Gradient Descent under strong convexity

Axel Böhm

October 13, 2021

1. Introduction

2. Convergence analysis

How fast can we go?

- So far we explored different smoothness properties.
- Error decreased with $1/k$ or $1/\sqrt{k}$
- call these rates sublinear
- Linear rate means

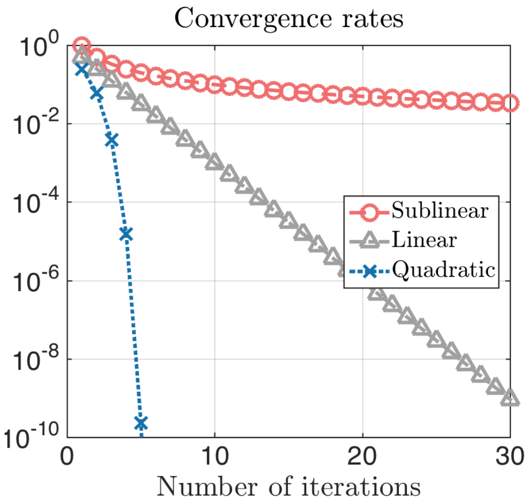
$$Err(x_k) \leq \frac{C}{\exp(k)}$$

or

$$Err(x_{k+1}) \leq qErr(x_k)$$

with $q < 1$.

Linear convergence



Example

- Consider $f(x) = x^2$. Clearly, f is $L = 2$ smooth.
  - So we can pick $\alpha = 1/L = 1/2$ for GD:

$$x_{k+1} = x_k - \frac{1}{2}\nabla f(x_k) = x_k - x_k = 0.$$

  - Converges **in one step!**
- Same $f(x) = x^2$, but is also $L = 4$ smooth.
  - So we can pick $\alpha = 1/L = 1/4$ for GD:

$$x_{k+1} = x_k - \frac{1}{4}\nabla f(x_k) = x_k - \frac{1}{2}x_k = \frac{1}{2}x_k.$$

  - Converges exponentially

$$f(x_k) = f\left(\frac{x_0}{2^k}\right) = \frac{1}{2^{2k}}x_0^2.$$

Strongly convexity

"Not too flat."

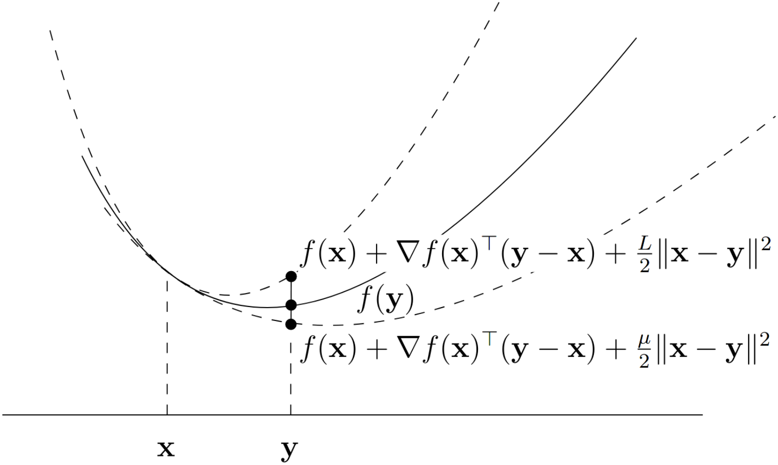
[

Recall] Let $f : \mathbb{R}^d \to \mathbb{R}$ be a differentiable function, then we say f is μ-strongly convex if

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{\mu}{2} \|y - x\|^2, \quad \forall x, y.$$

Strong convexity

Can be lowe bounded by a quadratic.



Smooth strongly convex functions

Theorem

Let $f : \mathbb{R}^d \to \mathbb{R}$ be L-smooth and μ-strongly convex. Then GD with stepsize $\alpha = 1/L$ and arbitrary starting point x_0 guarantees:

(i) *distance to solution decreases by a constant factor*

$$\|x_{k+1} - x^*\|^2 \leq \left(1 - \frac{\mu}{L}\right) \|x_k - x^*\|^2.$$

(ii) *Gives exponential decrease in function values*

$$f(x_k) - f^* \leq \left(1 - \frac{\mu}{L}\right)^k \frac{L\|x_0 - x^*\|^2}{2}.$$

Proof

$$\begin{aligned}\|x_{k+1} - x^*\|^2 &\leq \|x_k - \alpha \nabla f(x_k) - x^*\|^2 \\ &= \|x_k - x^*\|^2 + 2\alpha\langle \nabla f(x_k), x^* - x_k\rangle + \alpha^2 \|\nabla f(x_k)\|^2.\end{aligned}$$

Now we use the stronger version of the gradient inequality, namely

$$\langle \nabla f(x_k), x^* - x_k\rangle + \frac{\mu}{2}\|x^* - x_k\|^2 \leq f^* - f(x_k).$$

Combined we deduce

$$\begin{aligned}&\|x_{k+1} - x^*\|^2 \\ &\quad \leq \|x_k - x^*\|^2 + 2\alpha \left(f^* - f(x_k) - \frac{\mu}{2}\|x^* - x_k\|^2 \right) + \alpha^2 \|\nabla f(x_k)\|^2 \\ &\quad = \left(1 - \frac{\mu}{L}\right) \|x_k - x^*\|^2 + 2\alpha(f^* - f(x_k)) + \alpha^2 \|\nabla f(x_k)\|^2.\end{aligned}$$

Proof II

$$\|x_{k+1} - x^*\|^2 \leq \left(1 - \frac{\mu}{L}\right) \|x_k - x^*\|^2 \underline{+2\alpha(f^* - f(x_k)) + \alpha^2 \|\nabla f(x_k)\|^2}$$

is the desired statement up to an error which we can bound

$$\begin{aligned}\underline{2\alpha(f^* - f(x_k)) + \alpha^2 \|\nabla f(x_k)\|^2} &= \frac{2}{L}(f^* - f(x_k)) + \frac{1}{L^2}\|\nabla f(x_k)\|^2 \\ &\leq \frac{2}{L}(f(x_{k+1}) - f(x_k)) + \frac{1}{L^2}\|\nabla f(x_k)\|^2\end{aligned}$$

sufficient decrease

$$\leq -\frac{1}{L^2}\|\nabla f(x_k)\|^2 + \frac{1}{L^2}\|\nabla f(x_k)\|^2 = 0.$$

So we can ignore this extra term and get (i):

$$\|x_{k+1} - x^*\|^2 \leq \left(1 - \frac{\mu}{L}\right) \|x_k - x^*\|^2.$$

Proof III

Smoothness of f gives

$$f(x_k) - f^* \leq \langle \nabla f(x^*), x_k - x^* \rangle + \frac{L}{2}\|x_k - x^*\|^2$$

together with the fact that $\nabla f(x^*) = 0$ this gives

$$f(x_k) - f^* \leq \frac{L}{2}\|x_k - x^*\|^2.$$

If we combine this with (i)

$$f(x_k) - f^* \leq \frac{L}{2}\|x_k - x^*\|^2 \leq \frac{L}{2}\left(1 - \frac{\mu}{L}\right)^k \|x_0 - x^*\|^2.$$