Chapter 3
Approximations

Abstract Approximations are at the heart of calculus. In Chapter 1 we saw that the transformation of differentials $dx = \varphi'(s)\,ds$ can be traced back to the linear approximation $\Delta x \approx \varphi'(s)\,\Delta s$ (the microscope equation), and that the factor $\varphi'(s)$ represented a local length multiplier. We also suggested there that the transformation $dx\,dy = r\,dr\,d\theta$ of differentials from Cartesian to polar coordinates has the same explanation: the polar coordinate change map has a linear approximation (a two-variable "microscope" equation) and the factor r is the local area multiplier for that map. In this chapter we construct a variety of useful approximations to nonlinear functions of one or more variables. However, we save for the following chapter a discussion of the most important approximation, the *derivative* of a map.

3.1 Mean-value theorems

Measuring change with the derivative

The derivative indicates how much a function changes. It does this in the microscope equation, for example, and also in a similar equation called the law of the mean. First, consider the microscope equation, in this form:

$$f(x) - f(a) \approx f'(a)(x-a) \quad \text{for } x \approx a.$$

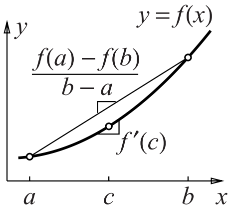
This says each change $x-a$ in input produces a change $f(x) - f(a)$ in output that is approximately $f'(a)$ times as large. Although this equation is only approximate, note that the multiplier $f'(a)$ that links the changes is the same for all x in some sufficiently small neighborhood of a.

Law of the mean

The **law of the mean** makes a similar statement, but there are important differences. It says that for each fixed x, say $x = b$, we can find a "mean value" c somewhere between a and b for which

$$f(b) - f(a) = f'(c)(b-a).$$

J.J. Callahan, *Advanced Calculus: A Geometric View*, Undergraduate Texts in Mathematics, DOI 10.1007/978-1-4419-7332-0_3, © Springer Science+Business Media, LLC 2010



The link between changes in input and output is now exact rather than approximate, and b need not be near a, as in the microscope equation. However, these benefits come at a cost: the location of the c varies with b and, in fact, is not usually explicitly known, even when a and b are. Also, f' has to be continuous from a to b. As the figure indicates, c is to be chosen so that the slope $f'(c)$ is equal to the slope of the line segment from $(a, f(a))$ to $(b, f(b))$. We can put the law this way:

When x changes from a to b, then $f(x)$ undergoes a change that is exactly $f'(c)$ times as large.

The law of the mean is the first of several mean-value theorems we consider, and it is the most basic: all others follow from it.

Bounding the magnitude of change

Our ignorance of the location of c makes it difficult to use the law of the mean in some circumstances. However, it is natural to assume that we do have information about f and its derivative. In particular, we can make use of a bound on the size of $|f'(x)|$.

Theorem 3.1 (Mean-value theorem). *Suppose $f(x)$ is continuously differentiable for all x between a and b; then*

$$|f(b) - f(a)| \leq \max_{a \leq x \leq b} |f'(x)||b - a|. \qquad \square$$

Although the law of the mean states explicitly—by means of an equality—how much the function grows (i.e., how $f(b) - f(a)$ depends on $b - a$), the mean-value theorem just provides a bound on that growth in terms of a bound on the derivative. As we show below, there is no direct extension of the law of the mean to *vector*-valued functions, that is, to functions $\mathbf{x} = \mathbf{f}(\mathbf{v})$ where the value $\mathbf{x}$ is a vector in $\mathbb{R}^p$, with $p \geq 2$. For a vector-valued function, we are only able to bound its growth by the size of its derivative, as in Theorem 3.1.

The law of the mean is frequently called the *mean-value theorem*; however, we reserve the latter term for the general theorem that extends to vector-valued functions.

Integral mean-value theorem

Continuing on, we formulate an integral law of the mean.

Theorem 3.2 (Law of the mean for integrals). *If $f(x)$ is a continuous function on the interval from a to b, then*

$$\int_a^b f(x)\,dx = f(c)(b - a)$$

for at least one point c in that interval.

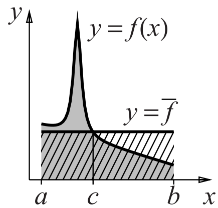
Proof. Let $F(x)$ be an antiderivative of $f(x)$ (i.e., $F'(x) = f(x)$). The fundamental theorem of calculus guarantees that F exists because f is continuous. If we now apply the law of the mean to F, we find

$$\int_a^b f(x)\,dx = F(b) - F(a) = F'(c)(b - a) = f(c)(b - a)$$

for some c between a and b. $\square$

Mean value of a function

We can connect this with the *mean value* of f. By definition, the **mean value** of a function $y = f(x)$ on the interval $a \le x \le b$ is the constant $\overline{f}$ whose integral over $[a,b]$ (the hatched region in the figure) is equal to the integral of $f(x)$ over $[a,b]$ (the shaded region):



$$\int_a^b f(x)\,dx = \int_a^b \overline{f}\,dx = \overline{f} \times (b-a).$$

This leads to the more familiar form of the definition:

$$\text{mean value of } f \text{ on } [a,b] = \overline{f} = \frac{1}{b-a}\int_a^b f(x)\,dx.$$

Because $\overline{f}$ lies between the maximum and minimum values of $f(x)$ on $[a,b]$, and because f is continuous, there is at least one c between a and b for which $f(c) = \overline{f}$. In other words, in the equation

$$\int_a^b f(x)\,dx = f(c)(b-a)$$

provided by the integral law of the mean, $f(c)$ is in fact the mean value of f on the interval $[a,b]$.

Because we know $f(c) = \overline{f}$ in the integral law of the mean, we can sometimes determine c explicitly. For example, let $f(x) = \sqrt{r^2 - x^2}$. The graph of $y = f(x)$ is a semicircle of radius r on the interval $[-r,r]$, so the area under it is $\pi r^2/2$ and

$$\frac{\pi r^2}{2} = \int_{-r}^{r} \sqrt{r^2 - x^2}\,dx = f(c)\,2r = 2r\sqrt{r^2 - c^2}.$$

We can solve this for c and get $c = \pm r\sqrt{1 - \pi^2/16}$; see the exercises.

Generalized law of the mean for integrals

The integral law of the mean has us compute an integral by extracting the mean value of its integrand. The following theorem makes a more general assertion: there are circumstances where we can compute an integral by extracting the mean value of just a part of its integrand.

Theorem 3.3 (Generalized law of the mean for integrals). *If $f(x)$ and $g(x)$ are continuous on the interval $[a,b]$ and $g(x) \ge 0$ there, then*

$$\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx$$

for at least one point c in that interval.

Proof. Our proof here echoes the previous argument, which said that $\overline{f} = f(c)$ for some c because $\overline{f}$ lies between the minimum and maximum values of the continuous function f.

Note first that if $g(x) = 0$ for all x in the interval $[a,b]$, then both integrals equal 0 and there is nothing to prove. So we assume $g(x) > 0$ for at least some x in the interval; because g is continuous and $g(x) \ge 0$, it also follows that

$$\int_a^b g(x)\,dx > 0.$$

Now let m and M be the minimum and maximum values of $f(x)$ on $[a,b]$; thus $m \le f(x) \le M$. Because $0 \le g(x)$, we also have

$$mg(x) \le f(x)g(x) \le Mg(x)$$

for all x in the interval, and therefore

$$m\int_a^b g(x)\,dx \le \int_a^b f(x)g(x)\,dx \le M\int_a^b g(x)\,dx.$$

We have already noted that the integral of g is nonzero, so we can divide these inequalities by it and conclude that the expression

$$\frac{\displaystyle\int_a^b f(x)g(x)\,dx}{\displaystyle\int_a^b g(x)\,dx}$$

lies between the minimum and maximum values of $f(x)$ on the interval. Therefore, because f is continuous, this expression must equal $f(c)$ for at least one value c in the interval. □

The condition $g(x) \ge 0$ can be changed to $g(x) \le 0$ without affecting the truth of the theorem; however, the theorem does fail if $g(x)$ changes sign on the interval. You can explore these points in the exercises. The proof of the generalized law of the mean can be adapted to have the result stated as an inequality, as with the mean-value theorem for functions.

Corollary 3.4 *If $f(x)$ and $g(x) \ge 0$ are continuous on $[a,b]$, then*

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \max_{a\le x\le b} |f(x)| \left|\int_a^b g(x)\,dx\right|. \qquad \square$$

Law of the mean for multivariable functions

Until now we have considered only functions of a single variable, but there are analogous mean-value theorems for functions of several variables. To extend the law of the mean to such functions, it is convenient for us to recast the law in a slightly different form. Let $\Delta x = b - a$; then the point c that lies between a and b can be written as $c = a + \theta\Delta x$ for some $0 \le \theta \le 1$, and the law itself can be written as

$$f(a+\Delta x) = f(a) + f'(a+\theta\Delta x)\,\Delta x.$$

Now consider a function of several variables $F(\mathbf{x}) = F(x_1,\ldots,x_n)$. The analogue of the ordinary derivative $f'(x)$ is the gradient of F (constructed with the differential operator ∇, "nabla"):

$$\nabla F(\mathbf{x}) = \operatorname{grad} F(\mathbf{x}) = \left(\frac{\partial F}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial F}{\partial x_n}(\mathbf{x})\right).$$

Theorem 3.5 (Law of the mean). *Suppose $F(\mathbf{x}) = F(x_1, \ldots, x_n)$ is a continuously differentiable function of $\mathbf{x}$; then*

$$F(\mathbf{a} + \Delta\mathbf{x}) - F(\mathbf{a}) = \nabla F(\mathbf{a} + \theta\Delta\mathbf{x}) \cdot \Delta\mathbf{x},$$

for some $0 \leq \theta \leq 1$.

