Gastby induction week: (very) short differentiability review

September 2018

We consider a function $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$, and let $f_1, f_2, \dots, f_p : \mathbb{R}^n \to \mathbb{R}$ be the components (or coordinates) of f.

Definition 1 (Derivative ($n = 1$))**.** We say that $f : \mathbb{R} \to \mathbb{R}^p$ has a derivative at $a \in U$ if

$$\lim_{t \to 0} \frac{f(a+t) - f(a)}{t} \in \mathbb{R}^p$$

exist. We denote it by $f'(a)$, and we have $f' : \mathbb{R} \to \mathbb{R}^p$.

Definition 2 (Partial derivative)**.** We say that $f : \mathbb{R}^n \to \mathbb{R}^p$ has a partial derivative with respect to the j-th coordinate at $a \in U$ if

$$\lim_{t \to 0} \frac{f(a_1, \dots, a_j + t, \dots, a_n) - f(a)}{t} \in \mathbb{R}^p$$

exists. We denote it by $\partial_j f(a)$ and we have $\partial_j f : \mathbb{R}^n \to \mathbb{R}^p$.

Definition 3 (Jacobian matrix)**.** If f has partial derivatives with respect to all coordinates at $a \in U$, then we define the Jacobian matrix of f at a as

$$J_f(a) = (\partial_j f_i(a))_{i \le p, j \le n} \in \mathcal{M}_{p,n}(\mathbb{R}).$$

If $p = 1$, then we define the gradient of $f : \mathbb{R}^n \to \mathbb{R}$ at a by

$$\nabla f(a) = grad(f)(a) = J_f(a)^T \in \mathbb{R}^n$$

.

Definition 4 (Differentiability)**.** We say that $f : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable at $a \in U$ if $\exists v \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ a linear map such that

$$\lim_{h \to 0} \frac{\|f(a+h) - f(a) - v(h)\|}{\|h\|} = 0.$$

We can prove v to be unique and we denote it by $df(a) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$. We have $df : \mathbb{R}^n \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$.

Proposition 1. If $f \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$, then it is differentiable and $df(a) = f \ \forall a \in \mathbb{R}^n$.

If $\phi : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^p$ is bilinear, then ϕ is differentiable and $\forall (x, y) \in \mathbb{R}^d \times \mathbb{R}^m$, $\forall (h, k) \in \mathbb{R}^d \times \mathbb{R}^m$, $d\phi(x, y)(h, k) = \phi(x, k) + \phi(h, y)$.

If $n = 1$, then f is differentiable at a if and only if f has a derivative at a, and we have $\forall h \in \mathbb{R}$, $df(a)(h) = hf'(a)$.

Theorem 1. Let $(e_1, \dots, e_n)$ be the canonical basis of $\mathbb{R}^n$. If $f : \mathbb{R}^n \to \mathbb{R}^p$ is differentiable, then f has partial derivatives with respect to all coordinates, and $\forall a \in U, \forall i, \partial_i f(a) = df(a)(e_i)$.

In particular, $\forall h \in \mathbb{R}^n, df(a)(h) = \sum_i h_i \partial_i f(a) = J_f(a)h$. If $p = 1$, then $df(a)(h) = J_f(a)h = \nabla f(a)^T h = \langle \nabla f(a), h \rangle$.

Theorem 2 (Chain rule)**.** Let $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ and $g : V \subset \mathbb{R}^p \to \mathbb{R}^d$. We suppose that f is differentiable at $a \in U$, and g is differentiable at $f(a) \in V$. Then $g \circ f : U \subset \mathbb{R}^n \to \mathbb{R}^d$ is differentiable at a and

$$\forall h \in \mathbb{R}^n, d(g \circ f)(a)(h) = dg(f(a))(df(a)(h))$$

or equivalently if $g \circ f$ is differentiable on U

$$\forall a \in U, d(g \circ f)(a) = dg(f(a)) \circ df(a).$$