Differentiability of Functions

January 16, 2017

Please read §3.1–§3.4 of Dr. Runde's notes.

1. Definitions

1.1. Directional and partial derivatives

- When $N=1$, the derivative at x_0 for a function f is defined as the following number:

$$\lim_{h\longrightarrow 0, h\neq 0} \frac{f(x_0+h)-f(x_0)}{h}. \tag{1}$$

  The function f is said to be not differentiable if this number does not exist.

- For $N>1$, the natural generalization of this idea gives "directional derivative": Let $v\in\mathbb{R}^N$, $v\neq 0$, and $f\colon \mathbb{R}^N \mapsto \mathbb{R}^M$. We say f is differentiable in the direction v if the function $g(h):=f(x_0+v\,h)$ is differentiable at $h=0$.

  Remark 1. In the special cases $v=e_1, e_2, \ldots, e_N$, the directional derivatives become the "partial derivatives" of f, $\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N}$, at x_0. Thus for example we see that

$$\frac{\partial f}{\partial x_1}(x_{01}, \ldots, x_{0N}) = g_1'(0) \tag{2}$$

  where $g_1(h):=f(x_{01}+h, x_{02}, \ldots, x_{0N})$.

  Note. Make sure you know how to calculate partial derivatives!

- Directional derivative is very useful, however it is not satisfactory, as the following important property for $N=1$,

  If f is differentiable at x_0, then it is continuous at x_0,

  does not hold anymore.

  Example 2. Consider $f(x,y):=\begin{cases} \frac{x^2\,y}{x^4+y^2} & (x,y)\neq(0,0) \\ 0 & x=y=0 \end{cases}$. At $(0,0)$, all the directional derivatives of $f(x,y)$ exist, but $f(x,y)$ is not continuous at $(0,0)$.

  Exercise 1. A function that is directional differentiable at all orders but not continuous?

1.2. Total differentiability

- Due to Example 2 directional differentiability (and of course the special case partial differentiability) is not appropriate for theoretical study.

  Exercise 2. Consider $f(x,y):=\begin{cases} \frac{x\,y}{x^2+y^2} & (x,y)\neq(0,0) \\ 0 & x=y=0 \end{cases}$. Prove that $\frac{\partial f}{\partial x}(0,0)=\frac{\partial f}{\partial y}(0,0)=0$ but f is not continuous at $(0,0)$.

- Alternative interpretation of differentiability for $N=1$.

  Let $N=1$. We say f is differentiable at x_0 if there is an affine function $A(x)=a+b\,x$ such that

$$\lim_{x\longrightarrow x_0} \frac{|f(x)-A(x)|}{|x-x_0|}=0. \tag{3}$$

  Exercise 3. Prove or disprove: f is continuous at x_0 if there is a constant function $C(x):=c_0$ such that

$$\lim_{x\longrightarrow x_0} |f(x)-C(x)|=0. \tag{4}$$

Exercise 4. Prove that such $A(x)$, if it exists, is unique.

Exercise 5. Prove that if $A(x)$ exists, then $f(x)$ is differentiable and x_0 and furthermore $a = f(x_0) - x_0\, f'(x_0)$, $b = f'(x_0)$.

Remark 3. We see that it is more illustruating to write $A(x) = f(x_0) + T(x - x_0)$ where T is a linear function $T(h) := b\,h$.

- Total differentiability in $\mathbb{R}^N$.

  We say f is differentiable at x_0 if there is a linear function T such that

$$\lim_{x \longrightarrow x_0, x \neq x_0} \frac{\|f(x) - f(x_0) - T(x - x_0)\|}{\|x - x_0\|} = 0. \tag{5}$$

  The linear function T is called the "differential" of f at x_0. We denote it by $Df(x_0)$.

  Remark 4. Note that $Df(x_0)$, not Df, denotes the linear function from $\mathbb{R}^N$ to $\mathbb{R}^M$. Df is a function with domain $\mathbb{R}^N$ and range "the space of linear functions".

  Exercise 6. Prove that if f is differentiable at x_0 then it is continuous at x_0.

- Linear function.

  Recall that $T: \mathbb{R}^N \mapsto \mathbb{R}^M$ is linear if there holds

$$T(a\,x + b\,y) = a\,T(x) + b\,T(y) \tag{6}$$

  for all $a, b \in \mathbb{R}$.

- Matrix representations of linear functions.