Proof. There is a simple way to reduce this to a single-variable question; let $f(t) = F(\mathbf{a} + t\Delta\mathbf{x}) = F(a_1 + t\Delta x_1, \ldots, a_n + t\Delta x_n)$. Then the chain rule gives us the derivative of f:

$$f'(t) = \frac{\partial F}{\partial x_1}(\mathbf{a} + t\Delta\mathbf{x})\,\Delta x_1 + \cdots + \frac{\partial F}{\partial x_n}(\mathbf{a} + t\Delta\mathbf{x})\,\Delta x_n = \nabla F(\mathbf{a} + t\Delta\mathbf{x}) \cdot \Delta\mathbf{x}.$$

This is the scalar (or dot) product of the gradient with the vector $\Delta\mathbf{x}$. By the law of the mean for $f(t)$, we know there is a $0 \leq \theta \leq 1$ for which the following is true:

$$\begin{aligned} F(\mathbf{a} + \Delta\mathbf{x}) - F(\mathbf{a}) &= f(1) - f(0) = f'(\theta)(1 - 0) \\ &= \nabla F(\mathbf{a} + \theta\Delta\mathbf{x}) \cdot \Delta\mathbf{x}. \end{aligned}$$ □

Thus, even for a function $F(\mathbf{x})$ of several variables, if we use the dot product for multiplication, we can express the law of the mean in the following way.

When $\mathbf{x}$ changes from $\mathbf{a}$ to $\mathbf{b} = \mathbf{a} + \Delta\mathbf{x}$, the change in $F(\mathbf{x})$ is exactly ∇F times as large, where the gradient ∇F is evaluated at some intermediate (or "mean") point along the line from $\mathbf{a}$ to $\mathbf{b}$.

We can rephrase the multivariable law of the mean as an inequality that bounds the growth of F in terms of a bound on its derivative (i.e., its gradient):

Corollary 3.6 (Mean-value theorem) *If $z = F(\mathbf{x})$ is continuously differentiable on the line that connects $\mathbf{a}$ and $\mathbf{b}$, then*

$$|F(\mathbf{b}) - F(\mathbf{a})| \leq \max_{\mathbf{x}} \|\nabla F(\mathbf{x})\| \|\mathbf{b} - \mathbf{a}\|,$$

where the maximum is taken over all points $\mathbf{x}$ on the line from $\mathbf{a}$ to $\mathbf{b}$. □

Double integrals

For functions of two variables, we have a natural extension of the integral law of the mean. The proof follows the pattern of all previous proofs; see the exercises. We deal with functions of three or more variables in a later chapter, after we discuss their integrals.

Definition 3.1 *The* ***mean value $\overline{F}$*** *of the function $F(x,y)$ on the domain D in $\mathbb{R}^2$ is*

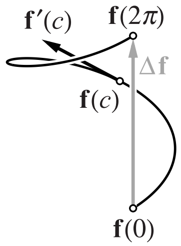
$$\overline{F} = \frac{1}{\operatorname{area} D} \iint_D F(x,y)\,dx\,dy.$$

Theorem 3.7 (Law of the mean for double integrals). *Let $F(x,y)$ be a continuous function on a connected domain D in $\mathbb{R}^2$. Then there is at least one point (c,d) in D where F takes on its mean value $\overline{F}$; thus*

$$\iint_D F(x,y)\,dx\,dy = F(c,d) \times \operatorname{area} D. \qquad \square$$

The problem with vector-valued functions

To complete the present study of mean-value theorems, let us consider *vector*-valued functions $\mathbf{x} = \mathbf{f}(\mathbf{v})$, where $\mathbf{x}$ is a vector in $\mathbb{R}^p$, $p \geq 2$. For even in the simplest such case—a vector-valued function $\mathbf{x} = \mathbf{f}(t)$ of a single variable, which defines a curve in $\mathbb{R}^p$—we now show there can be no direct extension of the law of the mean as an equality.



To see why, consider the helix $\mathbf{x} = \mathbf{f}(t) = (\cos t, \sin t, t)$ in $\mathbb{R}^3$. Let us try to express the change $\mathbf{f}(2\pi) - \mathbf{f}(0)$ in the form $\mathbf{f}'(c)(2\pi - 0)$ for some suitable mean value c between 0 and 2π. The vector $\Delta\mathbf{f} = \mathbf{f}(2\pi) - \mathbf{f}(0) = (0,0,2\pi)$ is vertical, but the derivative

$$\mathbf{f}'(c) = (-\sin c, \cos c, 1)$$

is never vertical, because its first two components are never simultaneously zero. Therefore, no scalar multiple of $\mathbf{f}'(c)$ will ever equal $\Delta\mathbf{f}$, even approximately. In particular, there is no number c for which

$$\mathbf{f}(2\pi) - \mathbf{f}(0) = \mathbf{f}'(c)(2\pi - 0).$$

Even though the law of the mean itself fails to hold, we can still get a *bound* on the size of the change in $\mathbf{f}$ that is exactly analogous to the bound provided by Theorem 3.1; in fact, we have

$$\|\mathbf{f}(2\pi) - \mathbf{f}(0)\| \leq \max \|\mathbf{f}'(t)\| |2\pi - 0|,$$

because the left-hand side equals 2π and the right-hand side equals $2\sqrt{2}\pi$.

Theorem 3.8 (Mean-value theorem). *If $\mathbf{f} : I \to \mathbb{R}^p$ has a continuous derivative on an interval I that contains a and b, then*

$$\|\mathbf{f}(b) - \mathbf{f}(a)\| \leq \max_{a \leq t \leq b} \|\mathbf{f}'(t)\| |b - a|.$$

Proof. Because $\mathbf{f}(b) - \mathbf{f}(a) = \displaystyle\int_a^b \mathbf{f}'(t)\,dt$, we have

$$\|\mathbf{f}(b) - \mathbf{f}(a)\| = \left\| \int_a^b \mathbf{f}'(t)\,dt \right\| \leq \max_{a \leq t \leq b} \|\mathbf{f}'(t)\| \left| \int_a^b dt \right| \leq \max_{a \leq t \leq b} \|\mathbf{f}'(t)\| |b - a|. \qquad \square$$

Extension to multivariable inputs

This theorem relies on the fact that we can regard the derivative $\mathbf{f}'(t)$ as a vector of the same sort as $\mathbf{f}(t)$ itself; this, in turn, is a consequence of the fact that the input is just a single variable t. If, instead, $\mathbf{x} = \mathbf{f}(\mathbf{v})$, where $\mathbf{v}$ is in $\mathbb{R}^p$, $p \geq 2$, then the *derivative of* $\mathbf{f}$ is something new. We define this new derivative below (Definition 3.16, p. 99) and analyze it in the following two chapters. At that time, we state and prove a natural extension of Theorem 3.8 for maps (Theorem 4.15, p. 140).

3.2 Taylor polynomials in one variable

New approximations

You are probably familiar with Taylor polynomials for $f(x)$ written in the form

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

These expressions give us new ways to approximate $f(x)$. One obvious benefit of any polynomial approximation is that it can be computed using only the four basic operations of arithmetic; most functions are not computable in this sense. Taylor polynomials become better approximations as n increases, and as x gets closer to a. We also see less obvious reasons that make them valuable.

Properties of the Taylor approximation

Consider successive terms in the Taylor polynomial to be separate contributions to the approximation. Then we find that the lowest-order term makes the largest contribution (at least when x is close to a) whereas the succeeeding terms, involving higher and higher powers of $x-a$, contribute less and less to the total. Also, because a Taylor polynomial gives a better approximation to $f(x)$ the closer x is to a, we write the polynomial instead in terms of the variable $\Delta x = x-a$ that indicates how close x is to a. We took the same approach when we reformulated the law of the mean (pp. 74ff.)

Taylor polynomial; $\Delta x = x-a$

Definition 3.2 *Suppose $y=f(x)$ has derivatives up to order n at $x=a$; then the* ***Taylor polynomial of degree n for f at x = a*** *is*

$$P_{n,a}(\Delta x)=f(a)+f'(a)\Delta x+\frac{f''(a)}{2!}(\Delta x)^2+\cdots+\frac{f^{(n)}(a)}{n!}(\Delta x)^n.$$

Notice that this expression for $P_{n,a}$ includes each of the Taylor polynomials $P_{0,a}$, $P_{1,a}, P_{2,a}, \ldots, P_{n-1,a}$ as an initial part.

