Descent lemma of gradient descent algorithm

Andersen Ang Email: ms3ang@uwaterloo.ca Homepage: angms.science
Department of Combinatorics and Optimization, University of Waterloo
First draft: July 11, 2020 version update: March 23, 2022

Setting

Unconstrained smooth optimization

$$P \ : \ \min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}).$$

- $f : \mathbb{R}^n \to \mathbb{R}^n$ and there is no constraint on $\mathbf{x}$.
- f is differentiable : $(\forall \mathbf{x} \in \mathbb{R}^n)\ \nabla f(\mathbf{x})$ exists.
- f is L-smooth with $L > 0$: $\nabla f(\mathbf{x})$ is L-Lipschitz.

Gradient Descent (GD) update

- Let k be the iteration counter, starting from $\mathbf{x}_0 \in \mathbb{R}^n$, update via
$$\mathbf{x}_{k+1} = \mathbf{x}_k - \frac{1}{L}\nabla f(\mathbf{x}_k), \tag{1}$$
here we use 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 at the lemma (2) tells:

- Every time we perform a GD update on variable $\mathbf{x}$, the cost function f goes down at least $\frac{1}{2L}\|\nabla f(\mathbf{x}_k)\|_2^2$.
$\implies$ f will go down as long as $\nabla f(\mathbf{x}_k) \neq \mathbf{0}$.
- At an optimum point $\mathbf{x}^*$, the gradient ∇f vanishes, so (2) gives the equality $f(\mathbf{x}_{k+1}) = f(\mathbf{x}_k)$.

The proof

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

Remark By (3), the descent lemma (2) is also equivalent to
$$f(\mathbf{x}_{k+1}) \leq f(\mathbf{x}_k) - \frac{L}{2}\|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2^2.$$

End of document