  Let $T: \mathbb{R}^N \mapsto \mathbb{R}^M$ be linear. Then there is a unique $M \times N$ matrix $\begin{pmatrix} t_{11} & \cdots & t_{1N} \\ \vdots & \ddots & \vdots \\ t_{M1} & \cdots & t_{MN} \end{pmatrix}$ such that

$$T(x) = \begin{pmatrix} t_{11} & \cdots & t_{1N} \\ \vdots & \ddots & \vdots \\ t_{M1} & \cdots & t_{MN} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_N \end{pmatrix}. \tag{7}$$

  Exercise 7. Prove this claim.

- Matrix representation of $Df(x_0)$.

  As $Df(x_0)$ is a linear function, it has a matrix representation, we denote it by $J_f(x_0)$, and call it the Jacobian matrix of f at x_0.

  The calculation of $J_f(x_0)$ is easy thanks to the following result:

  THEOREM 5. *Let $f: \mathbb{R}^N \mapsto \mathbb{R}^M$ be differentiable at x_0. Then it is directionally differentiable at x_0 in all directions. In particular all of its partial derivatives exist. Furthermore we have*

$$J_f(x_0) = \begin{pmatrix} t_{11} & \cdots & t_{1N} \\ \vdots & \ddots & \vdots \\ t_{M1} & \cdots & t_{MN} \end{pmatrix} \tag{8}$$

  with $t_{ij} = \frac{\partial f_i}{\partial x_j}$.

2. Further properties.

2.1. Tangent spaces.

- It is helpful to understand $Df(x_0)$ not as a function with the same domain as f. This understanding will be crucial in differential geometry, the application of calculus to geometry.
- Intuitively, let $f: \mathbb{R}^N \mapsto \mathbb{R}^M$. We can think of $f(x)$ as an "avatar" of the point x. When x moves in $\mathbb{R}^N$, its avatar moves in $\mathbb{R}^M$ according to the rule f.

- In this setting, the linear function $Df(x_0)$, is the relation between the velocity of x at x_0 and the velocity of its avatar at $f(x_0)$. More specifically, if the velocity of x at x_0 is $v = \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}$, then its avatar is moving with velocity

$$Df(x_0)(v) = J_f(x_0) \cdot v = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \cdots & \frac{\partial f_1}{\partial x_N}(x_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M(x_0)}{\partial x_1} & \cdots & \frac{\partial f_M}{\partial x_N}(x_0) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_N \end{pmatrix}. \tag{9}$$

- Thus rigorously speaking, the domain of $Df(x_0)$ is "the space of velocities at x_0", which is different from the domain of f, the "space of positions".
- This distinction becomes important in differential geometry, where the domain of f is not $\mathbb{R}^N$, but say some curved surfaces. In that case the "space of positions" is the curved surface, but the space of velocities at x_0 is not curved, it is the tangent plane of the surface at x_0.

2.2. Chain rule.

- If we use the "avatar" intuition, it is easy to understand the chain rule.
- So we have $x \in \mathbb{R}^N$, its "avatar" $g(x) \in \mathbb{R}^M$ and an "avatar of avatar" $f(g(x)) \in \mathbb{R}^K$. Thus at every $x_0 \in \mathbb{R}^N$, $Dg(x_0)$ relates the velocity of x to the velocity of the avatar $g(x)$, while at $y_0 \in \mathbb{R}^M$, $Df(y_0)$ relates the velocity of the avatar at y_0 to the velocity of the "avatar of avatar" at $f(y_0)$. Consequently, the relation between the velocity of x_0 to that of its "avatar of avatar" is given by the composite function:

$$v \mapsto Df(g(x_0))(Dg(x_0)(v)). \tag{10}$$

Exercise 8. Prove that the matrix of the composition of linear functions is the product of the matrix representations of these functions.

- Consequently we have

$$J_{f \circ g}(x_0) = J_f(y_0) \cdot J_g(x_0) \tag{11}$$

where $y_0 = g(x_0)$.

Exercise 9. Let $f: \mathbb{R}^M \mapsto \mathbb{R}$ and $g: \mathbb{R}^N \mapsto \mathbb{R}^M$, use (11) to prove the chain rule:

$$\frac{\partial (f \circ g)}{\partial x_i}(x_0) = \sum_{k=1}^{M} \frac{\partial f}{\partial y_k}(g(x_0)) \cdot \frac{\partial g_k}{\partial x_i}(x_0). \tag{12}$$