Estimates and errors

In many cases, it is easy to calculate these polynomials. For example, if $f(x) = \sqrt{x}$, $n=3$, and $a=100$, then

$$P_{3,100}(\Delta x)=10+\frac{\Delta x}{20}-\frac{(\Delta x)^2}{8\,000}+\frac{(\Delta x)^3}{1\,600\,000}.$$

Let us see how this gives us approximate values of $\sqrt{x}$ when $x \approx 100$. First we build a sequence of estimates of $\sqrt{102}$ (so $\Delta x = 2$) by adding in the terms of the polynomial, one at a time. This allows us to see how the approximation improves as the degree increases. Second, we then do the same for $\sqrt{120}$ ($\Delta x = 20$). Comparing the two sets of estimates allows us to see how the approximation improves as Δx decreases. In all cases our focus is on the **error**: that is, on the difference between the true value and the approximation.

Contribution made by each term in $P_{3,100}(2)$

First, we consider how each term in the cubic Taylor polynomial

$$P_{3,100}(2) = 10 + \frac{2}{20} - \frac{4}{8\,000} + \frac{8}{1\,600\,000} = 10.099\,505$$

contributes to the estimate of $\sqrt{102} = 10.099\,504\,938\,362\ldots$. Here are the results in a table:

Degree	Term	Sum	Error $= \sqrt{102} - \text{Sum}$	
0	10	10	$0.099\,5\ldots$	$\approx \; 1 \times 10^{-1}$
1	0.1	10.1	$-0.000\,495\ldots$	$\approx -5 \times 10^{-4}$
2	$-0.000\,5$	10.099 5	$0.000\,004\,938\ldots$	$\approx \; 5 \times 10^{-6}$
3	0.000 005	10.099 505	$-0.000\,000\,061\,6\ldots$	$\approx -6 \times 10^{-8}$

Thus we see that higher terms contribute less and less to the sum, but they effectively "fine-tune" the estimate. In fact, the contributions drive the error down *exponentially*; that is, the error at each stage made by the intermediate sum $P_{k,100}(2)$ is roughly of size 10^{-ak-b}, for some $a > 0$.

Comparative errors for $\sqrt{102}$ and $\sqrt{120}$

Of course the terms get smaller because their coefficients do, and this is clearly the result of the choice of the original function $f(x) = \sqrt{x}$. For a different function, the coefficients may not be so obliging. Nevertheless, by comparing the errors that $P_{3,100}(\Delta x)$ makes for different values of Δx, we largely eliminate the effect of the coefficients. At the same time, we see how the error is connected to the size of Δx, our second objective. The following table gives comparative information for $\sqrt{120} = 10.954\,451\,15\ldots$ ($\Delta x = 20$).

Degree	Term	Sum	Error $= \sqrt{120} - \text{Sum}$	
0	10	10	$0.954\ldots$	$\approx \; 1 \times 10^{0}$
1	1	11	$-0.045\,5\ldots$	$\approx -5 \times 10^{-2}$
2	-0.05	10.95	$0.004\,45\ldots$	$\approx \; 4 \times 10^{-3}$
3	0.005	10.955	$-0.000\,548\ldots$	$\approx -5 \times 10^{-4}$

Compare the rightmost columns of the two tables for $k = 2$ or 3: $x = 102$ is only 1/10 as far from 100 as $x = 120$, and the error that $P_{k,100}$ makes in estimating $\sqrt{102}$ is, roughly speaking, only about $1/10^{k+1}$ times as large as the error for $\sqrt{120}$. Later (p. 83), we confirm this is true even for $k > 3$.

Errors and Taylor's theorem

Our experiments with $\sqrt{x}$ suggest that that we should study how

$$\text{error} = f(a + \Delta x) - P_{n,a}(\Delta x)$$

depends on Δx and on n in general. The result is contained in Taylor's theorem, which we state and prove below. It spells out the error, and with it we are able to see that $P_{n,a}(\Delta x)$ makes a smaller error than any other polynomial of the same degree in approximating $f(x)$ near a. Before we state the theorem, let us look first at the simple case when $n = 0$.

The Taylor polynomial for $n = 0$ is just the constant function $P_{0,a}(\Delta x) = f(a)$, so we have the estimate

The error for $P_{0,a}(\Delta x)$

$$\text{error} = f(a+\Delta x) - f(a) \approx f'(a)\,\Delta x$$

by using the microscope equation. Because $f'(a)$ is fixed, this expression already tells us the error is roughly proportional to the size of Δx itself. So if x_1 is 1/10 as far from a as x_2 (and x_2 is still sufficiently close to a), then the error that $P_{0,a}$ makes in estimating $f(x_1)$ will be about 1/10 as large as the error in estimating $f(x_2)$.

By contrast, the law of the mean (p. 74) gives us the exact error, but in terms of a quantity $0 \le t \le 1$ whose value we may not be able to determine effectively:

$$f(a+\Delta x) - f(a) = f'(a+t\Delta x)\,\Delta x.$$

Because the derivative of $f(a+t\Delta x)$ with respect to t is $f'(a+t\Delta x)\,\Delta x$ (chain rule), we can also express the exact error as an integral:

$$\int_0^1 f'(a+t\Delta x)\,\Delta x\,dt = f(a+t\Delta x)\Big|_0^1 = f(a+\Delta x) - f(a).$$

Although this integral is more complicated-looking than the other ways of writing the error, it turns out to be the most useful. Let us rewrite the last equation as

$$f(a+\Delta x) = f(a) + \int_0^1 f'(a+t\Delta x)\,\Delta x\,dt = P_{0,a}(\Delta x) + R_{0,a}(\Delta x);$$

we call

$$R_{0,a}(\Delta x) = \Delta x \int_0^1 f'(a+t\Delta x)\,dt$$

the **remainder** because it is what is left after we subtract the Taylor polynomial from the function. Of course it is also the error we make in replacing the function value by the polynomial value. We are now ready to state Taylor's theorem; it expresses the remainder for a general $P_{n,a}$ as an integral.

Taylor's formula with remainder

Theorem 3.9 (Taylor). *If $f(x)$ has continuous derivatives up to order $n+1$ on an interval containing a and $a+\Delta x$, then*

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

where $R_{n,a}(\Delta x) = \dfrac{(\Delta x)^{n+1}}{n!} \displaystyle\int_0^1 f^{(n+1)}(a+t\Delta x)(1-t)^n\,dt.$

Proof by induction

Proof. Because the theorem is essentially a sequence of formulas—one for each value of n—we prove them one at a time, "by induction on n." That is, we first prove the formula involving $P_{0,a}$, then use it to prove the one involving $P_{1,a}$, then use that to prove the one involving $P_{2,a}$, and so on, generating each new remainder as we go. To prove Taylor's formula for $P_{0,a}$,

$$f(a+\Delta x) = f(a) + \Delta x \int_0^1 f'(a+t\Delta x)\,dt,$$

just set $\varphi(t) = f(a+t\Delta x)$ and use the fundamental theorem of calculus in the form

$$\varphi(1) = \varphi(0) + \int_0^1 \varphi'(t)\,dt.$$

First induction step

The induction that takes us from one formula to the next is just an integration by parts carried out on the remainder integral. The integration by parts formula we use is

$$\int_\alpha^\beta u\,dv = uv\Big|_\alpha^\beta - \int_\alpha^\beta v\,du.$$

To begin, we set $u = f'(a+t\Delta x)$ and $dv = dt$ in the formula for $P_{0,a}$. Then $du = f''(a+t\Delta x)\,\Delta x\;dt$ and $v = t + C$, where C is a constant of integration that we determine in a moment. We have

$$\begin{aligned} R_{0,a}(\Delta x) &= \Delta x \int_0^1 f'(a+t\Delta x)\,dt \\ &= \Delta x\, f'(a+t\Delta x)(t+C)\Big|_0^1 - \Delta x \int_0^1 (t+C) f''(a+t\Delta x)\,\Delta x\,dt \\ &= \Delta x\, f'(a+\Delta x)(1+C) - \Delta x\, f'(a)\,C - (\Delta x)^2 \int_0^1 (t+C) f''(a+t\Delta x)\,dt. \end{aligned}$$

If we now set $C = -1$ the first term on the right disappears and the second becomes $+f'(a)\,\Delta x$, which is exactly the linear term in $P_{1,a}(x)$. Thus the formula with $P_{0,a}$ becomes (note the sign changes)

$$f(a+\Delta x) = f(a) + R_{0,a}(\Delta x) = \underbrace{f(a) + f'(a)\,\Delta x}_{P_{1,a}(\Delta x)} + \underbrace{(\Delta x)^2 \int_0^1 f''(a+t\Delta x)(1-t)\,dt}_{R_{1,a}(\Delta x)}.$$

The error for $P_{1,a}(x)$

The error in replacing $f(a+\Delta x)$ by $P_{1,a}(\Delta x)$ is thus the new remainder,

$$R_{1,a}(\Delta x) = (\Delta x)^2 \int_0^1 f''(a+t\Delta x)(1-t)\,dt.$$

Second induction step

The second induction step starts with Taylor's formula when $n = 1$:

$$\begin{aligned} f(a+\Delta x) &= P_{1,a}(\Delta x) + R_{1,a}(\Delta x) \\ &= f(a) + f'(a)\,\Delta x + (\Delta x)^2 \int_0^1 f''(a+t\Delta x)(1-t)\,dt. \end{aligned}$$

