Descent lemma of gradient descent

a.k.a. Sufficient-decrease condition

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : July 11, 2020
Last update : July 12, 2020

Gradient Descent and the descent lemma

- Gradient descent (GD) update

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L}\nabla f(\mathbf{x}_k), \tag{1}$$

  where k is the iteration counter, L is the Lipschitz constant of ∇f. i.e., here we assume
  - f is smooth
  - GD update uses fix step size $\frac{1}{L}$

- The descent lemma associated with GD:

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|_2^2. \tag{2}$$

- This document: prove (2).

Look into the lemma

- The descent lemma of GD

$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|_2^2$$

  tells the following story about GD:
  - Every time we perform a GD update of optimization variable $\mathbf{x}$ as in (1), the cost function f goes down at least $\frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|_2^2$.
  - At the optimum point $\mathbf{x}^*$, the gradient ∇f vanish, so the lemma agree with the equality $f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k)$ which agrees with the optimality condition of gradient descent.

Proving the lemma

- The GD update $\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L}\nabla f(\mathbf{x}_k)$ gives

$$\mathbf{x}_{k+1} - \mathbf{x}_k = -\frac{1}{L}\nabla f(\mathbf{x}_k). \tag{3}$$

- As f is L-smooth, so

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 \tag{4}$$

  See sides 11, 12 here for details.

- Put $\mathbf{y} = \mathbf{x}_{k+1}$, $\mathbf{x} = \mathbf{x}_k$ in (4)

$$\begin{aligned} f(\mathbf{x}_{k+1}) &\leq f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle + \tfrac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2 \\ &\overset{(3)}{=} f(\mathbf{x}_k) + \langle \nabla f(\mathbf{x}_k), -\tfrac{1}{L}\nabla f(\mathbf{x}_k) \rangle + \tfrac{L}{2}\|\tfrac{1}{L}\nabla f(\mathbf{x}_k)\|_2^2 \\ &= f(\mathbf{x}_k) - \tfrac{1}{2L}\|\nabla f(\mathbf{x}_k)\| \quad \square \end{aligned}$$

End of document