March 4, 2019

LECTURE 11: DIFFERENTIALS AND TAYLOR SERIES.

110.211 HONORS MULTIVARIABLE CALCULUS
PROFESSOR RICHARD BROWN

Synopsis. Herein, we give a brief interpretation of the differential of a function. There are many interpretations of a function's differential, but we only deal with one currently. Then we delve into an even briefer description of the Taylor Series of a real-valued function on $\mathbb{R}^n$. The details, for now, are not important. But the relationship to the counterparts of both of these concepts to single variable calculus is quite important.

Helpful Documents. Mathematica: `TaylorPolynomials`.

The differential of a function. Recall that for a variable x, a small change in x is denoted $\Delta x = (x+h) - x = h$, where h is a number near 0. As the value of h tends to 0, Δx also vanishes. But we can mark the vanishing of Δx via what is called an *infinitesimal change* in x, and denote it dx, so that

$$\Delta x \xrightarrow{h\to 0} dx.$$

Really, this has meaning almost exclusively in the context of how other quantities change that depend on x or when compared to x. The quantity dx is called the *differential* of x.

Now let $f : X \subset \mathbb{R} \to \mathbb{R}$ be a differentiable function, and $a \in X$. For the graph $y = f(x)$, the quantity

$$\Delta y = \Delta f = f(x + \Delta x) - f(x)$$

represents a small change in y, as it depends on Δx, the small change in x. As $h \to 0$, of course, Δf also goes to 0. But measuring *how* Δf goes to zero is important in calculus. Hence we mark teh infinitesimal change in y or f by its differential: $df = dy$. Studying just how the dependent variable y is changing as one varies x is vitally important in the study of functional relationships between entities, and is the motivation behind the Liebniz notation in calculus $\frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x)$ representing the derivative of $f(x)$ with respect to x:

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} = \lim_{\Delta x \to 0} \frac{f\left(x + \Delta x\right) - f(x)}{\Delta x} = \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x} = \frac{df}{dx}$$

since $\Delta x = (x+h) - x = h$. To even be able to discuss ideas that involve passing to a limit, one needs to be able to discuss quantities that are *infinitesimally* close to 0 or close to each other. One can say that an infinitesimally small positive number represents a positive number closer to 0 than any real positive number.

We note here that, as an alternate definition, one can call the quantities dx, and dy actual new variables, whose relationship is tied to the relationship between y and x, namely $y = f(x)$. This alternate definition provides a much more concrete foundation for which to use these quantities, but structurally does not change their meaning. We will visit this more concrete notion of a differential later when we discuss differential forms.

The quantity df (the differential of f), represents an infinitesimal change in f given an infinitesimal change in its independent variable x: at $x = a$, we have

$$df(a) = f'(a)dx, \qquad \text{or} \qquad \frac{df}{dx}(a) = f'(a)$$

to reflect the idea that this differential will change as we vary the point $x = a$. More generally, $df = f'(x)\,dx$.

Some notes:

- This will make more sense later, when we discuss differential forms, but df, the differential of f, is an example of a *differential* 1*-form*.
- This concept embodies the Substitution Rule (the Anti-Chain Rule) in single variable calculus:

$$\int_a^b f\left(g(x)\right) g'(x)\, dx \xrightarrow[du=g'(x)\,dx]{u=g(x)} \int_{g(a)}^{g(b)} f(u)\ du.$$

Indeed, let f be a function of u, with $u = \alpha$ a point in its domain$f$, and F an antiderivative of f, so that $F'(u) = f(u)$. Then

$$dF(\alpha) = F'(\alpha)\ du = f(\alpha)\ du = \left(f(u)\Big|_{u=\alpha} \right) du.$$