To integrate by parts here, use $u = f''(a+t\Delta x)$, $dv = (1-t)\,dt$, and $v = -(1-t)^2/2$; then

$$\begin{aligned}
R_{1,a}(\Delta x) &= (\Delta x)^2 \int_0^1 f''(a+t\Delta x)(1-t)\,dt \\
&= -(\Delta x)^2 f''(a+t\Delta x)\frac{(1-t)^2}{2}\bigg|_0^1 + (\Delta x)^2 \int_0^1 f^{(3)}(a+t\Delta x)\,\Delta x\,\frac{(1-t)^2}{2}\,dt \\
&= \frac{f''(a)}{2}(\Delta x)^2 + \frac{(\Delta x)^3}{2}\int_0^1 f^{(3)}(a+t\Delta x)(1-t)^2\,dt.
\end{aligned}$$

With $R_{1,a}(\Delta x)$ replaced by the two terms in the last line, our previous equation for $f(a+\Delta x)$ (i.e., Taylor's formula when $n=1$) now reads

$$f(a+\Delta x) = \underbrace{f(a)+f'(a)\,\Delta x + \frac{f''(a)}{2}(\Delta x)^2}_{P_{2,a}(\Delta x)} + \underbrace{\frac{(\Delta x)^3}{2}\int_0^1 f^{(3)}(a+t\Delta x)(1-t)^2\,dt}_{R_{2,a}(\Delta x)}.$$

As we see, this has become Taylor's formula when $n=2$.

Third induction step

In the next step, we are able to see how the factorial expressions arise. Our starting point is $f(a+\Delta x) = P_{2,a}(\Delta x) + R_{2,a}(\Delta x)$, and we must integrate $R_{2,a}(\Delta x)$ by parts. If we use $u = f^{(3)}(a+t\Delta x)$ and $v = -(1-t)^3/3$, then

$$\begin{aligned}
R_{2,a}(\Delta x) &= -\frac{(\Delta x)^3}{3\cdot 2} f^{(3)}(a+t\Delta x)(1-t)^3\bigg|_0^1 + \frac{(\Delta x)^3}{3\cdot 2}\int_0^1 f^{(4)}(a+t\Delta x)\,\Delta x\,(1-t)^3\,dt \\
&= \frac{f^{(3)}(a)}{3!}(\Delta x)^3 + \frac{(\Delta x)^4}{3!}\int_0^1 f^{(4)}(a+t\Delta x)(1-t)^3\,dt.
\end{aligned}$$

Consequently, Taylor's formula when $n=2$ becomes

$$\begin{aligned}
f(a+\Delta x) &= P_{2,a}(\Delta x) + R_{2,a}(\Delta x) \\
&= \underbrace{P_{2,a}(\Delta x) + \frac{f^{(3)}(a)}{3!}(\Delta x)^3}_{P_{3,a}(\Delta x)} + \underbrace{\frac{(\Delta x)^4}{3!}\int_0^1 f^{(4)}(a+t\Delta x)(1-t)^3\,dt}_{R_{3,a}(\Delta x)},
\end{aligned}$$

which is just Taylor's formula when $n=3$.

General induction step

To complete the induction, we must transform Taylor's formula when $n=k$,

$$f(a+\Delta x) = P_{k,a}(\Delta x) + R_{k,a}(\Delta x),$$

(where k is any nonnegative integer) into the corresponding formula when $n=k+1$,

$$f(a+\Delta x) = P_{k+1,a}(\Delta x) + R_{k+1,a}(\Delta x).$$

This is another integration by parts (see the exercises):

$$
\begin{aligned}
R_{k,a}(\Delta x) &= \frac{(\Delta x)^{k+1}}{k!}\int_0^1 f^{(k+1)}(a+t\Delta x)(1-t)^k\,dt \\
&= \frac{f^{(k+1)}(a)}{(k+1)!}(\Delta x)^{k+1} + \frac{(\Delta x)^{k+2}}{(k+1)!}\int_0^1 f^{(k+2)}(a+t\Delta x)(1-t)^{k+1}\,dt.
\end{aligned}
$$

It implies $f(a+\Delta x)$

$$
= \underbrace{P_{k,a}(\Delta x) + \frac{f^{(k+1)}(a)}{(k+1)!}(\Delta x)^{k+1}}_{P_{k+1,a}(\Delta x)} + \underbrace{\frac{(\Delta x)^{k+2}}{(k+1)!}\int_0^1 f^{(k+2)}(a+t\Delta x)(1-t)^{k+1}\,dt}_{R_{k+1,a}(\Delta x)},
$$

completing the general induction step. □

Forms of the remainder

Our approximations of $\sqrt{102}$ and $\sqrt{120}$ suggested that the error $R_{n,a}(\Delta x)$ gets small as Δx does. In fact, in those examples we saw that the error got small faster than Δx; $R_{n,a}(\Delta x)$ vanished like $(\Delta x)^{n+1}$. This is true in general; to see why, we first write $R_{n,a}(\Delta x)$ in some alternate forms.

Corollary 3.10 (Lagrange's form of the remainder) *For each* $\Delta x \approx 0$*, there is a* $\theta = \theta(\Delta x)$ *with* $0 \le \theta \le 1$ *for which*

$$
R_{n,a}(\Delta x) = \frac{f^{(n+1)}(a+\theta\Delta x)}{(n+1)!}(\Delta x)^{n+1}.
$$

Proof. With the generalized integral law of the mean (Theorem 3.3), we can extract $f^{(n+1)}(a+t\Delta x)$ from the integral defining $R_{n,a}(\Delta x)$, and then compute the integral of the remaining function, $(1-t)^n$, exactly. Thus, for a given Δx, there is a point θ in the interval $[0,1]$ for which

$$
\begin{aligned}
R_{n,a}(\Delta x) &= \frac{(\Delta x)^{n+1}}{n!}\int_0^1 f^{(n+1)}(a+t\Delta x)(1-t)^n\,dt \\
&= (\Delta x)^{n+1}\,\frac{f^{(n+1)}(a+\theta\Delta x)}{n!}\int_0^1 (1-t)^n\,dt \\
&= (\Delta x)^{n+1}\,\frac{f^{(n+1)}(a+\theta\Delta x)}{n!}\left(-\frac{(1-t)^{n+1}}{n+1}\right)\Bigg|_0^1 \\
&= \frac{f^{(n+1)}(a+\theta\Delta x)}{(n+1)!}(\Delta x)^{n+1}.
\end{aligned}
$$

□

Taylor's formula and the law of the mean

If we write Taylor's formula for $n=0$ using the Lagrange form of the remainder, we get

$$
f(a+\Delta x) = f(a) + f'(a+\theta\Delta x)\,\Delta x,
$$

which is just the law of the mean. Thus we can see Taylor's formula with Lagrange's remainder as an extension of the law of the mean that incorporates higher powers of the displacement Δx.

Comparative estimates of $\sqrt{102}$ and $\sqrt{120}$

With the remainder in Lagrange's form we can see why we said, on page 78, that the error $P_{k,100}$ makes in estimating $\sqrt{102}$ would be only about $1/10^{k+1}$ times the error in estimating $\sqrt{120}$. Consider first $k=3$; because $f(x)=\sqrt{x}$,

$$R_{3,100}(\Delta x) = \frac{f^{(4)}(100+\theta\Delta x)}{4!}(\Delta x)^4 = \frac{-15}{16\times 24\,(100+\theta\Delta x)^{7/2}}(\Delta x)^4.$$

Because the number $100+\theta\Delta x$ certainly lies between 100 and 120, the coefficient $-15/(16\times 24\,(100+\theta\Delta x)^{7/2})$ will lie in the narrow range from -4×10^{-9} to -2×10^{-9} for both estimates. So the main cause of the difference between the two errors must be the factor $(\Delta x)^4$: in the two cases, its values are $2^4=16$ and $20^4=16\times 10^4$. This is why $R_{3,100}(2)$ is only about $1/10^4$ times as large as $R_{3,100}(20)$. Furthermore, because $100\le 100+\theta\Delta x\le 120$, we see that the errors themselves must lie in the following ranges:

$$-6.4\times 10^{-8} < R_{3,100}(2) < -3.2\times 10^{-8},$$
$$-6.4\times 10^{-4} < R_{3,100}(20) < -3.2\times 10^{-4}.$$

In fact, we already obtained by direct calculation the values $R_{3,100}(2)\approx -6\times 10^{-8}$ and $R_{3,100}(20)\approx -5\times 10^{-4}$; they fit into these ranges, as they should.

How the comparative error depends on k

Now take an arbitrary $k\ge 2$; then (see the exercises)

$$R_{k,100}(\Delta x) = \frac{\pm 1\cdot 3\cdots(2k-1)}{2^{k+1}(k+1)!(100+\theta\Delta x)^{k+1/2}}(\Delta x)^{k+1}.$$

The term $1/(100+\theta\Delta x)^{k+1/2}$ again varies over a small range of values when we have $0\le\Delta x\le 20$; the main cause of the variation of $R_{k,100}(\Delta x)$ with Δx is the factor $(\Delta x)^{k+1}$. Therefore,

$$\frac{R_{k,100}(2)}{R_{k,100}(20)} \approx \frac{2^{k+1}}{20^{k+1}} = \frac{1}{10^{k+1}},$$

so the error $P_{k,100}$ makes in estimating $\sqrt{102}$ is only about $1/10^{k+1}$ times as large as the error estimating $\sqrt{120}$.

