Infinitesimal Calculus 3

Lecture 14, Sunday December 4, 2022

Ari Feiglin

Proposition 14.0.1:

Suppose $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ and is defined in a neighborhood of x and differentiable at x. Specifically, $f(x+h) = f(x) + L(h) + \varepsilon(h)$ where L is a linear transform and ε is an ε function. L can be represented as a matrix $(A_1, \dots, A_n)$. Then

(1) f is continuous at x.

(2) For every $1 \leq k \leq n$, $\partial_{x_k} f$ exists and is equal to A_k.

(3) For every unit vector $u \in \mathbb{R}^n$, the directional derivative $D_u f(x)$ exists and is equal to $\nabla f(x) \cdot u$.

Proof:

(1) Notice that

$$\lim_{h\to 0} f(x+h) = f(x) + \lim_{h\to 0} L(h) + \lim_{h\to 0} \varepsilon(h)$$

And since L is a linear transformation on $\mathbb{R}^n$ so it is continuous (this can be shown directly since h_i converge to 0 so the sum of $A_i v_i$ converges to 0). And $\varepsilon(h)$ converges to 0, as explained previously, since $\varepsilon(h) = \|h\| \cdot \frac{\varepsilon}{\|h\|}$ which is the product of two limits which converge to 0. Thus $\lim_{h\to 0} f(x+h) = f(x)$ and it is therefore continuous.

(2) We will show this throught the definition of partial derivatives:

$$\partial_{x_k} f(x) = \lim_{\Delta x_k \to 0} \frac{f(x_1, \dots, x_k + \Delta x_k, \dots, x_n) - f(x_1, \dots, x_n)}{\Delta x_k}$$

If we define $h = \Delta x_k \cdot e_k$ then this is equal to

$$\lim_{\Delta x_k \to 0} \frac{f(x+h) - f(x)}{\|h\|} = \lim_{h\to 0} \frac{L(h) + \varepsilon(h)}{\|h\|} = \lim_{h\to} \frac{L(h)}{\|h\|}$$

Notice that $L(h) = A_k \cdot \Delta x_k = A_k \cdot \|h\|$, so this is equal to the limit of A_k, which is equal to A_k. So $\partial_{x_k} f(x) = A_k$ as required.

(3) Notice then that by above, L is represented by the gradient of f, ∇f, so $L(v) = \nabla f \cdot v$. By definition we know that

$$D_u f(x) = \lim_{t\to 0} \frac{f(x+tu) - f(x)}{t} = \lim_{t\to 0} \frac{L(tu) + \varepsilon(t)}{t} = \lim_{t\to} \frac{t \cdot \nabla f(x) \cdot u}{t} + \lim_{t\to 0} \frac{\varepsilon(tu)}{t} = \nabla f(x) \cdot u + \lim_{t\to 0} \frac{\varepsilon(tu)}{\pm\|tu\|} =$$
$$= \nabla f(x) \cdot u$$

Notice the $\pm$ before the $\|tu\|$ in the last transition. This is because $\|tu\| = |t| \cdot \|u\|$. But nonetheless, since the limit equals 0, multiplying it by ± 1 doesn't change it. So $D_u f(x) = \nabla f(x) \cdot u$ as required.

■

So by this above proposition, f is differentiable at x if and only if

$$f(x+h) = f(x) + \nabla f(x) \cdot h + \varepsilon(h)$$

Proposition 14.0.2:

If $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is defined and has partial derivatives in a neighborhood of $x \in \mathbb{R}^n$ and the partial derivatives are continuous at x, then f is differentiable at x.

The proof of this is identical to our earlier proof where $n = 2$.

Definition 14.0.3:

Suppose $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ is defined in a neighborhood of $x \in \mathbb{R}^n$ and is differentiable there. Then $f(x+h) = f(x) + L(h) + \varepsilon(h)$. We call the linear transform L f's **differential** at x and is denoted $df\big|_x$.

By our previous proposition, the differential of f and the gradient of f are related by the following equality:

$$df\big|_x(h) = \nabla f(x) \cdot h$$

Proposition 14.0.4:

Suppose $f, g\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ are defined in some neighborhood of $x \in \mathbb{R}^n$ and differentiable at x. Then for any $\alpha, \beta \in \mathbb{R}$:

(1) $d\alpha f + \beta g\big|_x = \alpha\, df\big|_x + \beta\, dg\big|_x$.

(2) $df \cdot g\big|_x = f(x) \cdot dg\big|_x + df\big|_x \cdot g(x)$.

(3) If $g(x) \neq 0$ then $d\frac{f}{g}\Big|_x = \frac{g(x)\, df\big|_x - dg\big|_x f(x)}{g^2(x)}$.

Proof:

(1) Since:

$$\begin{aligned}\alpha f(x+h) + \beta g(x+h) &= \alpha\big(f(x) + df\big|_x(h) + \varepsilon_1(h)\big) + \beta\big(g(x) + dg\big|_x(h) + \varepsilon_2(h)\big)\\ &= \alpha f(x) + \beta g(x) + \big(\alpha\, df\big|_x(h) + \beta\, dg\big|_x(h\big() + \alpha\varepsilon_1(h) + \beta\varepsilon_2(h)\end{aligned}$$

Since this is of the form $\alpha f(x+h) + \beta g(x+h) + L(h) + \varepsilon(h)$, we have that the differential is linear as required.

