Revisiting gradient descent : We will show convergence with

only Lipshitz continuity

Lipschitz continuity + strong convexity

We have $\forall x, c, y \in \mathbf{dom}\, f$,

- $f(y) = f(x) + \nabla^\top f(x)(y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(c)(y-x)$
- $\nabla^2 f(c) \preceq LI$
  - $\implies (y-x)^\top \nabla^2 f(c)(y-x) \leq L\|y-x\|^2$
- Thus we get

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

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

Generally, convergence analysis will require Lipshitz continuity

- Considering $x^k \equiv x$, and $x^{k+1} \equiv y$, and a fixed step size t, we get

$$f(x^{k+1}) \leq f(x^k) - t\nabla^\top f(x^k)\nabla f(x^k) + \frac{Lt^2}{2}\left\|\nabla f(x^k)\right\|^2$$

$$\implies f(x^{k+1}) \leq f(x^k) - (1 - \frac{Lt}{2})t\left\|\nabla f(x^k)\right\|^2$$

- Taking $0 < t \leq \frac{1}{L} \implies 1 - \frac{Lt}{2} \geq \frac{1}{2}$, we have

$$f(x^{k+1}) \leq f(x^k) - \frac{t}{2}\left\|\nabla f(x^k)\right\|^2$$

Note: $t < \frac{2}{L}$ is necessary for convergence

For a weaker notion of convergence discussed here $t \leq 2/L$ should be ok. So you can replace t with $t/2$ on this & next pages & results should still hold.

- Using convexity, we have $f(x^*) \geq f(x^k) + \nabla^\top f(x^k)(x^* - x^k)$
  $\implies f(x^k) \leq f(x^*) + \nabla^\top f(x^k)(x^k - x^*)$
- Thus,
  $f(x^{k+1}) \leq f(x^k) - \frac{t}{2}\left\|\nabla f(x^k)\right\|^2$
  $\implies f(x^{k+1}) \leq f(x^*) + \nabla^\top f(x^k)(x^k - x^*) - \frac{t}{2}\left\|\nabla f(x^k)\right\|^2$
  $\implies f(x^{k+1}) \leq f(x^*) - \frac{1}{2t}\left\|x^k - x^*\right\|^2 + \nabla^\top f(x^k)(x^k - x^*) -$
  $\frac{t}{2}\left\|\nabla f(x^k)\right\|^2 + \frac{1}{2t}\left\|x^k - x^*\right\|^2$
  $\implies f(x^{k+1}) \leq f(x^*) + \frac{1}{2t}(\left\|x^k - x^*\right\|^2 - \left\|x^k - x^* - t\nabla f(x^k)\right\|^2)$
  $\implies f(x^{k+1}) \leq f(x^*) + \frac{1}{2t}(\left\|x^k - x^*\right\|^2 - \left\|x^{k+1} - x^*\right\|^2)$

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

- Over all iterations, we have

$$\sum_{i=0}^{k}\left(f(x^i)-f(x^*)\right)\leq\frac{1}{2t}\left(\left\|x^{(0)}-x^*\right\|^2\right)-\frac{1}{2t}\left\|x^{(k+1)}-x^*\right\|^2$$

≤ 0

- Since $f(x^{k+1})\leq f(x^k)\ \forall k=0,1,\ldots,$ we get

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

Q: what abt rate of convergence?

Question: Could we analyze Gradient descent more generally?

- Assume backtracking line search OR exact line search?
- Continue assuming Lipschitz continuity
  - Curvature is upper bounded: $\nabla^2 f(x) \preceq LI$
- Assume **strong convexity**
  - Curvature is lower bounded: $\nabla^2 f(x) \succeq mI$
  - For instance, we wouldn't want to use gradient descent for a linear function (no curvature)

How can we get Q or R linear convergence?

$\theta \in (0,1)$

$$f(x^k) - f(x^*) \leq \theta^k \left(f(x^0) - f(x^*)\right) \to v^k \Rightarrow \text{R-linear}$$

$$f(x^k) - f(x^*) \leq \theta \left(f(x^{k-1}) - f(x^*)\right) \Rightarrow \text{Q-linear}$$

- Lipschitz continuity

$$\nabla^2 f(x) \preceq LI$$

$$\left\|\nabla f(x) - \nabla f(y)\right\| \leq L\|x - y\|$$

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

- Convexity
  - Curvature should **not** be negative

$$\nabla^2 f(x) \succeq 0$$

$$f(y) \geq f(x) + \nabla^\top f(x)(y - x)$$

- Strong convexity

$$\nabla^2 f(x) \succeq mI$$

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

  - For example, augmented Lagrangian is used to introduce strong convexity

Using strong convexity

- $f(y) \geq f(x) + \nabla^\top f(x)(y-x) + \frac{m}{2}\|y-x\|^2$
  $\geq$ minimum value the RHS can take as a function of y
- Minimum value of RHS
  $\nabla f(x) + my - mx = 0$
  $\implies y = x - \frac{1}{m}\nabla f(x)$
- Thus,
  $f(y) \geq f(x) + \nabla^\top f(x)\left(-\frac{1}{m}\nabla f(x)\right) + \frac{m}{2}\left\|-\frac{1}{m}\nabla f(x)\right\|^2$
  $\implies f(y) \geq f(x) - \frac{1}{2m}\left\|\nabla f(x)\right\|^2$
  - Here, LHS is independent of x, and RHS is independent of y

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

- If $\left\|\nabla f(x)\right\|$ is small, the point is nearly optimal
  - If $\left\|\nabla f(x)\right\| \leq \sqrt{2m\epsilon}$, then: $f(x) - f(x^*) \leq \epsilon$
  - As the gradient $\left\|\nabla f(x)\right\|$ approaches 0, we get closer to the optimal solution x^*

} A knob provided by strong convexity!

Analysis for Backtracking Line Search

- Backtracking line search exits when

$$f\left(x^k - t\nabla f(x^k)\right) \leq f(x^k) - \frac{t}{2}\left\|\nabla f(x^k)\right\|^2$$

  - where $t = (\beta)^r\, t_{orig}$
    - t_{orig} was the initial step size before the invocation of backtracking line search
    - r is the number of iterations before the loop terminated
- The margin of backtracking line search, $\frac{t}{2}\left\|\nabla f(x^k)\right\|^2$, is inspired by strong convexity

- Since f is strongly convex, and also Lipschitz continuous, we have for some L :

$$f(x^{k+1}) \leq f(x^k) + (\frac{Lt^2}{2} - t)\left\|\nabla f(x^k)\right\|^2$$

- We also consider
  $0 < t \leq \frac{1}{L} \implies t^2 \leq \frac{t}{L} \implies \frac{Lt^2}{2} \leq \frac{t}{2}$
  $\implies \frac{Lt^2}{2} - t \leq -\frac{t}{2}$
- Thus, we get the exit condition of backtracking line search

$$f(x^{k+1}) \leq f(x^k) - \frac{t}{2}\left\|\nabla f(x^k)\right\|^2$$

$$\implies f\left(x^k - t\nabla f(x^k)\right) \leq f(x^k) - \frac{t}{2}\left\|\nabla f(x^k)\right\|^2$$

- Convergence of gradient descent, given this condition, has been proved below