Class #16

Tuesday, March 25, 2025 8:59 AM

Convergence Analysis of Gradient Descent for strongly Convex functions

$\rightarrow$ strong convexity parameter m

Assumptions: $f \in C^2(\mathbb{R}^n)$, m-strongly convex (m-smooth)

$$\underbrace{mI}_{\text{global bound}} \preceq \nabla^2 f(x) \preceq \underbrace{MI}_{\text{local on the sublevel set}}$$

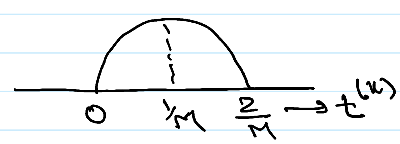
$$S = \{x : f(x) \le f(x^{(0)})\}$$

Iteration: $x^{(k+1)} = x^{(k)} - t^{(k)} \nabla f(x^{(k)})$ for $t^{(k)}$ s.t. $x^{(k+1)} \in \text{dom} f$

Quadratic upper bound $\forall x \in S$

$$f(x^{(k+1)}) \le f(x^{(k)}) - t^{(k)} \|\nabla f(x^{(k)})\|_2^2 + \frac{M t^{(k)^2}}{2} \|\nabla f(x^{(k)})\|_2^2$$

$$= f(x^{(k)}) - \underbrace{\left(-M\frac{t^{(k)^2}}{2} + t^{(k)}\right)}\|\nabla f(x^{(k)})\|_2^2$$



Exact line Search Analysis

When we do exact line search

$$f(x^{(k+1)}) \le \min_{t^{(k)}} \left(f(x^{(k)}) - \underbrace{\left(t^{(k)} - \frac{M\, t^{(k)2}}{2} \right)}_{t^{(k)} = \frac{1}{M}\ \star} \|\nabla f(x^{(k)})\|_2^2 \right)$$

$$\star \ - \ f(x^{(k+1)}) \le f(x^{(k)}) - \frac{1}{2M} \|\nabla f(x^{(k)})\|_2^2$$

Since f is m-strongly convex, we know that:

$$f(x^{(k)}) - p^* \le \frac{1}{2m} \|\nabla f(x^{(k)})\|_2^2$$

Subtract p^* from both sides of $\star$

$$f(x^{(k+1)}) - p^* \le f(x^{(k)}) - p^* - \frac{1}{2M} \underbrace{\|\nabla f(x^{(k)})\|_2^2}_{\ge 2m(f(x^{(k)}) - p^*)}$$

$$f(x^{(k+1)}) - p^* \le \underbrace{f(x^{(k)}) - p^*} - \frac{m}{M} \underbrace{(f(x^{(k)}) - p^*)}$$

$$\Rightarrow f(x^{(k+1)}) - p^* \le \underbrace{\left(1 - \frac{m}{M}\right)}_{0 \le c < 1} \left(f(x^{(k)}) - p^*\right)$$

If $\frac{M}{m} = 1 \Rightarrow$ Exact line search gives us the solution in one iteration.

$$f(x^{(k+1)}) - p^* \leq c\left(f(x^{(k)}) - p^*\right)$$

Let's recursively apply the above expression from $k=0$ to K.

$$f(x^{(K)}) - p^* \leq c^K\left(f(x^{(0)}) - p^*\right) \leq \epsilon$$

Fact 1: As $K \to \infty$, $f(x^{(K)}) \longrightarrow p^*$

Rate of Convergence:

We want $f(x^{(k)}) - p^* \leq \epsilon$

$$\Leftarrow \quad c^k\left(f(x^{(0)}) - p^*\right) \leq \epsilon$$

$$\Leftrightarrow \quad c^k \leq \frac{\epsilon}{f(x^{(0)}) - p^*}$$

$$\Leftrightarrow \quad k\ln(c) \leq \ln\left(\epsilon / f(x^{(0)}) - p^*\right)$$

$$\boxed{K \leq \frac{\ln\left(\epsilon / f(x^{(0)}) - p^*\right)}{\ln(c)}}$$

Ignoring all other constants: $K = O(\ln(\epsilon))$

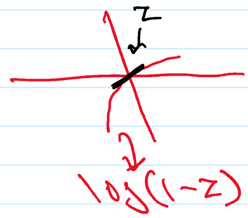
Linear Convergence

Implications: ① As $\epsilon \downarrow$, K has to increase only logarithmically with ϵ.

② As $f(x^{(0)}) - p^* \uparrow$, K has to increase, but only logarithmically.

③ K is proportional to the condition number of the problem $\tilde{K} = M/m$ $\approx$ linearly proportional

$\log(1-z) \approx -z$ when z is small



Remember: $c = 1 - \frac{m}{M}$; when $\frac{M}{m}$ is large,

$\Rightarrow \frac{m}{M}$ is small

$$\Rightarrow \ln(c) \approx \frac{m}{M}$$

$$\Rightarrow K \le \left|\ln\left(\epsilon / f(x^{(0)}) - p^*\right)\right| \cdot \underbrace{\frac{M}{m}}_{\tilde{K}}$$

Condition number has a fundamental role in determining the performance of gradient descent.

Analysis of Backtracking line Search

There are two key themes for this analysis.

① Backtracking line search accepts a step size that is (i) $\boxed{t \in [0, \frac{1}{M}]}$

(ii) It is either exactly 1 ✓

or it is in the interval $\underbrace{\left[\frac{\beta}{M}, \frac{1}{M}\right]}_{\star}$.

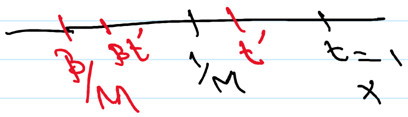
Backtracking evaluation:

Backtracking evaluation:

Accept t if

$$f(x^{(k+1)}) \leq f(x^{(k)}) - \alpha t \|\nabla f(x^{(k)})\|_2^2$$

for $\alpha \in (0, 1/2)$



Conclusion: $f(x^{(k+1)}) \leq f(x^{(k)}) - \min\left\{\alpha, \frac{\alpha\beta}{M}\right\} \|\nabla f(x^{(k)})\|_2^2$

Convergence result:

$$f(x^{(k)}) - p^* \leq c^k \left(f(x^{(0)}) - p^*\right)$$

where $c = 1 - \min\left\{2m\alpha, 2\alpha\beta \frac{m}{M}\right\}$

Compare to exact line search:

$$c = 1 - \frac{m}{M}$$

Why should α be $\leq 1/2$? $\Rightarrow$ Ensures $c \in [0,1)$.
Exact line search c is better than backtracking, but both give linear convergence.