Taylor's Theorem in One and Several Variables

MA 433
Kurt Bryan

Taylor's Theorem in 1D

The simplest case of Taylor's theorem is in one dimension, in the "first order" case, which is equivalent to the Mean Value Theorem from Calc I. In this case we have that

If $f(x)$ is differentiable on an open interval I containing a then for any x in I

$$f(x) = f(a) + (x-a)f'(b) \tag{1}$$

for some b between a and x.

Of course, b will depend on x. Now if x is close to a then b will also be close to a and we can write, to some approximation,

$$f(x) \approx f(a) + f'(a)(x-a),$$

at least if f' is continuous. In other words, near $x = a$ the function $f(x)$ is well-approximated by the LINEAR function $L(x) = f(a) + (x-a)f'(a)$. This is just the "tangent line approximation" from Calculus I.

The second order case of Taylor's theorem states that

If $f(x)$ is twice differentiable on an open interval I containing a then for any x in I

$$f(x) = f(a) + f'(a)(x-a) + \frac{1}{2}f''(b)(x-a)^2 \tag{2}$$

for some b between a and x.

Imprecisely, if x is close to a then b is close to a, so $f(x) \approx f(a) + f'(a)(x-a) + \frac{1}{2}f''(a)(x-a)^2$, at least if f'' is continuous.

Example: Consider the function $f(x) = x^2 \sin(x/2)$ with base point $a = 2$. In Figure 1 below is the tangent line approximation to f at a. In Figure 2 below is the quadratic approximation to f at $a = 2$; the quadratic approximation is clearly better than the tangent line.

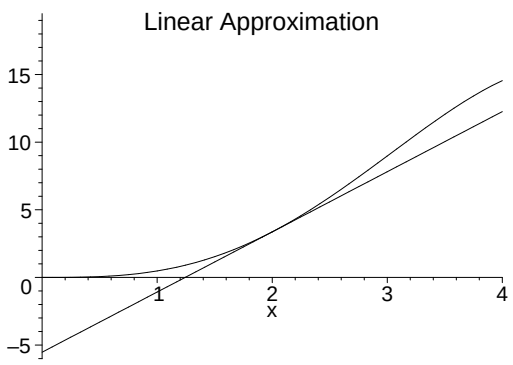

Figure 1: Graph $f(x) = x^2 \sin(x/2)$ and tangent line

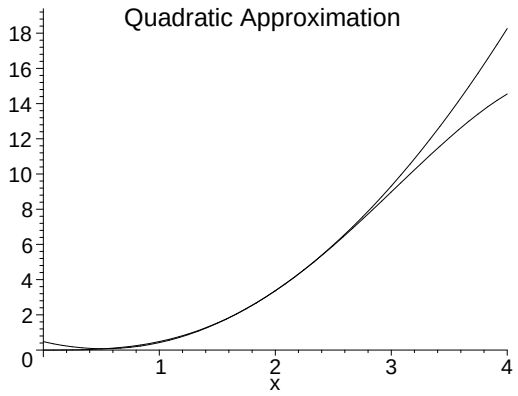

Figure 2: Graph $f(x) = x^2 \sin(x/2)$ and quadratic approximation

Proving Taylor's Theorem for the quadratic case is also pretty easy. Suppose we are given a function $f(x)$ which is twice differentiable on some interval containing a. Fix $x \neq a$ and choose some number M so that

$$f(x) - (f(a) + f'(a)(x-a) + M(x-a)^2) = 0. \tag{3}$$

We can always find such an M (just solve the above equation for M in terms of everything else!) Define a function $g(t)$ as

$$g(t) = f(t) - (f(a) + f'(a)(t-a) + M(t-a)^2).$$

Note that $g(a) = g'(a) = 0$ and $g(x) = 0$. By the Mean Value Theorem (or really, Rolle's Theorem) there must be some number x_1 between a and x such that $g'(x_1) = 0$. Applying Rolle's Theorem again, there must be some number b between a and x_1 such that $g''(b) = 0$. But this means that

$$g''(b) = f''(b) - 2M = 0,$$

or $M = f''(b)/2$. But if $M = f''(b)/2$ then equation (3) is exactly the statement of Taylor's Theorem.

Taylor's Theorem in Higher Dimensions

Let $\mathbf{x} = (x_1, \ldots, x_n)$ and consider a function $f(\mathbf{x})$. Let $\mathbf{a} = (a_1, \ldots, a_n)$ and suppose that f is differentiable (all first partials exists) in an open ball B around $\mathbf{a}$. Then the first order case of Taylor's Theorem says that

If $f(\mathbf{x})$ is differentiable on an open ball B around $\mathbf{a}$ and $\mathbf{x} \in B$ then

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{b})(x_k - a_k) = f(\mathbf{a}) + \nabla f(\mathbf{b}) \cdot (\mathbf{x} - \mathbf{a}) \tag{4}$$

for some $\mathbf{b}$ on the line segment joining $\mathbf{a}$ and $\mathbf{x}$.