Taylor's formula and the microscope equation

Because $f^{(n+1)}$ is a continuous function, the factor $f^{(n+1)}(a+\theta\Delta x)$ in the Lagrange remainder is as close as we wish to $f^{(n+1)}(a)$ if Δx is sufficiently close to 0. This gives us another form of the remainder.

Corollary 3.11 (Generalized microscope equation) *When $\Delta x\approx 0$,*

$$R_{n,a}(\Delta x) \approx \frac{f^{(n+1)}(a)}{(n+1)!}(\Delta x)^{n+1}. \qquad \square$$

Notice that when $n = 0$, $R_{0,a}(\Delta x) \approx f'(a)\,\Delta x$, and Taylor's formula becomes the microscope equation,

$$f(a+\Delta x) \approx f(a) + f'(a)\,\Delta x.$$

This is why we call the statement in the corollary the generalized microscope equation. Taylor's formula thus generalizes both the microscope equation and the law of the mean (Lagrange's form).

The next term is most of the remainder

The generalized microscope equation is a remarkable result. It says that the remainder looks more and more like the next term in the Taylor expansion of f, the more we magnify the graph of the remainder in a microscope window centered at $\Delta x = 0$. Thus, because most of the error at the nth stage is equal to the term of degree $n+1$, we can eliminate most of the error by adding that term to the nth stage (i.e., to $P_{n,a}$). The result is, of course, the next Taylor polynomial, $P_{n+1,a}$. This is perhaps the simplest and most intuitive way of seeing how the Taylor polynomials arise.

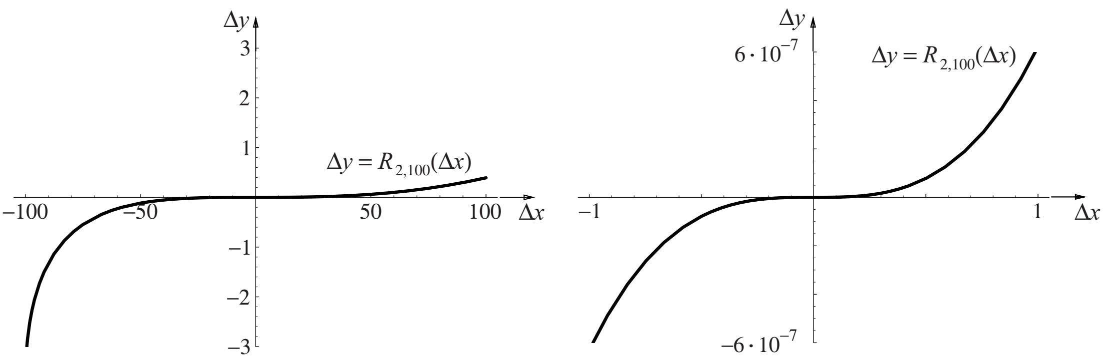
An example

Here is a visual example to help make it clear how much the remainder looks like the next term in the Taylor expansion. Let us again use the function $f(x) = \sqrt{x}$ at $a = 100$, but this time just the second-degree approximation instead of the third. According to the generalized microscope equation, the graph of the remainder,

$$R_{2,100}(\Delta x) = f(100+\Delta x) - P_{2,100}(\Delta x) = \sqrt{100+\Delta x} - \left(10 + \frac{\Delta x}{20} - \frac{(\Delta x)^2}{8\,000}\right),$$

Macroscopic versus microscopic

should look like the cubic term in $P_{3,100}$, at least if the domain is suitably restricted to a small interval around $\Delta x = 0$. Below we see two views of this graph.



On the left, where it is plotted over a large domain, $-100 < \Delta x < 100$, the graph looks only vaguely like a cubic. It fails to have the necessary symmetry, for example. But on the right, where its domain has been shrunk to $-1 \le \Delta x \le 1$ (and the vertical scale has been exaggerated), the graph is now indistinguishable from the graph of the cubic

$$\Delta y = \frac{(\Delta x)^3}{1\,600\,000} = 6.25 \times 10^{-7} \times (\Delta x)^3.$$

So $R_{2,100}(\Delta x)$ does indeed look like the cubic term $(\Delta x)^3/1\,600\,000$ in a sufficiently small window centered at $\Delta x = 0$. Note that we needed to exaggerate the vertical scale; the graph would otherwise have appeared to be just a horizontal line! The vertical scale is linked to the small coefficient $(1/1\,600\,000 \approx 6\times 10^{-7})$ of $(\Delta x)^3$.

A bound on $|R_{n,a}(\Delta x)|$

The following corollary says that $|R_{n,a}(\Delta x)|$ is bounded by (a multiple of) $|\Delta x|^{n+1}$. It relies on the continuity of the $(n+1)$st derivative of f, which was one of the original hypotheses of Taylor's theorem. The case $n=0$ is the ordinary mean-value theorem for f (Theorem 3.1, p. 72).

Corollary 3.12 *Let $f^{(n+1)}(a+\Delta x)$ be continuous for $|\Delta x| \le r$; then*

$$|R_{n,a}(\Delta x)| \le \max_x |f^{(n+1)}(x)|\,\frac{|\Delta x|^{n+1}}{(n+1)!},$$

where the maximum is taken over all x between a and $a+\Delta x$, inclusive.

Proof. When $0\le t\le 1$, then $x = a+t\Delta x$ lies between a and $a+\Delta x$, inclusive. The continuous function $f^{(n+1)}(a+t\Delta x)$ has a finite maximum on this closed interval. Therefore we have

$$\begin{aligned}|R_{n,a}(\Delta x)| &= \left|\frac{(\Delta x)^{n+1}}{n!}\int_0^1 f^{(n+1)}(a+t\Delta x)(1-t)^n\,dt\right| \\ &\le \frac{|\Delta x|^{n+1}}{n!}\max_{0\le t\le 1}|f^{(n+1)}(a+t\Delta x)|\int_0^1 (1-t)^n\,dt \\ &= \max_x |f^{(n+1)}(x)|\,\frac{|\Delta x|^{n+1}}{(n+1)!},\end{aligned}$$

because the value of the last integral is $1/(n+1)$. □

The order to which a power vanishes

With this corollary we finally have a simple and useful way to describe the size of the error and, in particular, how rapidly it vanishes as $\Delta x \to 0$. Here is the basic idea (expressed in terms of a variable t): although any positive power of a variable *t vanishes* (i.e., tends to 0) as $t\to 0$, a higher power vanish more rapidly, or as we say, *to a higher order*. For example, t^3 vanishes to a higher order than t^2 because the quotient t^3/t^2 also vanishes as $t\to 0$. We say that ***t* vanishes to the first order**, and $\boldsymbol{t^p}$ **vanishes to order *p*** (for any positive power $p>0$).

Vanishing to order greater than p: *little oh* notation

To describe the order of vanishing of an *arbitrary* function $\varphi(t)$ as $t\to 0$, we define what it means for $\varphi(t)$ to vanish to higher order than a given power of t, mimicking the way we compared t^3 and t^2. For the moment, we are concerned with the order of vanishing of $\varphi(t)$ only as $t\to 0$; later in the section we generalize to the case $t\to a$ for a arbitrary.

Definition 3.3 *We say* $\boldsymbol{\varphi(t)}$ ***vanishes to order greater than p (at the origin)****, and write* $\boldsymbol{\varphi(t) = o(p)}$*, if*

$$\lim_{t\to 0}\frac{\varphi(t)}{t^p} = 0.$$

The symbol "o" is called *little oh* and is meant to suggest the word *order*. Read "$o(p)$" as either "of order greater than p" or just "*little oh* of p."

Using ratios to compare orders of vanishing

The condition $\varphi(t) = o(p)$ is imprecise: when $\varphi(t)$ vanishes to higher order than t^p, we do not know how much higher: if $\varphi(t) = o(p)$, then $\varphi(t) = o(q)$ for every $q < p$. To get a more precise condition, let us look more closely at the ratio $\varphi(t)/t^p$ in the case when $\varphi(t) = Ct^m$ ($C \neq 0$); then, for $0 < p < \infty$, we have

$$\lim_{t\to 0} \frac{\varphi(t)}{t^p} = \begin{cases} 0 & p < m, \\ C & p = m, \\ \infty & p > m. \end{cases}$$

It is evident that we get the most information about $\varphi(t)$ not when the limit is zero but when it takes a finite nonzero value: that is, when $p = m$. Our example suggests that, to gain additional precision about the order of vanishing of an arbitrary $\varphi(t)$, we should focus on the value of p for which the ratio $\varphi(t)/t^p$ has a finite nonzero limit. This is certainly the right idea, but there are two technical stumbling blocks.

Functions can vanish unlike any power

First, consider the example of $\varphi(t) = 1/\ln|t|$. This vanishes at 0, but there is no $p > 0$ for which the limit $\varphi(t)/t^p$ is finite and nonzero (see the exercises). There is no way around this problem; some functions that do vanish still fail to vanish like any positive power of t.

