DERIVATIVES OF MULTIVARIATE FUNCTIONS AND THE CONTINUITY OF MIXED PARTIAL DERIVATIVES

Math 267F, Autumn 1995
Frank Sottile
18 September, 1995

1. The Derivative

We will briefly review the ordinary derivative of one variable calculus, then try to generalize this to the case of multivariate functions.

1.1. **The single variable derivative.** Let $f : \mathbb{R} \to \mathbb{R}$ be a function and $a \in \mathbb{R}$ be a real number. Recall that f is differentiable at a with derivative $Df(a)$ if

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

This may be rewritten in the following manner:

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

Or, replacing h by its absolute value, $|h|$, in the denominator,

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

Note that $Df(a)$ is the unique number (if such a number exists) which satisfies $(*)$.

1.2. **Multivariate functions: The case $f : \mathbb{R} \to \mathbb{R}^n$.** Now suppose that $f : \mathbb{R} \to \mathbb{R}^n$. Then f is an n-tuple of functions $(f_1, \ldots, f_n)$ where each $f_j : \mathbb{R} \to \mathbb{R}$. It is reasonable to define the derivative of f at $a \in \mathbb{R}$ to be the n-tuple $Df(a) = (Df_1(a), \ldots, Df_n(a))$. This satisfies the analog of $(*)$:

$$\lim_{h \to 0} \frac{f(a+h) - f(a) - hDf(a)}{|h|} = 0, \qquad (**)$$

as a limit of vectors in $\mathbb{R}^n$. Here, if $g : \mathbb{R} \to \mathbb{R}^n$ is a vector valued function and $L \in \mathbb{R}^n$, then $\lim_{t \to 0} g(t) = L$ means that for every $1 \leq j \leq n$ we have the limit of ordinary functions $\lim_{t \to 0} g_j(t) = L_j$, where g_j is the j^{th} component of g, and L_j the j^{th} component of L.

1.3. **Multivariate functions: The case $f : \mathbb{R}^n \to \mathbb{R}$.** Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$. What should be the derivative of f at a? We will reason by analogy with $(**)$. In $(**)$, $f(a)$ is a vector, $|h|$ is the magnitude of h, and $h \mapsto h \cdot Df(a)$ defines a linear map is a linear map from $\mathbb{R} \to \mathbb{R}^n$.

For $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, we let $||x||$ denote the length of the vector x, that is, $(x_1^2 + \cdots + x_n^2)^{\frac{1}{2}}$. We use $||\cdot||$ instead of $|\cdot|$ to differentiate between the multivariate and single variable situations.

Definition: Let $f : \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}^n$. Then the derivative of f at a is the linear map $Df(a) : \mathbb{R}^n \to \mathbb{R}$ such that

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

Here for a function $g : \mathbb{R}^n \to \mathbb{R}$, $\lim\limits_{h \to 0} g(h) = 0$ means that $\forall \epsilon > 0\ \exists \delta > 0$ such that if $||h|| < \delta$, then $|g(h)| < \epsilon$.

2. Partial Derivatives

A linear map $M : \mathbb{R}^n \to \mathbb{R}$ determines a vector $(m_1, \ldots, m_n) \in \mathbb{R}^n$ such that if $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then $M(x) = m_1 x_1 + \cdots + m_n x_n$. Likewise a vector in $\mathbb{R}^n$ determines a (unique) linear map. We call the components of this vector the *components* of the linear map.

2.1. **Definition.** Let $f : \mathbb{R}^n \to \mathbb{R}$. The components of the map $Df(a)$ are called the *partial derivatives* of f at a, and the j^{th} component is written $\dfrac{\partial f}{\partial x_j}(a)$ or $D_j f(a)$.

We show how to compute these partial derivatives. Let e_j be the vector with a 1 in the j^{th} place and 0's elsewhere, $(0, \ldots, 0, \overset{j}{1}, 0, \ldots, 0)$.

2.2. **Proposition.** $\dfrac{\partial f}{\partial x_j}(a) = \lim\limits_{t \to 0} \dfrac{f(a + te_j) - f(a)}{t}$.

In simple terms, we compute $D_j f(a)$ by treating the variables $x_1, \ldots, \widehat{x_j}, \ldots, x_n$[1] as constants and 'taking the derivative with respect to the x_j variable only'.

Proof: We know that

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

Now, $Df(a)(te_j) = tD_j f(a)$ and $||te_j|| = |t|$, so

$$\lim_{h \to 0} \frac{f(a + te_j) - f(a) - tD_j f(a)}{|t|} = 0. \qquad (*)$$

[1]We often the use the shorthand $x_1, \ldots, \widehat{x_j}, \ldots, x_n$ to indicate that x_j is omitted from the list: $x_1, \ldots, x_{j-1},\ x_{j+1}, \ldots, x_n$

Let $g : \mathbb{R} \to \mathbb{R}$ be defined by $g(t) = f(a + te_j)$. Then by $(*)$, $Dg(0) = D_j f(a)$. But

$$Dg(0) = \lim_{t\to 0} \frac{g(t) - g(0)}{t} = \lim_{t\to 0} \frac{f(a + te_j) - f(a)}{t}. \quad \blacksquare$$

3. The Continuity of Mixed Partial Derivatives

3.1. **Theorem.** *Let $\mathcal{D}$ be an open subset of $\mathbb{R}^2$ and let $f : \mathcal{D} \to \mathbb{R}$ be a differentiable function such that the mixed partial derivatives D_1D_2f and D_2D_1f exists on $\mathcal{D}$, and let $(a, b) \in \mathcal{D}$ be a point where they are both continuous. Then $D_1D_2f(a, b) = D_2D_1f(a, b)$.*

This is usually expressed as $\dfrac{\partial^2 f}{\partial x \partial y} = \dfrac{\partial^2 f}{\partial y \partial x}$.

