TD: Gradient Descent for convex and smooth functions

Faïcel Chamroukhi

week 4-5 - Nov. 28 (lecture). Dec 05. 2024

Convergence Analysis

We study the convergence for a fixed step size α. Prove the following result.

Theorem Assume that $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L-smooth. If x^* is a critical point of f, i.e., $\nabla f(x^*) = 0$, then the the sequence $\{x^{(k)}\}$ generated by gradient descent

$$x^{(k+1)} = x^{(k)} + \alpha \nabla f(x^{(k)}),$$

with fixed step size $0 \leq \alpha \leq \frac{1}{L}$ satisfies:

$$f(x^{(k)}) - f(x^\star) \leq \frac{\|x^{(0)} - x^\star\|^2}{2\alpha k}.$$

Optimization TD 05.12.2024

Convergence Analysis

Assume that $f: \mathbb{R}^n \to \mathbb{R}$ is convex and L-smooth. If x^* is a critical point of f, i.e., $\nabla f(x^*) = 0$, then the sequence $\{x^{(k)}\}$ generated by gradient descent $x^{(k+1)} = x^{(k)} + \alpha \nabla f(x^{(k)})$, with fixed step size $0 \le \alpha \le \frac{1}{L}$ satisfies

$$f(x^{(k)}) - f(x^*) \le \frac{\| x^{(0)} - x^* \|^2}{2 \alpha k}$$

Proof: L-smooth ($L>0$) when $f(x) \le f(\theta) + \nabla(\theta)^T(x-\theta) + \frac{L}{2}\|\theta + x\|^2, \ \forall \theta, x \in \mathbb{R}^n$

f is convex: $f(x) \le f(x^*) + \nabla^T f(x)(x - x^*)$

Using smoothness property we have:

$$f(y) \le f(x) + \nabla^T f(x)(y-x) + \frac{L}{2}\| y - x\|_2^2$$

* proof: f is L-smooth, then ∇f is L-lipshite continues: $\exists L>0$ such that:

$\nabla^2 f(x) \le L I$, equiv: $\nabla^2 f(x) - L I \le 0$

$\nabla^2 f(x) - L I$ is semi-defined negotaile: $\forall x, y, z: (x-y)^T(\nabla^2 f(z) - L I)(x-y) \le 0$

$\Leftrightarrow (x-y)^T \nabla^2 f(z)(x-y) - L \| y - x\|^2 \le 0$

normalizing $(x-y)^T \nabla^2 f(z)(x-y) \le L \|y-x\|^2$

Based on the Taylor Reminder Theorem, we have

$\forall x, y, \exists z \in [x, y]$ such that

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(x-y)^T \nabla^2 f(z)(x-y)$$

Sutisfituting the found (1) into this Taylor approximation we get: what we need to find

$$f(y) \le f(x) + \nabla^T f(x)(y-x) + \frac{1}{2} L \| y - x\|^2$$

Gradient

$$x^{(k+1)} = x^k - x^k \nabla f(x_k)$$

$$x^* = x - \alpha \nabla f(x)$$

let $y = x^t$

Pluging in the smoothing:

$$f(x) + \nabla^T f(x)(x^t - x) + L/2 \|x^t - x\|^2$$

we get $f(x^*) \le f(x) + \nabla^T f(x)(x - \alpha \nabla f(x) - x) + \frac{L}{2}\| \underbrace{x - \alpha \nabla f(x)}_{x^t} - x\|^2$

$$\le f(x) + \nabla^T f(x)(\underbrace{x - \alpha \nabla f(x)}_{x^t} - x) + \frac{L}{2}\alpha^2 \| \nabla f(x)\|^2$$

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

$$= f(x) - \alpha \|\nabla f(x)\|^2 + \frac{L}{2}\alpha^2 \|\nabla f(x)\|_2^2$$

$$= f(x) - \left(1 - \frac{L\alpha}{2}\right)\alpha \| \nabla f(x)\|$$

for $0 \le \alpha \le \frac{1}{L}$, we have $1 - \frac{L\alpha}{2} \ge \frac{1}{2}$

Therefore: $f(x^+) \leq f(x) - \frac{\alpha}{2} \|\nabla f(x)\|^2$; $f(x^+) - f(x) \leq -\frac{\alpha}{2}\|\nabla f(x)\|^2$

by the convexity property of f we have $f(x) \leq f(x^*) + \nabla f(x)^T (x - x^*)$

$\forall x$ in dom

Thus: $\underbrace{f(x^+) \leq f(x) - \frac{\alpha}{2}\|\nabla f(x)\|^2 \leq}$ $f(x^*) + \nabla^T f(x)(x - x^*) - \frac{\alpha}{2}\|\nabla f(x)\|^2$

$$f(x^*) + \frac{1}{2\alpha}\left(2\alpha \nabla^T f(x)(x - x^*) - \alpha^2 \|\nabla f(x)\|^2\right.$$

$2\alpha \nabla^T f(x)(x - x^*) - \alpha^2 \|\nabla f(x)\|^2$ is apart of $\|a - b\|^2$ with $a = x - x^*$, $b = \alpha \nabla f(x)$

So we have : $2\alpha \nabla^T f(x)(x - x^*) - \alpha^2 \|\nabla f(x)\|^2 = \|x - x^*\|^2 - \|x - x^* - \alpha \nabla f(x)\|^2$

$\underbrace{(2ab - b^2)}$ $= \|x - x^*\|^2 - \|x^+ - x^*\|^2$ We finally got:

$\Rightarrow f(x^+) - f(x^*) \leq \frac{1}{2}\left(\|x - x^*\|^2 - \|x^+ - x^*\|^2\right)$

This inequelity holds for x^+ at every iteration

Summing over iteration we get:

$$\sum_{k=1}^{k} (f(x^k))\, f(x^*) \leq \sum_{k=1}^{k} \frac{1}{2\alpha}\left(\|x^{(i-1)} - x^*\|^2 - \|x^{(i)} - x^*\|^2\right)$$

Telescoping

$$= \frac{1}{2\alpha}\left(\|x^{(0)} - x^*\|^2 - \|x^k - x^*\|^2\right) \leq \frac{1}{2\alpha}\|x^{(0)} - x^*\|^2$$

We obtain: $\sum_{i=1}^{k} \left(f(x^{(i)}) - f(x^*)\right) \leq \frac{1}{2\alpha}\|x^{(0)} - x^*\|^2$

Sine f is non increasing we can write:

$$k f(x^{(k)}) \leq \sum_{k=1}^{k} f(x^{(i)}) \Rightarrow k\left(f(x^{(k)}) - f(x^*)\right) \leq \sum_{i=1}^{k} \left(f(x^{(i)}) - f(x)\right) \Rightarrow$$

$$\Rightarrow f(x^{(k)}) - f(x^*) \leq \frac{1}{k}\sum_{i=1}^{k} \left(f(x^{(i)})\, f(x^*)\right) \Rightarrow$$

$$f(x^{(k)}) - f(x^*) \leq \frac{1}{2\alpha k}\|x^{(0)} - x^{(*)}\|^2$$ What we need