With several variables, the ratio usually has no limit

Second, consider the two-variable function $\varphi(x,y) = x^2 + 2y^2$. (In the next section we extend Taylor's theorem to functions of several variables; the remainder is likewise a function of several variables, and we have to consider its order of vanishing.) It is reasonable to say $\varphi(x,y)$ vanishes to the same order as $x^2 + y^2$; they are both homogeneous quadratic polynomials. However,

$$\lim_{(x,y)\to(0,0)} \frac{x^2 + 2y^2}{x^2 + y^2}$$

does not exist. One way to see this is to note that, on the radial line $y = mx$, $\varphi(x, mx) = (1 + 2m^2)/(1 + m^2)$, a quantity that takes values between 1 and 2 as m varies. Nevertheless, we do have

$$1 \le \frac{x^2 + 2y^2}{x^2 + y^2} \le 2 \ \text{ for all } (x,y) \neq (0,0),$$

and this is enough to guarantee that $x^2 + 2y^2$ and $x^2 + y^2$ each vanishes as rapidly as the other. In fact, it is sufficient if such upper and lower bounds exist for all (x,y) sufficiently close to $(0,0)$.

Vanishing to the same order

Definition 3.4 *The functions* $\boldsymbol{\varphi(t)}$ ***and*** $\boldsymbol{\psi(t)}$ ***vanish to the same order (at the origin)*** *if there are positive constants δ, C_1, and C_2 for which*

$$C_1 \le \left| \frac{\varphi(t)}{\psi(t)} \right| \le C_2 \ \textit{ for all } 0 < |t| < \delta.$$

We can rewrite the inequalities one at a time so as to indicate that each function "dominates" the other in a completely symmetric way:

$$|\psi(t)| \le \frac{1}{C_1}\,|\varphi(t)|; \quad |\varphi(t)| \le C_2\,|\psi(t)|.$$

According to the first, $\psi(t)$ vanishes at least to the same order as $\varphi(t)$; according to the second, $\varphi(t)$ vanishes at least to the same order as $\psi(t)$. In the following definition, we have only one of these two inequalities, and the comparison is being made with a power.

Vanishing at least to order p: *big oh* notation

Definition 3.5 *We say* $\boldsymbol{\varphi(t)}$ ***vanishes at least to order p (at the origin)****, and write* $\boldsymbol{\varphi(t) = O(p)}$*, if there are positive constants* δ*, C for which* $|\varphi(t)| \le C|t|^p$ *when* $|t| < \delta$*. Otherwise, we say* $\boldsymbol{\varphi(t)}$ ***fails to vanish to order p****, and write* $\boldsymbol{\varphi(t) \neq O(p)}$*.*

The symbol "O" is called *big oh* and like "o" it is meant to suggest the word *order*. Read "$O(p)$" as either "of order at least p" or as "*big oh* of p." Note the following.

- $\varphi(t) = O(p)$ implies $\varphi(t) = O(\alpha)$ for all $0 < \alpha < p$.
- $\varphi(t) \neq O(p)$ implies $\varphi(t) \neq O(\beta)$ for all $\beta > p$.
- $\varphi(t) = o(p)$ implies $\varphi(t) = O(p)$, but the converse is not true: if $\varphi(t)$ vanishes at least to order p, there is no reason to think $\varphi(t)$ vanishes to higher order than p (e.g., $t^p = O(p)$ but $t^p \neq o(p)$). See also Exercise 3.15.

(We use O and o to indicate "order of vanishing;" however, in other settings they are used to indicate "order of magnitude." We avoid this phrase, though, because *magnitude* implies, etymologically at least, "largeness," not the "smallness" with which we are dealing.)

Big oh notation gives us the right level of precision to describe the order of vanishing of an approximation error. With it we get a convenient and vivid way to rewrite Taylor's formula. The first step is to restate Corollary 3.12 (p. 85) in the new language.

Corollary 3.13 $R_{n,a}(\Delta x) = O(n+1)$. □

$O(p)$ as a function; Taylor's formula

Next, we enlarge the meaning of $O(p)$ to allow it to stand for an otherwise unspecified function that vanishes at least to order p (or even allow it to stand for the set of such functions). Then, with this in mind, we can rewrite Taylor's formula in the following simple form that indicates just the order of the remainder:

$$f(a+\Delta x) = f(a) + f'(a)\,\Delta x + \frac{f''(a)}{2!}\,(\Delta x)^2 + \cdots + \frac{f^{(n)}(a)}{n!}\,(\Delta x)^n + O(n+1).$$

In words: $f(a+\Delta x)$ equals the Taylor polynomial of degree n plus *some unspecified function that vanishes to order* $n+1$ *in* Δx. Often this level of precision is all we need. Consider, for example, the infinite Taylor series for $f(x) = \ln x$ at $a = 1$:

$$\ln(1+\Delta x)$$
$$= \underbrace{\Delta x - \frac{(\Delta x)^2}{2} + \cdots + (-1)^{n-1}\frac{(\Delta x)^n}{n}}_{P_{n,1}(\Delta x)} + \underbrace{(-1)^n\frac{(\Delta x)^{n+1}}{n+1} + (-1)^{n+1}\frac{(\Delta x)^{n+2}}{n+2} + \cdots}_{O(n+1)}.$$

The first n terms constitute the Taylor polynomial $P_{n,1}(\Delta x)$; the rest are the remainder $O(n+1)$ in an explicit form. This shows how apt it is to think of $O(n+1)$ as a shorthand for "the terms that vanish at least to order $n+1$."

Functions that agree at least to order p

Definition 3.6 *We say* $\boldsymbol{\varphi(t)}$ ***and*** $\boldsymbol{\psi(t)}$ ***agree at least to order*** $\boldsymbol{p}$ ***in*** $\boldsymbol{t}$, *and write* $\boldsymbol{\varphi(t) = \psi(t) + O(p)}$, *if* $\varphi(t) - \psi(t) = O(p)$.

With this definition, we can put Taylor's formula,

$$f(a+\Delta x) = P_{n,a}(\Delta x) + O(n+1),$$

into these words: "$f(a+\Delta x)$ and $P_{n,a}(\Delta x)$ agree (or are equal) at least to order $n+1$ in Δx when Δx is near 0."

The "best fitting" approximation

Taylor's theorem tells us just half the story about the Taylor polynomial, namely, how well it approximates a given function. The other half of the story is that the Taylor polynomial is unique: no other polynomial of the same degree approximates the function as well. Theorem 3.14, below, explains just what this means.

Lemma 3.1. *If* $\varphi(t)/t^p \to \infty$ *as* $t \to 0$, *then* $\varphi(t) \neq O(p)$.

Proof. (By contradiction.) Suppose that $\varphi(t) = O(p)$; then $|\varphi(t)/t^p|$ would be bounded when $t \approx 0$. However, this contradicts the hypothesis that $\varphi(t)/t^p \to \infty$ as $t \to 0$. □

Theorem 3.14. *Suppose* $Q(\Delta x)$ *is a polynomial of degree* n *that differs from the Taylor polynomial* $P_{n,a}(\Delta x)$ *at least in the term of degree* k*; then* $f(a+\Delta x) - Q(\Delta x)$ *fails to vanish to order* $k+1$, *and hence*

$$f(a+\Delta x) - Q(\Delta x) \neq O(n+1).$$

Proof. The difference $S(\Delta x) = Q(\Delta x) - P_{n,a}(\Delta x)$ is also a polynomial of degree n:

$$S(\Delta x) = a_0 + a_1\Delta x + \cdots + a_k(\Delta x)^k + a_{k+1}(\Delta x)^{k+1} + \cdots + a_n(\Delta x)^n,$$

and $a_k \neq 0$ by hypothesis. Therefore

$$\frac{S(\Delta x)}{(\Delta x)^{k+1}} = \frac{a_0}{(\Delta x)^{k+1}} + \frac{a_1}{(\Delta x)^k} + \cdots + \frac{a_k}{\Delta x} + a_{k+1} + \cdots + a_n(\Delta x)^{n-k-1},$$

and this becomes infinite as $\Delta x \to 0$ (even if $a_0 = a_1 = \cdots = a_{k-1} = 0$), because $a_k \neq 0$. The error made by using $Q(\Delta x)$ to approximate $f(a+\Delta x)$ is

$$f(a+\Delta x)-Q(\Delta x)=f(a+\Delta x)-P_{n,a}(\Delta x)-S(\Delta x)=R_{n,a}(\Delta x)-S(\Delta x).$$

Therefore

$$\frac{f(a+\Delta x)-Q(\Delta x)}{(\Delta x)^{k+1}}=\frac{R_{n,a}(\Delta x)}{(\Delta x)^{k+1}}-\frac{S(\Delta x)}{(\Delta x)^{k+1}}$$

and, as we have seen, the second term becomes infinite as $t\to 0$. However, the first term remains bounded, because $R_{n,a}(\Delta x)=O(k+1)$ for all $k\le n$. Therefore, the two terms together become infinite. It follows from the lemma that

$$f(a+\Delta x)-Q(\Delta x)\ne O(k+1),$$

and hence $f(a+\Delta x)-Q(\Delta x)\ne O(n+1)$. □

Corollary 3.15 *If $K\le n$ is the degree of the lowest order term where Q and $P_{n,a}$ differ, then*

$$f(a+\Delta x)-Q(\Delta x)=O(K),\quad f(a+\Delta x)-Q(\Delta x)\ne O(K+1).$$

Proof. Exercise 3.21. □

Comparing $P_{n-1,a}$ and $P_{n,a}$

Finally, we see how a Taylor polynomial becomes a better approximation as its degree increases. There is one case where this fails to happen, namely when an increase in the degree leaves the polynomial unchanged. For example, with $f(x)=\sin x$, the first- and second-degree Taylor polynomials at the origin are

$$P_1(\Delta x)=\Delta x\ \text{ and }\ P_2(\Delta x)=\Delta x,$$

so $P_2(\Delta x)$ will be no better than $P_1(\Delta x)$ in approximating $f(0+\Delta x)=\sin\Delta x$. The problem is that P_2 lacks a quadratic term, because $f''(0)=0$. The following corollary avoids this case by requiring $f^{(n)}(a)\ne 0$, guaranteeing that the two polynomials are indeed different.

