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^*

Did not require Lipschitz continuity

Analysis for Backtracking Line Search

- Backtracking line search exits when

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

sufficient decrease condition

$\rightarrow$ (4.90) for gradient descent with $c_1 = \frac{1}{2}$

  - 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

Note: The analysis that follows should hold $\forall\ c_1 < 1$

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

Not required immediately

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

- 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 (BTLS)

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

If ∇f is is Lipschitz

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

then exit condition of BTLS should be met for some $t \in (0, 1/L]$ if $c_1 = \frac{1}{2}$

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

- Let $p^* = f(x^*)$
- $f\left(x - t\nabla f(x)\right) \leq \overbrace{f(x) - t\left\|\nabla f(x)\right\|^2 + \frac{Lt^2}{2}\left\|\nabla f(x)\right\|^2}^{\text{Maximize RHS }(t)}$
  - RHS here will be maximum for $t = \frac{1}{L}$

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

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

(Subtracted p^* from both sides)

- From strong convexity, we had

$f(y) \geq f(x) - \frac{1}{2m}\left\|\nabla f(x)\right\|^2$ → already proved

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

$$\implies \left\|\nabla f(x)\right\|^2 \geq 2m\left(f(x) - p^*\right)$$

Note: We know $m < L$. Now show Q-linear | R-linear convergence

- Thus,

$$f\left(x - t^* \nabla f(x)\right) - p^* \leq f(x) - \frac{1}{2L} \left\|\nabla f(x)\right\|^2 - p^*$$
$$\implies f\left(x - t^* \nabla f(x)\right) - p^* \leq f(x) - \frac{2m}{2L}\left(f(x) - p^*\right) - p^*$$
$$\implies f\left(x - t^* \nabla f(x)\right) - p^* \leq \left(1 - \frac{m}{L}\right)\left(f(x) - p^*\right)$$ → Q-Linear

- Which is,

$$f(x^k) - p^* \leq \left(1 - \frac{m}{L}\right)\left(f(x^{k-1}) - p^*\right)$$
$$\leq \left(1 - \frac{m}{L}\right)^2 \left(f(x^{k-2}) - p^*\right)$$
$$\vdots$$
$$\leq \left(1 - \frac{m}{L}\right)^k \left(f(x^{(0)}) - p^*\right)$$ → R-Linear

- We get linear convergence

$$f(x^k) - p^* \leq \left(1 - \frac{m}{L}\right)^k \left(f(x^{(0)}) - p^*\right)$$

  - Here, $\frac{m}{L} \in (0, 1)$
  - This is, loosely speaking, faster than what we got using only Lipschitz continuity, which was:
    $f(x^k) - p^* \leq \frac{\left\|x^{(0)} - x^*\right\|^2}{2tk}$
    (sublinear convergence)