Multi-Dimensional Taylor Series

Wayne Hacker

Contents

Higher-order approximations to $f(x, y)$

Recall that in Calculus I, you approximated a function f by its tangent line: if $|x - x_0|$ was sufficiently small,

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0)\,.$$

This is the first two terms in the Taylor expansion of f about the point x_0. If you want more accuracy, you keep more terms in the Taylor series. In particular, by keeping one additional term, we get what is called a "second-order approximation". It has the form

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \frac{1}{6}f'''(\xi)(x - x_0)^3\,. \qquad (0.1)$$

The first two terms make up the tangent-line, or linear, approximation. The first three terms make up the second-order approximation. The fourth term is called the error term, and it allows us to use "=" instead of "≈" in the equation. In it, ξ is between x_0 and x: either $x_0 < \xi < x$ or $x < \xi < x_0$, depending on whether x_0 is greater or smaller than x. We don't know exactly what value ξ has, but we can use it to estimate the maximum possible error in our approximation.

Why use the second-order approximation? There are two approaches to answering this question: a geometric and an algebraic one.

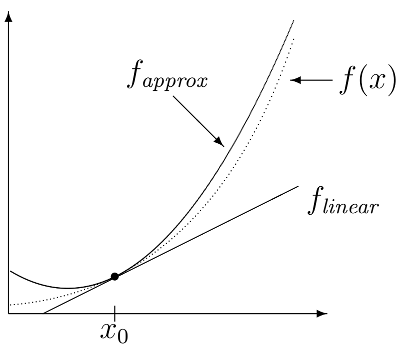
The geometric approach is more intuitive. The first-order approximation, or linear approximation,

$$f_{linear}(x) = f(x_0) + f'(x_0)(x - x_0)$$

approximates $f(x)$ by a line passing through $(x_0, f(x_0))$ and tangent to $f(x)$ at that point. It's a good approximation as long as x is close enough to x_0 that the curve of $f(x)$ between them can be regarded as a straight line. The second-order approximation

$$f_{approx}(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2$$

approximates $f(x)$ near x_0 as a parabola passing through $(x_0, f(x_0))$, with the same tangent line at x_0, and also with the same concavity at x_0. Thus even as $f(x)$ curves away from the tangent line to x_0, the parabolic approximation can curve with it.



The algebraic approach is based on the error terms in the Taylor expansion. We saw in (0.1) that the error term was $\frac{1}{6}f'''(\xi)(x-x_0)^3$. The corresponding equation for the first-order approximation is

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{1}{2}f''(\tilde{\xi})(x-x_0)^2 \,.$$

Note that $\tilde{\xi}$ in the first-order equation need not be the same as ξ in (0.1). However, like ξ, it must be the case that $\tilde{\xi}$ is between x_0 and x.

Thus

$$f_x - f_{linear}(x) = (\text{some constant})(x-x_0)^2 \,;$$
$$f_x - f_{approx}(x) = (\text{some other constant})(x-x_0)^3 \,.$$

If $|x-x_0|$ is "small", i.e. much smaller than 1, then $|(x-x_0)^3|$ is much smaller than $(x-x_0)^2$. You can see that if you let $x - x_0 = 10^{-n}$ for $n = 1, 2, 3, ...$

Now, let us extend this idea to functions of higher dimensions. Recall that the tangent-plane approximation to the function $z = f(x, y)$ at the point (x_0, y_0) is

$$f(x,y) \approx z_{TP}(x_0,y_0) = f(x_0,y_0) + \vec{\nabla} f(x_0,y_0)\cdot d\vec{x} \,,$$

where $d\vec{x} = \langle x - x_0, y - y_0 \rangle$.

The second-order approximation is

$$\begin{aligned} f(x,y) \approx f(x_0,y_0) + \vec{\nabla} f(x_0,y_0)\cdot d\vec{x} + \frac{1}{2}f_{xx}(x_0,y_0)(x-x_0)^2 \\ + f_{xy}(x_0,y_0)(x-x_0)(y-y_0) + \frac{1}{2}f_{yy}(x_0,y_0)(y-y_0)^2 \,. \end{aligned} \tag{0.2}$$

How did we get this formula? We know how to work with a one-dimensional Taylor series; and we know a directional derivative is just a one-dimensional derivative: the slope of a curve in the z-$\vec{u}$ plane, where $\vec{u}$ is the direction in which we take the derivative. For example, f_x is the same thing as $f_{\hat{\imath}}$, taken in the plane containing $\hat{\imath}$ (and therefore the x-axis) and the z-axis. By analogy, we might expect a "two-dimensional" Taylor series to look like a "one-dimensional" one when viewed in the proper way.

Let (x_0, y_0) be a fixed point in the plane. Suppose we want to approximate $f(x, y)$ at some other point (x, y). Since Taylor series are constructed from derivatives, and since the derivative for a general direction is a directional derivative, it makes sense to parameterize (x, y) to be on the same line as (x_0, y_0). In that way, the domain is reduced to one dimension, just as it is for $f_{\vec{u}}$.