Corollary 3.16 *Suppose $f^{(n)}(a)\ne 0$; then $R_{n,a}(\Delta x)$ vanishes to a higher order than $R_{n-1,a}(\Delta x)$:*

$$R_{n,a}(\Delta x)=O(n+1)\ \textit{ but }\ R_{n-1,a}(\Delta x)\ne O(n+1).$$

Proof. Take $Q(\Delta x)=P_{n-1,a}(\Delta x)$ in the previous corollary; then $K=n$, because the term of degree n in $Q=P_{n-1,a}$ is 0, but in $R_{n,a}$ it is $f^{(n)}(a)\,(\Delta x)^n/n!\ne 0$. Therefore $R_{n-1,a}(\Delta x)\ne O(n+1)$. □

Order of vanishing at an arbitrary point

We end with a summary of definitions and results about the order of vanishing of functions at an arbitrary point.

Definition 3.7 *We say* $\boldsymbol{\varphi(t)}$ ***vanishes to order greater than*** $\boldsymbol{p}$ ***at*** $\boldsymbol{t=a}$, *and write* $\boldsymbol{\varphi(t)=o(p)}$ ***as*** $\boldsymbol{t\to a}$, *if*

$$\lim_{t\to a}\frac{\varphi(t)}{(t-a)^p}=0.$$

Thus, $\varphi(t) = o(p)$ is an abbreviation for $\varphi(t) = o(p)$ as $t \to 0$. We continue to use the briefer form unless clarity requires the longer one.

Definition 3.8 *The functions* $\boldsymbol{\varphi(t)}$ ***and*** $\boldsymbol{\psi(t)}$ ***vanish to the same order at*** $\boldsymbol{t = a}$ *if there are positive constants* δ, C_1, *and* C_2 *for which*

$$C_1 \le \left|\frac{\varphi(t)}{\psi(t)}\right| \le C_2 \ \textit{for all}\ 0 < |t-a| < \delta.$$

Definition 3.9 *We say* $\boldsymbol{\varphi(t)}$ ***vanishes at least to order p at*** $\boldsymbol{t = a}$, *and write* $\boldsymbol{\varphi(t) = O(p)}$ ***as*** $\boldsymbol{t \to a}$, *if there are positive constants* δ, C *for which* $|\varphi(t)| \le C|t-a|^p$ *when* $|t-a| < \delta$. *Otherwise, we say* $\boldsymbol{\varphi(t)}$ ***fails to vanish to order p at*** $\boldsymbol{t = a}$, *and write* $\boldsymbol{\varphi(t) \ne O(p)}$ ***as*** $\boldsymbol{t \to a}$.

Thus $\varphi(t) = O(p)$ is an abbreviation for $\varphi(t) = O(p)$ as $t \to 0$.

Definition 3.10 *We say* $\boldsymbol{\varphi(t)}$ ***and*** $\boldsymbol{\psi(t)}$ ***agree at least to order p at*** $\boldsymbol{t = a}$, *and write* $\boldsymbol{\varphi(t) = \psi(t) + O(p)}$ ***as*** $\boldsymbol{t \to a}$, *if* $\varphi(t) - \psi(t) = O(p)$ *as* $t \to a$.

3.3 Taylor polynomials in several variables

Taylor polynomials in two variables

We obtain Taylor polynomials for a function of two variables in a natural way from the one-variable version. However, the formulas are messy and therefore harder to interpret. For example, a polynomial in two or more variables can have several terms of the same degree. The collection of terms of a given degree forms a *homogeneous* polynomial. A homogeneous polynomial of degree k in two variables has the general form

$$\begin{aligned} Q(x,y) &= A_0x^k + A_1x^{k-1}y + A_2x^{k-2}y^2 + \cdots + A_{k-1}xy^{k-1} + A_ky^k \\ &= \sum_{i+j=k} A_jx^iy^j. \end{aligned}$$

The Taylor polynomial of degree n for a function $z = f(x,y)$ at $(x,y) = (a,b)$ consists of terms that are homogeneous polynomials in $\Delta x = x - a$ and $\Delta y = y - b$; there is a homogeneous polynomial of every degree between 0 and n. The terms involve the binomial coefficients

$$\binom{k}{j} = \frac{k!}{j!(k-j)!} = \binom{k}{k-j}$$

and partial derivatives of f. For the sake of visual clarity, we use subscripts to write the partial derivatives (e.g., $\partial^3 f/\partial x^2\,\partial y = f_{x^2y}$).

Definition 3.11 *Suppose all partial derivatives of* $f(x,y)$ *up to order* n *exist at* $(x,y) = (a,b)$; *then the* ***Taylor polynomial of degree n for f at*** $\boldsymbol{(a,b)}$ *is*

$$\begin{aligned}
P_{n,(a,b)}(\Delta x, \Delta y) &= f(a,b) + f_x(a,b)\,\Delta x + f_y(a,b)\,\Delta y \\
&\quad + \frac{1}{2!}\left(f_{xx}(a,b)\,(\Delta x)^2 + 2 f_{xy}(a,b)\,\Delta x\,\Delta y + f_{yy}(a,b)\,(\Delta y)^2\right) \\
&\quad + \cdots + \frac{1}{n!}\sum_{i+j=n}\binom{n}{j} f_{x^i y^j}(a,b)\,(\Delta x)^i\,(\Delta y)^j
\end{aligned}$$

Theorem 3.17 (Taylor). *If $f(x,y)$ has continuous partial derivatives up to order $n+1$ on an open set that contains the line segment from (a,b) to $(a+\Delta x, b+\Delta y)$, then*

Taylor's formula for functions of two variables

$$f(a+\Delta x, b+\Delta y) = P_{n,(a,b)}(\Delta x, \Delta y) + R_{n,(a,b)}(\Delta x, \Delta y),$$

where $R_{n,(a,b)}(\Delta x, \Delta y)$

$$= \frac{1}{n!}\sum_{i+j=n+1}\binom{n+1}{j}(\Delta x)^i\,(\Delta y)^j \int_0^1 f_{x^i y^j}(a+t\Delta x, b+t\Delta y)(1-t)^n\,dt.$$

Proof. The idea is to have the two-variable formula emerge from Taylor's formula for a suitably chosen function of one variable. We can assume $(\Delta x, \Delta y) \neq (0,0)$, for otherwise there is nothing to prove. In that case, there is a unique unit vector (α, β) for which $(\Delta x, \Delta y) = s(\alpha, \beta)$ with $s > 0$. Let

$$F(s) = f(a+s\alpha, b+s\beta) = f(a+\Delta x, b+\Delta y).$$

Taylor's formula for $F(s)$ at $s=0$ is

$$F(s) = F(0) + F'(0)\,s + \cdots + \frac{F^{(n)}(0)}{n!}s^n + \frac{s^{n+1}}{n!}\int_0^1 F^{(n+1)}(ts)(1-t)^n\,dt.$$

We claim this will turn into Taylor's formula for $f(a+\Delta x, b+\Delta y)$ when we express each derivative of F in terms of α and β and partial derivatives of f. One application of the chain rule gives

Derivatives of $F(s)$

$$F'(s) = f_x(a+s\alpha, b+s\beta)\,\alpha + f_y(a+s\alpha, b+s\beta)\,\beta.$$

A second application gives

$$\begin{aligned}
F''(s) &= f_{xx}(a+s\alpha, b+s\beta)\,\alpha^2 + f_{xy}(a+s\alpha, b+s\beta)\,\alpha\beta \\
&\quad + f_{yx}(a+s\alpha, b+s\beta)\,\beta\alpha + f_{yy}(a+s\alpha, b+s\beta)\,\beta^2, \\
&= f_{xx}(a+s\alpha, b+s\beta)\,\alpha^2 + 2 f_{xy}(a+s\alpha, b+s\beta)\,\alpha\beta \\
&\quad + f_{yy}(a+s\alpha, b+s\beta)\,\beta^2.
\end{aligned}$$

To get a clearer idea of the patterns being generated here, we calculate one more derivative. Applying the chain rule to each of the functions $f_{xx}(a+s\alpha, b+s\beta)$, $f_{xy}(a+s\alpha, b+s\beta)$, and $f_{yy}(a+s\alpha, b+s\beta)$, we get

$$\begin{aligned}F'''(s) &= f_{xxx}(a+s\alpha,b+s\beta)\,\alpha^3 + f_{xxy}(a+s\alpha,b+s\beta)\,\alpha^2\beta\\ &\qquad + 2f_{xyx}(a+s\alpha,b+s\beta)\,\alpha^2\beta + 2f_{xyy}(a+s\alpha,b+s\beta)\,\alpha\beta^2\\ &\qquad + f_{yyx}(a+s\alpha,b+s\beta)\,\alpha\beta^2 + f_{yyy}(a+s\alpha,b+s\beta)\,\beta^3\\ &= f_{xxx}(a+s\alpha,b+s\beta)\,\alpha^3 + 3f_{xxy}(a+s\alpha,b+s\beta)\,\alpha^2\beta\\ &\qquad + 3f_{xyy}(a+s\alpha,b+s\beta)\,\alpha^2\beta + f_{yyy}(a+s\alpha,b+s\beta)\,\beta^3.\end{aligned}$$

We have used $f_{xyx} = f_{xxy}$, and so forth, to combine terms.

