I. Chain Rule

1. *Theorem:* Let $f : A \subset \mathbb{R}^n \to \mathbb{R}^m$ and let $g : B \subset \mathbb{R}^m \to \mathbb{R}^p$ be such that f is differentiable at $\mathbf{x} \in A$ and g is differentiable at $f(\mathbf{x}) \in B$, then the composition function $g \circ f$ is differentiable at $\mathbf{x}$. Furthermore, the derivative matrix of the composition is given as a product of the derivative matrices of g and f as follows, $D(g \circ f)(\mathbf{x}) = D(g)(f(\mathbf{x})) \cdot D(f)(\mathbf{x})$. Proof: omitted.

2. Example 1: We have already seen chain rule for functions $\mathbb{R} \to \mathbb{R} \to \mathbb{R}^n$. Recall.

3. Example 2: Given $u = x^2 + y^2$, $v = 3x + 5xy^2$ and $w = 3\sin 2u + 3v + 5\cos 4v - 2u$, Find $\frac{\partial w}{\partial x}$ and $\frac{\partial w}{\partial x}$ at the point $(x, y) = (2, 2)$.

4. Example 3: Given $f(x, y, z) = 3xyz^2$ and $g(t) = (t, t^2, t^3)$. Find $D(g \circ f)$ at the point $(1, 1, 1)$.

II. Gradients and Directional Derivatives

1. Let us return to the case of $f : A \subset \mathbb{R}^2 \to \mathbb{R}$. Suppose that f is differentiable at $\mathbf{a} \in A$. In this case, the gradient of f at $\mathbf{a}$ denoted $\nabla f(\mathbf{a})$ is the vector in $\mathbb{R}^2$ given by $\left(\frac{\partial f}{\partial x}(\mathbf{a}), \frac{\partial f}{\partial y}(\mathbf{a})\right)$.

2. *Definition:* Let $f : A \subset \mathbb{R}^2 \to \mathbb{R}$ and let $(a_1, a_2) \in A$. Further assume that there is an $\epsilon > 0$ such that $B_\epsilon(\mathbf{a}) \subset A$. Let (b_1, b_2) be a unit vector in $\mathbb{R}^2$. Then the limit

$$\lim_{t \to 0} \frac{f((a_1, a_2) + t(b_1, b_2)) - f(a_1, a_2)}{t}$$

   when it exists is defined to be the directional derivative of f along the direction (b_1, b_2) at the point (a_1, a_2).

   (a) Since $\nabla f(\mathbf{a}) \cdot \mathbf{h} = Df(\mathbf{a})(\mathbf{h})$ for an error vector $\mathbf{h} \in \mathbb{R}^2$, we can write the LHS of this equation as $\|\nabla f(\mathbf{a})\| \|\mathbf{h}\| \cos\theta$, where θ is the angle between $\nabla f(\mathbf{a})$ and $\mathbf{h}$. If we let $\mathbf{h}$ vary over the unit vectors, the this LHS has a maximum value when $\theta = 0$, that is when $\mathbf{h}$ is parallel to $\nabla f(\mathbf{a})$. Likewise it is minimum when $\theta = \pi$, that is when $\mathbf{h}$ is anti-parallel to $\nabla f(\mathbf{a})$.

   (b) Significance: At $\mathbf{a}$, the vector ∇f indicates the direction in which the function changes the fastest.

   (c) For a given direction, *i.e.* a unit vector $\mathbf{h} \in \mathbb{R}^2$, the directional derivative of the function f along this direction at the point $\mathbf{a}$ is given by $\nabla f(\mathbf{a}) \cdot \mathbf{h}$, when the function f is differentiable at that point.

3. In the case of $f : A \subset \mathbb{R}^3 \to \mathbb{R}$, if f is differentiable at $\mathbf{a} \in A$ then the gradient of f at $\mathbf{a}$ denoted $\nabla f(\mathbf{a})$ is the vector in $\mathbb{R}^3$ given by $\left(\frac{\partial f}{\partial x}(\mathbf{a}), \frac{\partial f}{\partial y}(\mathbf{a}), \frac{\partial f}{\partial z}(\mathbf{a})\right)$

   (a) Like in the above case, At $\mathbf{a}$, the vector ∇f indicates the direction in which the function changes the fastest.

   (b) For a given direction, *i.e.* a unit vector $\mathbf{h} \in \mathbb{R}^3$, the directional derivative of the function f along this direction at the point $\mathbf{a}$ is defined to be $\nabla f(\mathbf{a}) \cdot \mathbf{h}$.

4. One can likewise define the gradient vector for a real-valued function of n variables as well.