Proof: Our plan is to show that for all $\epsilon > 0$, $|D_1D_2f(a, b) - D_2D_1f(a, b)| < \epsilon$.

Let $\epsilon > 0$. Because both D_1D_2f and D_2D_1f are continuous at (a, b), there exists $\delta > 0$ such that if $||(a, b) - (x, y)|| < \delta$, then we have

$$\begin{aligned} |D_1D_2f(a, b) - D_1D_2f(x, y)| &< \epsilon/2 \\ |D_2D_1f(a, b) - D_2D_1f(x, y)| &< \epsilon/2 \end{aligned}$$

Let $0 < h, k \le \delta/\sqrt{2}$. Then if $a < c < a + h$, and $b < d < b + k$, we have $||(a, b) - (c, d)|| < \delta$.

Define

$$\begin{aligned} (*) &= f(a + h, b + k) - f(a + h, b) - f(a, b + h) + f(a, b) \\ G(y) &= f(a + h, y) - f(a, y) \\ F(x) &= f(x, b + k) - f(x, b) \end{aligned}$$

Then

$$F(a + h) - F(a) = (*) = G(b + k) - G(b).$$

F is continuous on $[a, a+h]$ and differentiable on $(a, a+h)$ and G is continuous on $[b, b+k]$ and differentiable on $(b, b + k)$. By Lagrange's Theorem (Mean Value Theorem), there exists c, d' with $a < c < a + h$ and $b < d' < b + k$ such that

$$hDF(c) = (*) = kDG(d').$$

But

$$\begin{aligned} DF(c) &= D_1f(c, b + k) - D_1f(c, b) \\ DG(d') &= D_2f(a + h, d') - D_2f(a, d') \end{aligned}$$

Applying Lagrange's Theorem once again, we see that there exists c', d with $a < c' < a + h$ and $b < d < b + k$ such that

$$DF(c) = kD_2D_1f(c, d) \quad \text{and} \quad DG(d') = hD_1D_2f(c', d')$$

Thus

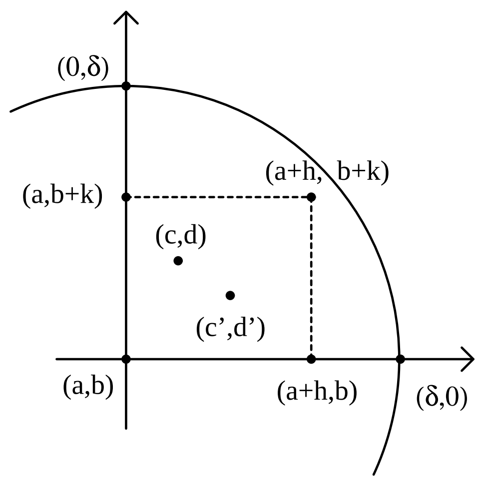
$$hkD_2D_1f(c,d) = (*) = khD_1D_2f(c',d')$$

and so $D_2D_1f(c,d) = D_1D_2f(c',d')$.

We use this to estimate $|D_1D_2f(a,b) - D_2D_1f(a,b)|$.

$$\begin{aligned}|D_1D_2f(a,b) - D_2D_1f(a,b)| \leq \\ &|D_1D_2f(a,b) - D_1D_2f(c',d')| + |D_2D_1f(c,d) - D_2D_1f(a,b)| \\ &< \epsilon/2 + \epsilon/2 = \epsilon.\end{aligned}$$ ∎

The figure below should help to visualize the situation of the proof.



4. Exercises

These are to be contemplated.

(1) **Multivariate Derivative** Give a reasonable definition for the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}^m$.

(2) **Multivariate Chain Rule** If $f : \mathbb{R}^n \to \mathbb{R}^m$ and $g : \mathbb{R}^m \to \mathbb{R}^p$, we have the composition $g \circ f : \mathbb{R}^n \to \mathbb{R}^p$. What is $D(g \circ f)$?

(3) Define $f : \mathbb{R}^2 \to \mathbb{R}$ as follows: For $(x,y) \in \mathbb{R}^2 - \{(0,0)\}$, set $f(x,y) = \dfrac{xy(x^2-y^2)}{x^2+y^2}$, and $f(0,0) = 0$. Show that f is everywhere continuous, that D_1f, D_2f, $D_{12}f$, and $D_{21}f$ exist everywhere, but that $D_{12}f(0,0) \neq D_{21}f(0,0)$.

Reconcile this with the Theorem of §3.1.