Quick Note on Gradients and Directional Derivatives

David S. Rosenberg

1 Directional Derivative and First Order Approximations

- Let f be a differentiable function $f : \mathbf{R}^d \to \mathbf{R}$. We the directional derivative of f at the point $x \in \mathbf{R}^d$ in the direction $v \in \mathbf{R}^d$ as

  $$\nabla_v f(x) = \lim_{\varepsilon \to 0} \frac{f(x + \varepsilon v) - f(x)}{\varepsilon}.$$

  Note that $\nabla_v f(x)$ 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 ε is very small. Then we can write

  $$\frac{f(x + \varepsilon v) - f(x)}{\varepsilon} \quad \approx \quad \nabla_v f(x).$$

  Rearranging, this implies that

  $$f(x + \varepsilon v) - f(x) \approx \varepsilon \nabla_v f(x).$$

  In words, we can interpret this as follows: If we start at x and move to $x + \varepsilon v$, then the value of f increases by approximately $\varepsilon \nabla_v f(x)$. This is called a **first order** approximation, because we used the first derivative information at x.

- Rearranging again, we can write

  $$f(x + \varepsilon v) \approx f(x) + \varepsilon \nabla_v f(x).$$

The expression $f(x) + \varepsilon\nabla_v f(x)$ is a **first order approximation** to $f(x + \varepsilon v)$. Note that we are approximating the value of f at the location $x + \varepsilon v$ using only information about f at the location x. This approximation becomes exact as $\varepsilon \to 0$.

2 Gradients

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

$$\nabla f(x) = \begin{pmatrix} \nabla_{e_1} f(x) \\ \vdots \\ \nabla e_d f(x) \end{pmatrix},$$

  and where $e_1, \ldots, e_d \in \mathbf{R}^d$ are the unit coordinate vectors: $e_1 = (1, 0, 0, \ldots, 0) \in \mathbf{R}^d$, and in general $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 is in the ith coordinate.

- One fact you should recall from calculus is that we can get any directional derivative from the gradient:

$$\nabla_v f(x) = \nabla f(x)^T v$$

- Thus we can also write

$$f(x + \varepsilon v) \approx f(x) + \varepsilon\nabla f(x)^T v$$