If $u = g(x)$ is a function of x, then du, the differential of u, is related to dx, and $du = g'(x)\,dx$. But also, f and hence F are functions of x, via composition: $f(u) = f\left(g(x)\right)$ and $F(u) = F\left(g(x)\right)$. Thus, their differentials also vary with respect to x,

$$\begin{aligned} f(\alpha)\ du = dF(\alpha) = F'(\alpha)\ du &= \left(F'(u)\Big|_{u=\alpha} \right) du \\ &= \left(F'(u)\Big|_{u=g(a)=\alpha} \right) \left(g'(x)\Big|_{x=a} \right) dx \\ &= \left(f(u)\Big|_{u=g(a)} \right) \left(g'(x)\Big|_{x=a} \right) dx \\ &= f'\left(g(a)\right) g'(a)\, dx. \end{aligned}$$

Hence, we are left with, on their appropriate domains,

$$f(u)\ du = f'\left(g(x)\right) g'(x)\, dx.$$

Finally, recall that integration is just a form of infinitesimal addition. Now does the form of the Substitution Rule make more sense?

In many variables, let $f : X \subset \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and $\mathbf{a} \in X$. df is the sum of the *partial differentials* (differentials in the coordinate directions), $\frac{\partial f}{\partial x_i}\ dx_i$, and

$$df = \frac{\partial f}{\partial x_1}\ dx_1 + \ldots + \frac{\partial f}{\partial x_n}\ dx_n = \sum_{n=1}^{n} \frac{\partial f}{\partial x_i}\ dx_i = Df(\mathbf{a})d\mathbf{x} \tag{11.1}$$

represents the *total differential* of f. This quantity represents an infinitesimal change in f in terms of the infinitesimal changes in its coordinate directions dx_i. The use of the vector term $d\mathbf{x} = \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}$ will make more sense later on in the course.

As a function, $\Delta f = f(\mathbf{a} + \Delta\mathbf{x}) - f(\mathbf{a}) = f(\mathbf{a} + \mathbf{h}) - f(\mathbf{a})$, where $\Delta\mathbf{x} = \mathbf{h}$ is a vector of small changes in each of the coordinate directions. Written out, Δf will contain many terms which are not linear in $\Delta\mathbf{x}$. As $\Delta\mathbf{x}$ tends to 0, any terms which contain products of the various small changes in the coordinate directions will tend toward 0 much faster, so that only the linear parts of these terms will contribute to the limit (the higher-degree terms will die off quickly, leaving only the linear terms). One can then see directly how the differential of a function operates:

Example 11.1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by $f(x, y) = x^2 + xy - x - y + \sin x$. Here $\Delta\mathbf{x} = (\Delta x, \Delta y))^T$, and

$$\begin{aligned}
\Delta f(\pi, 0) &= f\left((\pi, 0)^T + (\Delta x, \Delta y)^T\right) - f(\pi, 0) \\
&= (\pi + \Delta x)^2 + (\pi + \Delta x)(\Delta y) - (\pi + \Delta x) - \Delta y + \sin(\pi + \Delta x) - \pi^2 + \pi \\
&= \pi^2 + 2\pi\Delta x + (\Delta x)^2 + \pi\Delta y + \Delta x\Delta y - \pi - \Delta x - \Delta y - \sin(\Delta x) - \pi^2 + \pi.
\end{aligned}$$

Notice here that all of the terms not containing a Δx or a Δy cancel out. Recall also that for very small values of Δx, the function $\sin(\Delta x) \approx \Delta x$. This is called a first-order approximation of the sine function near $x = 0$, and reflects the idea that the sine function has a Taylor expansion at $x = 0$ containing a linear term with coefficient 1 (its first Taylor polynomial is $T_1(x) = x$). Likewise, for very small values of Δx and Δy, all of the other higher-order terms vanish double fast, leaving only the linear terms:

$$\Delta f(\pi, 0) = (2\pi - 1)\Delta x - \Delta x + (\pi - 1)\Delta y = (2\pi - 2)\Delta x + (\pi - 1)\Delta y.$$

