum_{k=0}^{n-1}\|\nabla f(x_k)\|^2 < \infty$$

(or) $$\lim_{n\to\infty}\|\nabla f(x_n)\| = 0$$ $\rightarrow$ Convergence to a Stationary Point

Also, min $\leq$ avg $\Rightarrow$

using $(**)$

$$\min_{k=0,1,\dots,n-1}\|\nabla f(x_k)\|^2 \leq \frac{1}{n}\sum_{k=0}^{n-1}\|\nabla f(x_k)\|^2 \leq \frac{2L\left(f(x_0)-\bar{f}\right)}{n}$$

(or) $$\boxed{\min_{k=0,1,\dots,n-1}\|\nabla f(x_k)\| \leq \sqrt{\frac{2L\left(f(x_0)-\bar{f}\right)}{n}}}$$

"Stationary Convergence"

After n steps, we have a point with $\|\text{gradient}\| \le \frac{\text{const}}{\sqrt{n}}$

$$x_R = \begin{cases} x_i & \text{w.p. } \frac{1}{n}, \ \forall i \end{cases}$$

x_k picked unif. at random $\rightarrow \{x_0, \dots, x_{n-1}\}$

$$\mathbb{E}\|\nabla f(x_R)\|^2 = \frac{1}{n}\sum_{k=0}^{n-1}\|\nabla f(x_k)\|^2 \le 2L(f(x_0) - \bar{f})/n$$

Convex Case.

+ L-smooth

$x^* \rightarrow$ optima

First-order convexity: $\quad f(y) \ge f(x) + \nabla f(x)^T (y - x)$

Setting $y = x^*$, $x = x_k$

$$f(x^*) \ge f(x_k) + \nabla f(x_k)^T (x^* - x_k) \quad — ②$$

Also,

$$f(x_{k+1}) \le f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2 \quad — ③$$

② & ③ $\Rightarrow$

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

$$= f(x^*) + \frac{L}{2}\left(\|x_k - x^*\|^2 - \|x_k - x^* - \frac{1}{L}\nabla f(x_k)\|^2\right)$$

GD $\rightarrow$ $x_{k+1} = x_k - \alpha \nabla f(x_k)$, $\alpha = 1/L$

$$= f(x^*) + \frac{L}{2}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right) \quad ④$$

Repeated application of ④ leads to

$$\sum_{k=0}^{n-1} f(x_{k+1}) - f(x^*) \le \frac{L}{2}\sum_{k=0}^{n-1}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right)$$

telescopic sum

$$= \frac{L}{2}\left(\|x_0 - x^*\|^2 - \|x_n - x^*\|^2\right)$$

$$\leq \frac{L}{2}\|x_0 - x^*\|^2$$

Since $\quad f(x_0) \geq f(x_1) \geq f(x_2) \cdots \geq f(x_n)$

$$\underbrace{f(x_n) - f(x^*)}_{\text{optimization error}} \leq \frac{1}{n}\sum_{k=0}^{n-1} f(x_{k+1}) - f(x^*) \leq \frac{L}{2n}\|x_0 - x^*\|^2$$

$$\boxed{f(x_n) - f(x^*) \leq \frac{L}{2n}\|x_0 - x^*\|^2} \leftarrow \text{Bound for Convex Case}$$

Strongly-Convex Case
+ L-smooth

Recall A differentiable function f is m-strongly convex if

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

Claim: A m-strongly convex function f satisfies

$$f(x) - f(x^*) \leq \frac{1}{2m}\|\nabla f(x)\|^2 \rightarrow \text{PL condition}$$

(Polyak-Lojasiewicz)

Pf: We know $\quad f(y) - f(x) \geq \nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|^2$

$$\underbrace{\min_y \left(f(y) - f(x)\right)}_{\substack{\shortparallel \\ f(x^*) - f(x)}} \geq \min_y \underbrace{\left(\nabla f(x)^T(y - x) + \frac{m}{2}\|y - x\|^2\right)}_{\text{differentiate}}$$

$$\nabla f(x) + m(y^* - x) = 0$$

$$y^* = -\frac{1}{m}\nabla f(x) + x$$

m

$$f(x^*) - f(x) \geq -\frac{1}{m}\|\nabla f(x)\|^2 + \frac{1}{2m}\|\nabla f(x)\|^2$$

(or) $$f(x^*) - f(x) \geq -\frac{1}{2m}\|\nabla f(x)\|^2$$

Remark:

From PL-condition,

$$f(x) - f(x^*) \leq \frac{\|\nabla f(x)\|^2}{2m}$$

If $\|\nabla f(x)\|$ is small, say $\|\nabla f(x)\| \leq \delta$, then

$$f(x) - f(x^*) \leq \frac{\delta^2}{2m}$$ $\Leftarrow$ Closeness in function value.