The proof of this is easy from the single variable case of Taylor's Theorem. consider the function

$$g(t) = f(\mathbf{a} + t(\mathbf{x} - \mathbf{a})). \tag{5}$$

Note that $g(0) = f(\mathbf{a})$, while $g(1) = f(\mathbf{x})$. If f has all first partial derivatives then g is differentiable, and according to the Mean Value Theorem, for some $t^* \in (0,1)$ we have

$$g(1) = g(0) + g'(t^*). \tag{6}$$

But from the chain rule

$$g'(t^*) = \nabla f(\mathbf{a} + t^*(\mathbf{x} - \mathbf{a})) \cdot (\mathbf{x} - \mathbf{a}). \tag{7}$$

With $\mathbf{b} = \mathbf{a} + t^*(\mathbf{x} - \mathbf{a})$ equations (6) and (7) yield equation (4).

As in the one dimensional case, if $\mathbf{x}$ is close to $\mathbf{a}$ then $\mathbf{b}$ is also close to $\mathbf{a}$ and so if the first partial derivatives of f are continuous we have that

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a}).$$

Example: Let $f(x_1, x_2) = x_1 e^{x_2} + \cos(x_1 x_2)$, with $\mathbf{a} = (1, 0)$. Then near $\mathbf{a}$ we have

$$\begin{aligned} f(x_1, x_2) &= f(\mathbf{a}) + \frac{\partial f}{\partial x_1}(\mathbf{b})(x_1 - a_1) + \frac{\partial f}{\partial x_2}(\mathbf{b})(x_2 - a_2) \\ &\approx 2 + \frac{\partial f}{\partial x_1}(\mathbf{a})(x_1 - a_1) + \frac{\partial f}{\partial x_2}(\mathbf{a})(x_2 - a_2) \\ &= 2 + (e^{a_2} - a_2 \sin(a_1 a_2))(x_1 - a_1) + (a_1 e^{a_2} - a_1 \sin(a_1 a_2))(x_2 - a_2) \\ &= 2 + (x_1 - 1) + x_2 = 1 + x_1 + x_2, \end{aligned}$$

so near $(1, 0)$ the function is well-approximated by the linear function $1 + x_1 + x_2$. You can check, for example, that $f(1.1, -0.2) = 1.8765$, while the linearized approximation gives $1 + 1.1 - 0.2 = 1.9$.

The second order case of Taylor's Theorem in n dimensions is

If $f(\mathbf{x})$ is twice differentiable (all second partials exist) on a ball B around $\mathbf{a}$ and $\mathbf{x} \in B$ then

$$f(\mathbf{x}) = f(\mathbf{a}) + \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{a})(x_k - a_k) + \frac{1}{2} \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(\mathbf{b})(x_j - a_j)(x_k - a_k) \tag{8}$$

for some $\mathbf{b}$ on the line segment joining a and x.

Despite the apparent complexity, this really just says that f is approximated by a quadratic polynomial. The proof of the second order case for n variables is just like the proof of the first order case. Define $g(t)$ as in equation (5) and note that from the second order case of Taylor's Theorem we have

$$g(1) = g(0) + g'(0) + \frac{1}{2} g''(t^*) \tag{9}$$

for some $t^* \in (0, 1)$. Also, $g(0) = f(\mathbf{a}), g'(0) = \nabla f(\mathbf{a}) \cdot (\mathbf{x} - \mathbf{a})$, and you can check (bit of a mess) that

$$g''(t) = \sum_{j,k=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_k}(\mathbf{a} + t^*(\mathbf{x} - \mathbf{a}))(x_j - a_j)(x_k - a_k) \tag{10}$$

Equations (9) and (10) together with the definition $\mathbf{b} = \mathbf{a} + t^*(\mathbf{x} - \mathbf{a})$ yields the second order case of Taylor's Theorem.

Example: As in the above example, let $f(x_1, x_2) = x_1 e^{x_2} + \cos(x_1 x_2)$. We've already computed the first two terms in equation (8); the terms involving the second partials yield (when evaluated at $\mathbf{a} = (1, 0)$)

$$f(\mathbf{x}) \approx 1 + x_1 + x_2 + (x_1 - 1)x_2.$$

The Matrix Form of Taylor's Theorem

There is a nicer way to write the Taylor's approximation to a function of several variables. Let's write all vectors like $\mathbf{x} = (x_1, x_2)$ as columns, e.g.,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Recall that the *transpose* of a vector $\mathbf{x}$ is written as $\mathbf{x}^T$ and just means $\mathbf{x}^T = [x_1 \;\; x_2]$. In this case we can write the first order Taylor approximation (4) as

$$f(\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x} - \mathbf{a})^T \nabla f(\mathbf{b})$$

where the second term is computed as a matrix multiplication.

The second order case can be written out as

$$f(\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x} - \mathbf{a})^T \nabla f(\mathbf{a}) + \frac{1}{2}(\mathbf{x} - \mathbf{a})^T \mathbf{H}(\mathbf{b})(\mathbf{x} - \mathbf{a})$$

where $\mathbf{H}$ is the so-called *Hessian* matrix of second partial derivatives defined as

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ & & \vdots & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

Notice that since $\frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j}$ for functions which are twice continuously differentiable, the matrix $\mathbf{H}$ is symmetric.