Directional Derivatives and Gradients

David S. Rosenberg

1 Directional Derivatives and First Order Approximations

- Let f be a differentiable function $f : \mathbf{R}^d \to \mathbf{R}$. We define the **directional derivative** of f at the point $x \in \mathbf{R}^d$ in the direction $u \in \mathbf{R}^d$ as
$$f'(x;u) = \lim_{h \to 0} \frac{f(x+hu) - f(x)}{h}.$$
Note[1] that $f'(x,u)$ is a scalar (i.e. an element of $\mathbf{R}$).

- This expression is easy to interpret if we drop the limit and replace equality with approximate equality. So, let's suppose that h is very small. Then we can write
$$\frac{f(x+hu) - f(x)}{h} \quad \approx \quad f'(x;u).$$
Rearranging, this implies that
$$f(x+hu) - f(x) \approx hf'(x;u).$$
So if we start at x and move to $x+hu$, then the value of f increases by approximately $hf'(x;u)$. This is called a **first order** approximation, because we used the first derivative information at x.

Exercise 1. Suppose we find a direction u for which the directional derivative is negative, i.e. $f'(x;u) < 0$. Show that for small enough step size $h > 0$, we'll always have $f(x+hu) < f(x)$. Such a direction u is called a **descent direction**. (Hint: This is straightforward from the llimit definition (ε's and δ's) of the directional derivative.

[1] Sometimes people require that u be a unit vector, but that is not necessary, and we do not assume that here.

- Rearranging again, we can write

$$f(x + hu) \approx f(x) + hf'(x; u).$$

Think of each side of the $\approx$ as a function of $h \in \mathbf{R}$. As we change h, we're evaluating f at different points on the line $\{x + hu \mid h \in \mathbf{R}\}$. Note that on the right hand side we have a linear function of h (or more precisely an affine function). The affine function $h \mapsto f(x) + hf'(x; u)$ is called a **first order approximation** to the function $h \mapsto f(x + hu)$ at the point x. Note that we are approximating the value of f at the location $x + hu$ using only information about f at the location x.

- Let $e_i = \Big(\underbrace{0, \ldots, 0}, 1, 0, \ldots, 0\Big)$, where the 1 is in the ith position. This vector is called the ith **coordinate vector** or the ith **standard basis vector**. The directional derivative in the direction e_i is called the ith **partial derivative**, which you'll see written in various ways:

$$\frac{\partial}{\partial x_i} f(x) = \partial_{x_i} f(x) = \partial_i f(x) = f'(x; e_i).$$

2 Gradients

- The gradient of f at x can be written as a column vector $\nabla f(x) \in \mathbf{R}^d$, where

$$\nabla f(x) = \begin{pmatrix} \partial_{x_1} f(x) \\ \vdots \\ \partial_{x_d} f(x) \end{pmatrix}.$$

- If f is differentiable, the gradient can be very useful! We can get the directional derivative in any direction u simply by taking the inner product between the gradient and the direction vector:

$$f'(x; u) = \nabla f(x)^T u.$$

(So once have the gradient, we don't need to take limits to find a directional derivative.)

- Thus we can also write the first order approximation as

$$f(x + hu) \approx f(x) + h\nabla f(x)^T u.$$

Exercise 2. Show that we can get the directions of steepest descent and steepest ascent from $\nabla f(x)$. More precisely, show that

$$\arg\min_{\|u\|_2=1} f'(x;u) = \frac{\nabla f(x)}{\|\nabla f(x)\|_2} \quad \text{and} \quad \arg\max_{\|u\|_2=1} f'(x;u) = -\frac{\nabla f(x)}{\|\nabla f(x)\|_2}.$$

(Hint: $|f'(x;u)| = \left|\nabla f(x)^T u\right|$ and then consider the Cauchy-Schwarz inequaltity.)