"Closeness in the parameter"

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

Cauchy-Schwarz inequality $|u^T v| \leq \|u\|\|v\|$

$$\geq f(x) - \|\nabla f(x)\|\|x^* - x\| + \frac{m}{2}\|x - x^*\|^2$$

$$\cancel{f}(x) \geq f(x^*) \geq \cancel{f}(x) - \|\nabla f(x)\|\|x^* - x\| + \frac{m}{2}\|x - x^*\|^2$$

$\Rightarrow$ $$\frac{2}{m}\|\nabla f(x)\|\|x - x^*\| \geq \|x - x^*\|^2$$

(or) $$\boxed{\|x - x^*\| \leq \frac{2\|\nabla f(x)\|}{m}}$$ $\Leftarrow$ Closeness in parameter

End of lecture

Lecture-15

GD for strongly-convex functions:

We know

(1) $f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2$ — holds for GD on a smooth f

Using $\|\nabla f(x_k)\|^2 \geq 2m(f(x_k) - f(x^*))$ — PL-condition

$f(x_{k+1}) \leq f(x_k) - \frac{m}{L}(f(x_k) - f(x^*))$ → using PL condition in (1)

$f(x_{k+1}) - f(x^*) \leq f(x_k) - f(x^*) - \frac{m}{L}(f(x_k) - f(x^*))$

(or) $$\boxed{f(x_{k+1}) - f(x^*) \leq \left(1 - \frac{m}{L}\right)(f(x_k) - f(x^*))}$$

$$\leq \left(1 - \frac{m}{L}\right)^2 (f(x_{k-1}) - f(x^*))$$

$\vdots$

$$f(x_n) - f(x^*) \leq \left(1 - \frac{m}{L}\right)^n (f(x_0) - f(x^*))$$

$$\boxed{f(x_n) - f(x^*) \leq \exp\left(-\frac{nm}{L}\right)(f(x_0) - f(x^*))}$$

Remark:-

(I) For convex + smooth f,

$$f(x_n) - f(x^*) = O\left(\frac{1}{n}\right)$$

(2) m- strongly convex + L-smooth

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

$$\leq f(y)$$

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

$$\Rightarrow \quad m \leq L \quad (\text{or}) \quad \frac{L}{m} \geq 1$$

from GD bound:

$$f(x_n) - f(x^*) \leq \exp\left(-\frac{n}{\left(\frac{L}{m}\right)}\right)\left(f(x_0) - f(x^*)\right)$$

(3) Recall m-strong convexity

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

At $x = x^*$, $y = x$,

$$f(x) \geq f(x^*) + \frac{m}{2}\|x-x^*\|^2$$

$$\boxed{\|x-x^*\|^2 \leq \frac{2}{m}\left(f(x) - f(x^*)\right)}$$

$\rightarrow$ useful to relate parameter error to optimization error

Using GD bound, $\|x_n - x^*\|^2 \leq \frac{2}{m}\left(f(x_n) - f(x^*)\right)$

$$\leq \frac{2}{m} \exp\left(-\frac{n}{\left(\frac{L}{m}\right)}\right) \left(f(x_0) - f(x^*)\right)$$

$$\leq \frac{2}{m} \exp\left(-\frac{n}{(L/m)}\right) \times \frac{L}{2} \|x_0 - x^*\|^2$$

Iteration Complexities

	Condition	Condition satisfied for
Smooth f	$\|\nabla f(x_k)\| \leq \epsilon$ for some $k \leq n$ → ϵ - First-order stationary point	$n \geq \frac{2L(f(x_0) - f(x^*))}{\epsilon^2}$ since $\min_k \|\nabla f(x_k)\| \leq \sqrt{\frac{2L(f(x_0) - f(x^*))}{n}}$
Smooth + Convex f	$f(x_n) - f(x^*) \leq \epsilon$	$n \geq \frac{L\|x_0 - x^*\|^2}{2\epsilon}$ since $f(x_n) - f(x^*) \leq \frac{L}{2n}\|x_0 - x^*\|^2$
Smooth + Strongly Convex f	$f(x_n) - f(x^*) \leq \epsilon$	$n \geq \frac{L}{m} \log\left(\frac{f(x_0) - f(x^*)}{\epsilon}\right)$ Since $f(x_n) - f(x^*) \leq \exp\left(-\frac{n}{\left(\frac{L}{m}\right)}\right)(f(x_0) - f(x^*))$