Passing to the infinitesimals, we get $\Delta f \longrightarrow df$, and $\Delta\mathbf{x} \longrightarrow d\mathbf{x} = \begin{bmatrix} dx \\ dy \end{bmatrix}$, and we get

$$df(\pi, 0) = (2\pi - 2)\, dx + (\pi - 1)\, dy.$$

Now compare this to the direct computation, using Equation 11.1 above. Here

$$\frac{\partial f}{\partial x}(\pi, 0) = (2x - y - 1 + \cos x)\Big|_{\substack{x=\pi \\ y=0}} = (2\pi - 2)$$

and

$$\frac{\partial f}{\partial y}(\pi, 0) = (x - 1)\Big|_{\substack{x=\pi \\ y=0}} = (\pi - 1)$$

so that

$$df(\pi, 0) = Df(\pi, 0)\, d\mathbf{x} = \begin{bmatrix} 2\pi - 2 & \pi - 1 \end{bmatrix} \begin{bmatrix} dx \\ dy \end{bmatrix} = (2\pi - 2)\, dx + (\pi - 1)\, dy.$$

The result is the same.

And finally, going back to the notion of dx and dy being actual coordinates, tied together via $y = f(x)$ so that $dy = f'(x)\,dx$, we can extend this notion to the multidimensional case. Here, we can view each of the dx_i as an actual coordinate on the local linearization of f at the point $\mathbf{x} = \mathbf{a}$. Thus, the set $\{\, dx_1, \ldots, dx_n\}$ become a set of coordinates on each tangent space to the domain $X \subset \mathbb{R}^n$, and $df(\mathbf{a})$ becomes a vector measuring just how f is changing infinitesimally. Again,we will explore this notion in detail at the end of the course.

The Taylor series. Recall that the Taylor series of a C^∞-function $f : I \subset \mathbb{R} \to \mathbb{R}$ at a point $a \in I$ is

$$f(x) = \sum_{i=0}^{\infty} \frac{f^{(i)}(a)}{i\,!}(x-a)^i,$$

and is defined on the largest interval where the series converges. Here, one may also truncate this series to obtain the mth Taylor polynomial

$$T_m(x) = \sum_{i=0}^{m} \frac{f^{(i)}(a)}{i\,!}(x-a)^i.$$

The mth Taylor polynomial is considered the "best" mth-degree polynomial that approximates $f(x)$ near $x = a$, and we define the term "best" to mean that all of the derivatives of f and T_m are the same up to and including the mth derivative. So, for $i = 0, 1, \ldots, m$,

$$\frac{d^i}{dx^i} T_m(a) = f^{(i)}(a).$$

Now let $g : X \subset \mathbb{R}^n \to \mathbb{R}$ also be C^∞. We may ask very similar questions, like: What is the best mth degree polynomial (in the variables defining X) that approximates g near $\mathbf{x} = \mathbf{a}$. Again, the criteria for "best" will be the one that matches g at $\mathbf{a}$ for all (partial) derivatives up to and including the order-m ones.

Obviously, the best 0-degree polynomial to approximate $g(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$ is the one whose function value is $g(\mathbf{a})$, so

$$T_0(\mathbf{x}) = g(\mathbf{a}).$$

And we have already calculated the best first-degree polynomial, where the derivative matrix of g played a vital role:

$$T_1(\mathbf{x}) = g(\mathbf{a}) + Dg(\mathbf{a})(\mathbf{x} - \mathbf{a}).$$

Perhaps a better way to write this is to appeal to the individual variables explicitly, so

$$\begin{aligned} T_1(\mathbf{x}) &= g(\mathbf{a}) + Dg(\mathbf{a})(\mathbf{x} - \mathbf{a}) \\ &= g(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) \\ &= g(\mathbf{a}) + \underbrace{g_{x_1}(\mathbf{a})(x_1 - a_1) + g_{x_2}(\mathbf{a})(x_2 - a_2) + \ldots + g_{x_n}(\mathbf{a})(x_n - a_n)}_{\text{all first partials}}. \end{aligned}$$

Now it should be straightforward to see that not only does $T_1(\mathbf{a}) = g(\mathbf{a})$, but

$$\frac{\partial T_1}{\partial x_i}(\mathbf{a}) = \frac{\partial g}{\partial x_i}(\mathbf{a}),$$

for all $i = 1, \ldots, n$.

