Lecture Notes: Gradient

Yufei Tao
Department of Computer Science and Engineering
Chinese University of Hong Kong
taoyf@cse.cuhk.edu.hk

Let $p(x_1, x_2, ..., x_d)$ be a point in $\mathbb{R}^d$. We will often view it as a d-dimensional vector $[x_1, x_2, ..., x_d]$. As a convention, if it has been clear from the context that p is a point, then $\boldsymbol{p}$ represents this corresponding vector.

Let $f(x_1, x_2, ..., x_d)$ be a scalar function of real-valued parameters $x_1, ..., x_d$. In other words, for each point $p(x_1, ..., x_d)$ of $\mathbb{R}^d$, $f(x_1, x_2, ..., x_d)$ returns a real value, if it is defined at p. For simplicity, sometimes we may write $f(x_1, x_2, ..., x_d)$ simply as $f(p)$. Next, we introduce a concept called *gradient* for such functions:

Definition 1. *Let $f(x_1, ..., x_d)$ be a function defined as above. Consider a point $(t_1, t_2, ..., t_d)$ at which the partial derivative $\frac{\partial f}{\partial x_i}(t_1, ..., t_d)$ exists for all $i \in [1, d]$. Then, the* **gradient** *of $f(x_1, ..., x_d)$ at $(t_1, t_2, ..., t_d)$ is the vector:*

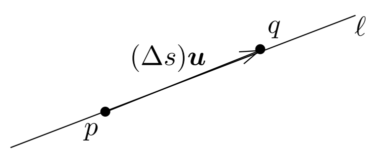
$$\nabla f(t_1, ..., t_d) = \left[\frac{\partial f}{\partial x_1}(t_1, ..., t_d), \frac{\partial f}{\partial x_2}(t_1, ..., t_d), ..., \frac{\partial f}{\partial x_d}(t_1, ..., t_d)\right].$$

For example, suppose that $f(x, y, z) = x^3 + 2xy + 3xz^2$. We know that $\frac{\partial f}{\partial x} = 3x^2 + 2y + 3z^2$, $\frac{\partial f}{\partial y} = 2x$, and $\frac{\partial f}{\partial z} = 6x$. Therefore,

$$\nabla f(x, y, z) = \left[3x^2 + 2y + 3z^2, 2x, 6x\right].$$

The gradient $\nabla f(t_1, ..., t_d)$ has an important geometric interpretation. Imagine that we are standing at the point $p(t_1, ..., t_d)$. Then the gradient points to the direction we should move in order to increase the value of function $f(x_1, ..., x_d)$ the *fastest*. Next, we will formalize the intuition.

Lemma 1. *Suppose that we decide to move from p towards the direction of a unit vector $\boldsymbol{u}$ by a distance Δs. Let q be the point we will reach, as shown below:*



We have:

$$\lim_{\Delta s \to 0} \frac{f(q) - f(p)}{\Delta s} = \Big(\nabla f(p)\Big) \cdot \boldsymbol{u}. \tag{1}$$

Proof. Suppose that $\boldsymbol{u} = [u_1, u_2, ..., u_d]$, and the coordinates of p are $(t_1, t_2, ..., t_d)$.

Let ℓ be the line that passes p and q. We know that we can represent any point on ℓ as $(x_1(s), x_2(s), ..., x_d(s))$, where for all $i \in [1, d]$:

$$x_i(s) = t_i + s \cdot u_i.$$

In particular, if $s = 0$, the above representation gives p, whereas if $s = \Delta s$, the above representation gives q.

Define $g(s) = f(x_1(s), ..., x_d(s))$. We can re-write the left hand side of (1) as:

$$\begin{aligned} \lim_{\Delta s \to 0} \frac{f(q) - f(p)}{\Delta s} &= \lim_{\Delta s \to 0} \frac{g(\Delta s) - g(0)}{\Delta s} \\ \text{(by def. of derivative)} &= g'(0). \end{aligned}$$

On the other hand, applying the chain rule[1], we know:

$$\begin{aligned} g'(s) &= \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}(x_1(s), ..., x_d(s)) \frac{dx_i}{ds} \\ &= \left[\frac{\partial f}{\partial x_1}(x_1(s), ..., x_d(s)), ..., \frac{\partial f}{\partial x_d}(x_1(s), ..., x_d(s))\right] \cdot \left[x_1'(s), ..., x_d'(s)\right] \\ &= (\nabla f(x_1(s), ..., x_d(s))) \cdot [u_1, ..., u_d] \\ &= (\nabla f(x_1(s), ..., x_d(s))) \cdot \boldsymbol{u}. \end{aligned}$$

Therefore, $g'(0) = (\nabla f(x_1(0), ..., x_d(0))) \cdot \boldsymbol{u} = (\nabla f(p)) \cdot \boldsymbol{u}$. □

As a corollary of the above lemma, we obtain

$$\lim_{\Delta s \to 0} \frac{f(q) - f(p)}{\Delta s} = \left|\nabla f(p)\right| |\boldsymbol{u}| \cos \gamma.$$

where γ is the angle between the directions of $\nabla f(p)$ and $\boldsymbol{u}$. Hence, the limit is maximized if $\gamma = 0$, namely, $\boldsymbol{u}$ has the same direction as $\nabla f(p)$.

It is worth mentioning that the limit on the left hand side of (1) is called the *directional derivative* in the direction of $\boldsymbol{u}$, and is denoted as $D_{\boldsymbol{u}} f$. Note that this is a function of p. In other words, $D_{\boldsymbol{u}} f(p)$ gives the directional derivative in the direction of $\boldsymbol{u}$ at point p.

[1] For example, suppose that $f(x, y) = xy$ with $x = \sin t$ and $y = t$. The chain rule states that $\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}$. To verify this, let us first compute $\frac{\mathrm{d}f}{\mathrm{d}t}$ directly: since $f = (\sin t) \cdot t$, we have $\frac{\mathrm{d}f}{\mathrm{d}t} = (\cos t)t + \sin t$. We can get the same using the chain rule: $\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} = y \cdot \cos t + x = (\cos t)t + \sin t$. In general, given a function $f(x_1, x_2, ..., x_d)$ where each x_i ($i \in [1, d]$) is a function of t, the chain rule states that $\frac{\mathrm{d}f}{\mathrm{d}t} = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial t}$