CLARK UNIVERSITY 1887

Differentials and Taylor polynomials
Math 131 Multivariate Calculus
D Joyce, Spring 2014

Linear approximations. Let's start with a differentiable function, $f : \mathbf{R} \to \mathbf{R}$, of one variable. You know that the best linear approximation for f at the point $x = a$ is the linear polynomial

$$p_1(x) = f(a) + f'(a)\,(x - a),$$

and its graph $y = p_1(x)$ is the straight line tangent to the graph $y = f(x)$ of f at the point $(x, y) = (a, f(a))$. The definition of the derivative justifies the statement

$$\lim_{x \to a} \frac{f(x) - p_1(x)}{x - a} = 0,$$

and that statement is the algebraic formulation of the tangency of the line.

We've already generalized this to a differentiable function $f : \mathbf{R}^2 \to \mathbf{R}$ of two variables. We've seen that the linear polynomial

$$\begin{aligned} p_1(x, y) &= f(a, b) + Df(a, b)\,((x, y) - (a, b)) \\ &= f(a, b) + \frac{\partial f}{\partial x}(a, b)\,(x - a) + \frac{\partial f}{\partial y}(a, b)\,(y - b) \end{aligned}$$

as a graph which is the plane tangent to the graph $z = f(x, y)$ at the point $(x, y, z) = (a, b, f(a, b))$. More generally, for a function $f : \mathbf{R}^n \to \mathbf{R}$ of several variables, the linear polynomial

$$p_1(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})\,(\mathbf{x} - \mathbf{a})$$

has a graph which is the hyperplane tangent to the graph $z = f(\mathbf{x})$ at the point $(\mathbf{x}, z) = (\mathbf{a}, f(\mathbf{a}))$.

Differentials. Now, let $d\mathbf{x}$ be a differential of $\mathbf{x}$. For Leibniz that would have been an infinitesimal, but for us we can take $d\mathbf{x}$ to be a vector which we'll let approach $\mathbf{0}$. If we take the last paragraph and (1) replace $\mathbf{x} - \mathbf{a}$ by $d\mathbf{x}$, (2) replace $\mathbf{a}$ by $\mathbf{x}$, and (3) let df, which is called the *total differential* of f, be the difference $f(\mathbf{x} + d\mathbf{x}) - f(\mathbf{x})$, then we see it suggests

$$df = Df(\mathbf{x})\,d\mathbf{x},$$

which looks like

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_n} dx_n$$

when written out is full. The df describes the change in f when the various dx_i's give the changes in the x_i's. For Leibniz, the equation was an exact equality of infinitesimals. We can interpret it in three ways. First, the right hand side is the exact value of the linear approximating polynomial $p_1(\mathbf{x})$. Second, it's an approximation of the change in $f(\mathbf{x})$. Third, it's just a reformulation of the chain rule

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \frac{\partial f}{\partial x_2}\frac{dx_2}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt}.$$

Taylor polynomials for functions of one variable. The function $p_1(\mathbf{x})$ described above is the first-order Taylor polynomial for a function of several variables. Before looking at higher-order Taylor polynomials for functions of several variables, let's recall the higher-order Taylor polynomials for functions of one variable.

Let $f : \mathbf{R} \to \mathbf{R}$ be a function of one variable with derivatives of whatever order we need. The first-order Taylor polynomial, $p_1(x) = f(a) + f'(a)\,(x-a)$, is the best linear approximation to f. The n^{th} order Taylor polynomial, $p_n(x)$ will be the best n^{th} degree polynomial approximation to f. The second-order Taylor polynomial for f at $x = a$ is

$$p_2(x) = f(a) + f'(a)\,(x-a) + \frac{f''(a)}{2}\,(x-a)^2.$$

It's the only quadratic polynomial that has (1) the same value at $x = a$ that f has, (2) the same first derivative at $x = a$ that f has, and (3) the same second derivative at $x = a$ that f has. It is so close to f near $x = a$ that the difference between f and p_2, called the *remainder term*, $R_2(x) = f(x) - p_2(x)$, approaches 0 as $x \to a$ much faster than $(x-a)^2$ approaches 0, that is,

$$\lim_{x\to a} \frac{R_2(x)}{(x-a)^2} = 0.$$

Analogous arguments show that the n^{th} order Taylor polynomial for f at $x = a$ is