So follow the pattern. What would be the "best" degree-2 polynomial to approximate $g(\mathbf{x})$ at $\mathbf{x} = \mathbf{a}$? Of course, one such that $T_2(\mathbf{a}) = g(\mathbf{a})$, and for all $i, j = 1, \ldots, n$, we have

$$\frac{\partial T_2}{\partial x_i}(\mathbf{a}) = \frac{\partial g}{\partial x_i}(\mathbf{a}), \quad \text{and} \quad \frac{\partial^2 T_2}{\partial x_j \partial x_i}(\mathbf{a}) = \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a}).$$

It is apparent by reverse engineering that the only degree-2 polynomial that would work is

$$T_2(\mathbf{x}) = g(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j).$$

A big question to answer here is: Why the $\frac{1}{2}$ coefficient? Think about this.

There is a much better way to view the form of T_2:

Definition 11.2. Given $f : X \subset \mathbb{R}^n \to \mathbb{R}$ a C^2 function, the $n \times n$ matrix whose ijth entry is $\frac{\partial^2 f}{\partial x_j \partial x_i}$,

$$Hf = \begin{bmatrix} f_{x_1 x_1} & \cdots & f_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ f_{x_n x_1} & \cdots & f_{x_n x_n} \end{bmatrix}$$

is called the *Hessian* of f.

Now, denote by $\mathbf{h} = \mathbf{x} - \mathbf{a}$, so that $h_i = x_i - a_i$. Then we can write

$$\begin{aligned} T_2(\mathbf{x}) &= g(\mathbf{a}) + \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j) \\ &= g(\mathbf{a}) + \begin{bmatrix} g_{x_1}(\mathbf{a}) & \cdots & g_{x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} + \frac{1}{2} \begin{bmatrix} h_1 & \cdots & h_n \end{bmatrix} \begin{bmatrix} g_{x_1 x_1}(\mathbf{a}) & \cdots & g_{x_1 x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ g_{x_n x_1}(\mathbf{a}) & \cdots & g_{x_n x_n}(\mathbf{a}) \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ h_n \end{bmatrix} \\ &= g(\mathbf{a}) + Dg(\mathbf{a})\mathbf{h} + \frac{1}{2}\mathbf{h}^T Hf \mathbf{h}. \end{aligned}$$

So what would the third Taylor polynomial look like? Generalize in the obvious fashion, and get

$$\begin{aligned} T_3(\mathbf{x}) = g(\mathbf{a}) &+ \sum_{i=1}^{n} \frac{\partial g}{\partial x_i}(\mathbf{a})(x_i - a_i) + \frac{1}{2} \sum_{i,j=1}^{n} \frac{\partial^2 g}{\partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j) \\ &+ \frac{1}{6} \sum_{i,j,k=1}^{n} \frac{\partial^3 g}{\partial x_k \partial x_j \partial x_i}(\mathbf{a})(x_i - a_i)(x_j - a_j)(x_k - a_k), \end{aligned}$$

of course. And the polynomial $T_\ell(\mathbf{x})$, for natural number $\ell > 3$?

You may notice that I did not write $T_3(\mathbf{x})$ in a more elegant fashion, using some three dimensional verion of the derivative matrix or the Hessian matrix. It get difficult now since we would be creating and using objects that are higher dimensional arrays (All $8 = 2^3$ of the third-order partials of $f : \mathbb{R}^2 \to \mathbb{R}$ would be arranged into a three dimensional array. These objects do exist and are manifestations of what are called tensors. Getting a handle on the notation and working with these objects would involve a bit more time than we can devote to it at the moment. So we rely simply on the summation notation, and basically stop here.

Exercise 1. Devise a mathematical notation that would provide an array-based version of the third-order terms in $T_3(\mathbf{x})$.