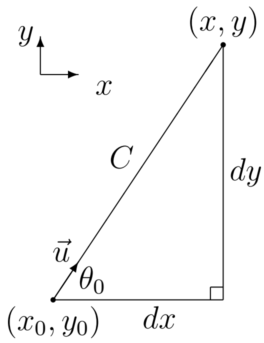
We parameterize the line segment joining (x_0, y_0) and (x, y) by s and write it in terms of the direction vector $\vec{u} = \langle \cos\theta_0, \sin\theta_0 \rangle$, where θ_0 is the direction from (x_0, y_0) to (x, y). Then $x = x(s)$; $y = y(s)$; and $f(x(s), y(s)) = F(s)$. We want to expand $F(s)$ about $s = 0$, i.e. $(x(0) = x_0, y(0) = y_0)$. This parameterization reduces a two-dimensional domain to a one-dimensional one, and a two-dimensional function $f(x, y)$ to a one-dimensional function $F(s)$. Instead of taking ∂_x and ∂_y, we take ∂_s. The situation is illustrated below:

The parameter s is the distance from (x_0, y_0) in the direction of (x, y). This direction is

$$\vec{u} = \langle \cos\theta_0, \sin\theta_0 \rangle \, .$$

The curve C is the line segment from (x_0, y_0) to (x, y). It is parameterized by

$$\begin{aligned} x &= x_0 + s\cos\theta_0 \\ y &= y_0 + s\sin\theta_0 \, . \end{aligned}$$



We expand $F(s)$ in a one-dimensional Taylor series about $s = 0$:

$$F(s) = F(0) + \partial_s F(0) s + \frac{1}{2}\partial_s^2 F(0) s^2 + \frac{1}{6}\partial_s^3 F(\bar{s}) s^3 \, , \tag{0.3}$$

where $\bar{s}$ is analogous to ξ in (0.1): $0 < \bar{s} < s$.

Consider the second term on the right side of (0.3). By the chain rule,

$$\begin{aligned} \partial_s F(s) = \partial_s f(x(s), y(s)) &= f_x \partial_s x + f_y \partial_s y \\ &= f_x \cos\theta_0 + f_y \sin\theta_0 \\ &= \vec{\nabla} f \cdot \vec{u} = f_{\vec{u}}(x, y) \\ &= \partial_{\vec{u}} f(x(s), y(s)) \, . \end{aligned}$$

Thus $\partial_s = \partial_{\vec{u}}$. By a similar argument, we can show that

$$\partial_s^2 F(s) = \partial_{\vec{u}}^2 f(x(s), y(s)) = f_{\vec{u}\vec{u}} \, . \tag{0.4}$$

At $s = 0$, $(x(s), y(s)) = (x_0, y_0)$; so

$$\partial_s F(0) = f_{\vec{u}}(x_0, y_0) = \vec{\nabla} f(x_0, y_0) \cdot \vec{u} = f_x(x_0, y_0)\cos\theta_0 + f_y(x_0, y_0)\sin\theta_0 \, .$$

Therefore

$$\begin{aligned}\partial_s F(0)s &= f_x(x_0,y_0)(s\cos\theta_0) + f_y(x_0,y_0)(s\sin\theta_0)\\ &= f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)\,,\end{aligned} \tag{0.5}$$

since $x - x_0 = s\cos\theta_0$ and $y - y_0 = s\sin\theta_0$.

Now, let's look at the third term on the right side of (0.3). From (0.4), we have $\partial_s^2 F = f_{\vec{u}\vec{u}}$. Now

$$\begin{aligned}f_{\vec{u}\vec{u}} &= \partial_{\vec{u}} f_{\vec{u}} = \partial_{\vec{u}}\left(\vec{\nabla} f\cdot\vec{u}\right) = \partial_{\vec{u}}\left(f_x\cos\theta_0 + f_y\sin\theta_0\right)\\ &= \partial_{\vec{u}} f_x\cos\theta_0 + \partial_{\vec{u}} f_y\sin\theta_0\,.\end{aligned} \tag{0.6}$$

Recall that $\partial_{\vec{u}} g = g_{\vec{u}} = \vec{\nabla} g\cdot\vec{u}$. If we apply this to $g = f_x$ and then $g = f_y$, we get

$$\begin{aligned}\partial_{\vec{u}} f_x &= \vec{\nabla} f_x\cdot\vec{u} = f_{xx}\cos\theta_0 + f_{xy}\sin\theta_0\\ \partial_{\vec{u}} f_y &= \vec{\nabla} f_y\cdot\vec{u} = f_{yx}\cos\theta_0 + f_{yy}\sin\theta_0\,.\end{aligned}$$

When we put these results into (0.6), assuming $f_{xy} = f_{yx}$, we get

$$\begin{aligned}f_{\vec{u}\vec{u}} &= (f_{xx}\cos\theta_0 + f_{yy}\sin\theta_0)\cos\theta_0 + (f_{xy}\cos\theta_0 + f_{yy}\sin\theta_0)\sin\theta_0\\ &= f_{xx}\cos^2\theta_0 + 2f_{xy}\sin\theta_0\cos\theta_0 + f_{yy}\sin^2\theta_0\end{aligned}$$

Thus

$$\begin{aligned}\partial_s^2 F(0)s^2 &= f_{\vec{u}\vec{u}}(x_0,y_0)s^2\\ &= f_{xx}(s\cos\theta_0)^2 + 2f_{xy}(s\cos\theta_0)(s\sin\theta_0) + f_{yy}(s\sin\theta_0)^2\\ &= f_{xx}(x-x_0)^2 + 2f_{xy}(x-x_0)(y-y_0) + f_{yy}(y-y_0)^2\,,\end{aligned} \tag{0.7}$$

where we have used $x - x_0 = s\cos\theta_0$ and $y - y_0 = s\sin\theta_0$.

Now, substitute (0.5) and (0.7) into (0.3), along with the fact that $F(0) = f(x_0, y_0)$.

$$\begin{aligned}F(s) = F(x,y) = f(x_0,y_0) &+ f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)\\ &+ \frac{1}{2}f_{xx}(x_0,y_0)(x-x_0)^2 + f_{xy}(x_0,y_0)(x-x_0)(y-y_0)\\ &+ \frac{1}{2}f_{yy}(x_0,y_0)(y-y_0)^2 + (\text{error term})\,.\end{aligned}$$

Assuming that the error term is small, this is equivalent to (0.2).