Convergence of gradient descent

- iteration complexity of gradient descent
- quadratic model
- strong convexity and linear convergence.

Lipschitz gradient and descent lemma

a function $f: \mathbb{R}^n \to \mathbb{R}$ is L-Lipschitz cont. differentiable ($f \in C_L^{1,1}$):

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\| \quad \forall\, x,y \in \mathbb{R}^n$$

examples :
① Linear $f(x) = a^T x \Rightarrow L = 0$

② quadratic $f(x) = \frac{1}{2} x^T A x + b^T x \Rightarrow L = \|A\|_2 = \lambda_{max}(A)$

2nd-order equivalence: if f is twice cont. diff'l then
[Theorem 4.20 in Beck]

$$f \in C_L^{1,1} \iff \|\nabla^2 f(x)\| \leq L \quad \forall\, x \in \mathbb{R}^n$$

descent lemma : if $f \in C_L^{1,1}$ over $\mathbb{R}^n$ then
[Lemma 4.2 in Beck]

$$f(z) \leq f(x) + \nabla f(x)^T (z-x) + \frac{L}{2}\|z-x\|^2 \quad \forall\, x,z \in \mathbb{R}^n$$

quadratic upper bound

choose any $\alpha \in (0, \frac{1}{L})$. Then by descent Lemma

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

$$\le f(x) + \nabla f(x)^T (z-x) + \frac{1}{2\alpha} \|z-x\|^2 \quad \forall\ x, y \in \mathbb{R}^n$$

(projected) gradient descent step minimizes quadratic upper bound:

$$\underset{z \in C}{\operatorname{argmin}}\ f(x) + \nabla f(x)^T (z-x) + \frac{1}{2\alpha} \|z-x\|^2 =$$

$$= \underset{z \in C}{\operatorname{argmin}}\ \frac{\alpha}{2} \|\nabla f(x)\|^2 + \nabla f(x)^T (z-x) + \frac{1}{2\alpha} \|z-x\|^2$$

$$= \underset{z \in C}{\operatorname{argmin}}\ \frac{1}{2\alpha} \| z - x + \alpha \nabla f(x) \|^2$$

$$= \operatorname{proj}_C (x - \alpha \nabla f(x))$$

(valid for any convex set $C \subseteq \mathbb{R}^n$)

convergence gradient descent (take $C = \mathbb{R}^n$ for simplicity)

by descent lemma: $f_k := f(x_k)$ and $\nabla f_k := \nabla f(x_k)$

$$f_{k+1} \leq f_k + \nabla f_k^T (x_{k+1} - x_k) + \frac{L}{2} \| x_{k+1} - x_k \|^2$$

take $x_{k+1} := x_k - \alpha \nabla f_k \implies x_{k+1} - x_k = -\alpha \nabla f_k$

$$\begin{aligned} f_{k+1} &\leq f_k - \alpha \nabla f_k^T \nabla f_k + \frac{L}{2} \| -\alpha \nabla f_k \|^2 \\ &= f_k - \alpha \| \nabla f_k \|^2 + \frac{\alpha^2 L}{2} \| \nabla f_k \|^2 \\ &= f_k - \alpha \left(1 - \frac{\alpha L}{2} \right) \| \nabla f_k \|^2 \end{aligned}$$

objective reduces if

$f_{k+1} < f_k$ if $\alpha \in (0, \frac{2}{L})$ and $\nabla f_k \neq 0$

$(*)\quad f_{k+1} \leq f_k - \alpha\left(1 - \frac{\alpha L}{2}\right) \|\nabla f_k\|^2$

take steplength $\alpha \in (0, \frac{2}{L})$ so that

$(**)\quad f_{k+1} \leq f_k - \frac{1}{2}\alpha \|\nabla f_k\|^2 \implies \frac{1}{2}\alpha \|\nabla f_k\|^2 \leq f_k - f_{k+1}$

sum across iterations $k = 0, 1, 2, \cdots, T$:

$$\frac{1}{2}\alpha \sum_{k=0}^{T} \|\nabla f_k\|^2 \leq f(x_0) - f(x_T) \leq f(x_0) - f^* \qquad f^* = \text{min value}$$

$$\implies \min_{k \in \{0, \cdots, T\}} \|\nabla f(x_k)\|^2 \leq \frac{1}{T} \sum_{k=0}^{T} \|\nabla f(x_k)\|^2 \leq \frac{2(f(x_0) - f^*)}{\alpha T}$$

<u>NOTE</u> : Choosing $\alpha = \frac{1}{L}$ minimizes RHS of $(*)$.

strong convexity

a function $f: \mathbb{R}^n \to \mathbb{R}$ is μ strongly convex $(\mu > 0)$ if

$$f(z) \geq f(x) + \nabla f(x)^T (z-x) + \frac{\mu}{2} \|z-x\|^2 \quad \forall\, x,y \in \mathbb{R}^n$$

if f is twice continuously diff'l, then equivalent to

$$\| d^T \nabla^2 f(x) d \| \geq \mu \|d\|^2 \iff \lambda_{min}(\nabla^2 f(x)) \geq \mu \quad \forall\, x \in \mathbb{R}^n$$

implies

$$\|\nabla f(x)\|^2 \geq 2\mu \left(f(x) - f^* \right)$$

compare with L-Lipschitz gradient:

$$f(z) \leq f(x) + \nabla f(x)^T (z-x) + \frac{L}{2} \|z-x\|^2 \quad \forall\, x,y \in \mathbb{R}^n$$

and

$$\lambda_{max}(\nabla^2 f(x)) \leq L \quad \forall\, x \in \mathbb{R}^n$$

convergence under strong convexity

from (**) above,

$$f_{k+1} \leq f_k - \frac{1}{2}\alpha \|\nabla f_k\|^2$$

using strong convexity

$$f_{k+1} \leq f_k - \alpha\mu (f_k - f^*)$$

subtract f^* from both sides

$$f_{k+1} - f^* \leq f_k - f^* - \alpha\mu(f_k - f^*) = (1-\alpha\mu)(f_k - f^*)$$

recursing from $k = T, T-1, \ldots, 2, 1, 0$:

$$f_T - f^* \leq (1-\alpha\mu)^T (f_0 - f^*) \leq \exp(-\alpha\mu T)(f_0 - f^*)$$

take $\alpha = \frac{1}{L}$ and solve $\quad \varepsilon \leq \exp\left(-\frac{\mu T}{L}\right)(f_0 - f^*)$

$$\Rightarrow \quad T \geq \frac{L}{\mu} \log\left(\frac{f(x_0) - f^*}{\varepsilon}\right) \quad \leftarrow \text{linear convergence}$$

"condition number" of f