(2) We will do some algebraic manipulation:

$$\begin{aligned}(fg)(x+h) - (fg)(x) &= \big(f(x+h)g(x+h) - f(x+h)g(x)\big) + \big(f(x+h)g(x) - f(x)g(x)\big)\\ &= f(x+h)\big(g(x+h) - g(x)\big) + g(x)\big(f(x+h) - f(x)\big)\\ &= \big(f(x) + df\big|_x(h) + \varepsilon_1(h)\big)\big(dg\big|_x(h) + \varepsilon_2(h)\big) + g(x)\big(df\big|_x(h) + \varepsilon_3(h)\big)\\ &= f(x)\, dg\big|_x(h) + df\big|_x(h) \cdot g(x) + \big(df\big|_x(h) \cdot dg\big|_x(h) + f\varepsilon_2 + df\big|_x\varepsilon_2 + \varepsilon_1\, dg\big|_x + \varepsilon_1\varepsilon_2 + g\varepsilon_3\big)\end{aligned}$$

The rightmost side is an ε function since either every summand is the product of something (either a constant like $f(x)$ or an ε function) and another ε function, or it is $df\big|_x \cdot dg\big|_x$. For the first option it is obvious why these are all ε functions, and for the latter, since linear transforms in $\mathbb{R}^n$ are bounded:

$$df\big|_x(h) \cdot \frac{dg\big|_x(h)}{\|h\|} \leq M \cdot df\big|_x(h)$$

which converges to 0 so it is an ε function.

(3) This proof is computational and similar to the one above.

■

Notice that by the relation between the differential and gradient:

$$\nabla(fg) = f\nabla g + g\nabla f$$

Definition 14.0.5:

Suppose $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^k$ is a function where $f(x) = \big(f_1(x), \dots, f_k(x)\big)$ where $f_j\colon \mathbb{R}^n \longrightarrow \mathbb{R}$. Then f is **differentiable** if $f(x+h) = f(x) + L(h) + \varepsilon(h)$ where L is a linear transform $\mathbb{R}^n \longrightarrow \mathbb{R}^k$. The linear transform L is f's **differential** at h.

Proposition 14.0.6:

Suppose $f = (f_1, \dots, f_k)$ is defined around some neighborhood of $x \in \mathbb{R}^n$. Then f is differential at x if and only if f_j is differential at x for every $1 \le j \le k$. And in this case

$$df\big|_x = \left(df_1\big|_x, \dots, df_k\big|_x\right)^T$$

Proof:

Suppose f is differentiable at x, recall the definition of $\chi_i \colon (x_1, \dots, x_n) \mapsto x_i$. So there exists a linear transform $L \colon \mathbb{R}^n \longrightarrow \mathbb{R}^k$ and an ε function such that

$$f(x+h) = f(x) + L(h) + \varepsilon(h)$$

And so $f_j(x+h) = \chi\big(f(x+h)\big)$ so:

$$f_j(x+h) = f_j(x) + \chi_j(L(h)) + \chi_j(\varepsilon(h))$$

Since both χ_j and L are linear transforms, so is their composition. Since convergence in $\mathbb{R}^n$ is pointwise, if $\frac{|epsilon(h)}{\|h\|}$ converges to 0 so does $\chi_j\left(\frac{\varepsilon(h)}{\|h\|} = \frac{\chi_j(\varepsilon(h))}{\|h\|}\right)$. Therefore $\chi_j \circ \varepsilon$ is an ε function, so f_j is differentiable.

To show the converse, suppose $f_j(x+h) = f_j(x) + L_j(h) + \varepsilon_j(h)$ then $f(x+h) = f(x) + \big(L_1(h), \dots, L_k(h)\big)^T + \big(\varepsilon_1(h), \dots, \varepsilon_k(h)\big)^T$. Now, the vector $L = \big(L_1, \dots, L_k\big)^T$ represents a linear transform, since it is a vector of one dimensional linear transforms, which can be represented as a matrix. And the vector of ε functions is itself an epsilon function since if $\frac{\varepsilon_j(h)}{\|h\|}$ converges to 0 for each j, then since convergence is pointwise, $\frac{\varepsilon(h)}{\|h\|}$ converges to 0 as well for $\varepsilon = \big(\varepsilon_1, \dots, \varepsilon_k\big)$. So $f(x+h) = f(x) + L(h) + \varepsilon(h)$ as required. And notice that we showed $L = \big(L_1, \dots, L_k\big)^T$, that is

$$df\big|_x = \left(df_1\big|_x, \dots, df_k\big|_x\right)$$

as required.

■

Notice that the matrix described can be written as:

$$\begin{pmatrix} \nabla f_1 \\ \vdots \\ \nabla f_k \end{pmatrix}$$

Since the representation of the differential is the gradient. By definition this is equal to

$$\begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_k}{\partial x_1} & \cdots & \frac{\partial f_k}{\partial x_n} \end{pmatrix}$$

Definition 14.0.7:

If $f \colon \mathbb{R}^n \longrightarrow \mathbb{R}^k$ is differentiable, then we define the above matrix to be the **Jacobian matrix**, denoted $\dfrac{\partial(f_1, \dots, f_k)}{\partial(x_1, \dots, x_n)}$.