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

where $f^{(k)}$ indicates the k^{th} derivative and $k!$ is k factorial. As n increases, the approximations $p_n(x)$ usually get closer to $f(x)$ for values of x near enough to a, so

$$\lim_{n\to\infty} p_n(x) = f(x)$$

for most functions f that we encounter. The "infinite"-order polynomial is called the *Taylor series* for f. When the sum of the Taylor series is equal to f for values of x near a, which it usually is, we say f is *analytic* near a. But for the time being, we're only interested in the polynomials, and to begin with, only the quadratic polynomials.

Second-order Taylor polynomials for functions of several variables. Let's start with functions of two variables. Using the same techniques that we used for a function of one variable, you can show that the best second-order approximation for $f : \mathbf{R}^2 \to \mathbf{R}$ at $\mathbf{x} = \mathbf{a}$ is the function

$$\begin{aligned} p_2(x,y) &= f(a,b) + f_x(a,b)\,(x-a) + f_y(a,b)\,(y-b) \\ &\quad + \frac{1}{2}\left(f_{xx}(a,b)\,(x-a)^2 + 2f_{xy}(a,b)(x-a)(y-b) + f_{yy}(a,b)\,(y-b)^2\right). \end{aligned}$$

The graph of the second-order Taylor polynomial $p_2(x, y)$ is the quadric surface that best approximates the graph of f at (a, b). Later in this chapter we'll use p_2 to help us determine whether a critical point of f (an $\mathbf{x}$ such that $\frac{\partial f}{dx}(\mathbf{x}) = 0$ and $\frac{\partial f}{dy}(\mathbf{x}) = 0$) gives a maximum of f, a minimum of f, or neither.

This result for functions of two variables generalizes to functions $f : \mathbf{R}^n \to \mathbf{R}$ of n variables. Using summation notation and vector notation,

$$\begin{aligned} p_2(\mathbf{x}) &= f(\mathbf{a}) + \sum_{i=1}^{n} f_{x_i}(\mathbf{a})\,(x_i - a_i) \\ &\quad + \tfrac{1}{2} \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{a})\,(x_i - a_i)(x_j - a_j) \end{aligned}$$

You can imagine how to generalize this to higher-order Taylor polynomials of functions of several variables.

The Hessian. When we looked at the first-order approximation p_1, we found a way to express the total differential df in terms of partial derivatives. Now we'll do something like that with the second-order approximation p_2. Define the *Hessian* H as this $n \times n$ matrix of second-order partial derivatives

$$Hf = \begin{bmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{bmatrix}$$

Then the last part of the second-order Taylor's formula, the part

$$\tfrac{1}{2} \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{a})\,(x_i - a_i)(x_j - a_j)$$

can be written as

$$\tfrac{1}{2}\mathbf{h}^T\, Hf(\mathbf{a})\,\mathbf{h}$$

where $\mathbf{h}$ is the column matrix of differences

$$\mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_n \end{bmatrix} = \begin{bmatrix} x_1 - a_1 \\ x_2 - a_2 \\ \vdots \\ x_n - a_n \end{bmatrix},$$

and $\mathbf{h}^T$ is the transpose of $\mathbf{h}$. Then the second-order Taylor polynomial can be written as

$$p_2(\mathbf{x}) = f(\mathbf{a}) + Df(\mathbf{a})\,\mathbf{h} + \tfrac{1}{2}\mathbf{h}^T\, Hf(\mathbf{a})\,\mathbf{h}.$$

This $p_2(\mathbf{x})$ is the second-order approximation for f at a.

Ludwig Otto Hesse (1811–1874) introduced this matrix and its determinant in his study of algebraic curves.

Math 131 Home Page at `http://math.clarku.edu/~djoyce/ma131/`