A binomial expansion

For $k = 1$, 2, and 3, the formula for $F^{(k)}(s)$ is a sum of partial derivatives of f in which the numerical coefficients are the binomial coefficients. For an arbitrary k, the formula is

$$F^{(k)}(s) = \sum_{i+j=k} \binom{k}{j} f_{x^i y^j}(a+s\alpha,b+s\beta)\,\alpha^i\beta^j.$$

The next step is to determine the factor $F^{(k)}(0)\,s^k$ that appears in the kth term of the Taylor polynomial for $F(s)$. We have

$$\begin{aligned}F^{(k)}(0)\,s^k &= \sum_{i+j=k}\binom{k}{j} f_{x^iy^j}(a,b)\,s^k\alpha^i\beta^j = \sum_{i+j=k}\binom{k}{j} f_{x^iy^j}(a,b)\,(s\alpha)^i(s\beta)^j\\ &= \sum_{i+j=k}\binom{k}{j} f_{x^iy^j}(a,b)\,(\Delta x)^i(\Delta y)^j.\end{aligned}$$

These expresssions are equal because $s^k\alpha^q\beta^p = s^{i+j}\alpha^i\beta^j = (s^i\alpha^i)(s^j\beta^j)$ and, furthermore, $s\alpha = \Delta x$, $s\beta = \Delta y$.

Determining $R_{n,(a,b)}(\Delta x,\Delta y)$

At this point we have found all the terms in the Taylor polynomial $P_{n,(a,b)}$. The final step is to see how the remainder $R_{n,(a,b)}$ emerges from the remainder for $F(s)$. That remainder is

$$\begin{aligned}&\frac{s^{n+1}}{n!}\int_0^1 F^{(n+1)}(ts)(1-t)^n\,dt\\ &\quad= \frac{1}{n!}\int_0^1\left(\sum_{i+j=n+1}\binom{n+1}{j} f_{x^iy^j}(a+ts\alpha,b+ts\beta)\,s^{n+1}\alpha^i\beta^j\right)(1-t)^n\,dt\\ &\quad= \frac{1}{n!}\sum_{i+j=n+1}\binom{n+1}{j}(s\alpha)^i(s\beta)^j\int_0^1 f_{x^iy^j}(a+t\Delta x,b+t\Delta y)(1-t)^n\,dt\\ &\quad= R_{n,(a,b)}(\Delta x,\Delta y)\end{aligned}$$

when we set $(s\alpha)^i = (\Delta x)^i$, $(s\beta)^j = (\Delta y)^j$. □

Simplified notation

The large and unwieldy expression for the Taylor polynomial of a function of two variables gets even worse when there are more input variables. Before moving on to this, we introduce a simplifying notation for the two-variable polynomial that makes the r-variable case clearer.

The differential operator $\Delta\mathbf{x}\cdot\nabla$

The first step is to use vector notation. Thus, we write $\mathbf{x} = (x,y)$, $\mathbf{a} = (a,b)$, $\Delta\mathbf{x} = (\Delta x, \Delta y) = \mathbf{x} - \mathbf{a}$. The second step is to express the various partial derivatives in vector fashion, as well. A familiar example is the vector differential operator "nabla" $\nabla = (\partial/\partial x, \partial/\partial y)$ that is used for the gradient: $\operatorname{grad} f = \nabla f$. The operator we need is the dot product of $\Delta\mathbf{x}$ and ∇:

$$\Delta\mathbf{x}\cdot\nabla = \Delta x\,\frac{\partial}{\partial x} + \Delta y\,\frac{\partial}{\partial y}$$

This operator produces a certain "mixture" of the partial derivatives of any function it operates on:

$$(\Delta\mathbf{x}\cdot\nabla)\,f(\mathbf{x}) = \Delta x\,\frac{\partial f}{\partial x}(x,y) + \Delta y\,\frac{\partial f}{\partial y}(x,y).$$

In particular, $(\Delta\mathbf{x}\cdot\nabla)f(\mathbf{a})$ is just the linear homogeneous (i.e., first-degree) part of the Taylor polynomial for f at $\mathbf{a}$.

Composing $\Delta\mathbf{x}\cdot\nabla$ with itself

To get the homogeneous parts of higher degree, just apply the same differential operator to its previous output. In other words, compose $\Delta\mathbf{x}\cdot\nabla$ with itself, treating Δx and Δy as constants with respect to the partial differential operators $\partial/\partial x$ and $\partial/\partial y$. The resulting operator involves second derivatives:

$$\begin{aligned}(\Delta\mathbf{x}\cdot\nabla)^2 &= (\Delta\mathbf{x}\cdot\nabla)\circ(\Delta\mathbf{x}\cdot\nabla)\\ &= \left(\Delta x\,\frac{\partial}{\partial x} + \Delta y\,\frac{\partial}{\partial y}\right)\circ\left(\Delta x\,\frac{\partial}{\partial x} + \Delta y\,\frac{\partial}{\partial y}\right)\\ &= (\Delta x)^2\,\frac{\partial^2}{\partial x^2} + 2\Delta x\,\Delta y\,\frac{\partial^2}{\partial x\,\partial y} + (\Delta y)^2\,\frac{\partial^2}{\partial y^2},\end{aligned}$$

You should check that $(\Delta\mathbf{x}\cdot\nabla)^2 f(\mathbf{a})/2!$ is the homogeneous quadratic part of the Taylor polynomial of f at $\mathbf{a}$.

A binomial expansion

Repeated composition produces operators $(\Delta\mathbf{x}\cdot\nabla)^k$ involving derivatives of order k for any positive integer k. Because $\Delta\mathbf{x}\cdot\nabla$ is a binomial expression, each such power of $\Delta\mathbf{x}\cdot\nabla$ can be expanded as if it were an ordinary binomial:

$$(\Delta\mathbf{x}\cdot\nabla)^k = \left(\Delta x\,\frac{\partial}{\partial x} + \Delta y\,\frac{\partial}{\partial y}\right)^k = \sum_{i+j=k}\binom{k}{j}(\Delta x)^i(\Delta y)^j\,\frac{\partial^k}{\partial x^i\,\partial y^j}.$$

Notice that this is indeed a homogeneous polynomial of degree k in the variables Δx and Δy.

Taylor's formula in terms of $\Delta\mathbf{x}\cdot\nabla$

In terms of $\Delta\mathbf{x}\cdot\nabla$, the Taylor polynomial for $f(\mathbf{x})$ at $\mathbf{a}$ is just

$$P_{n,\mathbf{a}}(\Delta\mathbf{x}) = f(\mathbf{a}) + (\Delta\mathbf{x}\cdot\nabla)\,f(\mathbf{a}) + \frac{(\Delta\mathbf{x}\cdot\nabla)^2 f(\mathbf{a})}{2!} + \cdots + \frac{(\Delta\mathbf{x}\cdot\nabla)^n f(\mathbf{a})}{n!},$$

a much simpler expression than in the original definition (Definition 3.11)! The formula for the remainder is simplified in the same way:

$$R_{n,\mathbf{a}}(\Delta\mathbf{x}) = \frac{1}{n!}\int_0^1 (\Delta\mathbf{x}\cdot\nabla)^{n+1} f(\mathbf{a}+t\Delta\mathbf{x})(1-t)^n\,dt.$$

Functions of r variables

Let us move, finally, to the case of a function of r variables. As we have seen, the differential operator $\Delta\mathbf{x}\cdot\nabla$ plays the crucial role in the new notation. When there are r variables instead of two, so that

$$\mathbf{x} = (x_1, x_2, \ldots, x_r), \quad \mathbf{a} = (a_1, a_2, \ldots, a_r), \quad \Delta\mathbf{x} = \mathbf{x}-\mathbf{a},$$

our differential operator becomes a multinomial,

$$\Delta\mathbf{x}\cdot\nabla = \Delta x_1 \frac{\partial}{\partial x_1} + \cdots + \Delta x_r \frac{\partial}{\partial x_r},$$

instead of a binomial. Consequently, we can no longer represent the higher powers $(\Delta\mathbf{x}\cdot\nabla)^k$ using the binomial expansion.

A multinomial expansion

However, there is a way to expand multinomials that is exactly analogous to the binomial expansion. It uses the **multinomial coefficients**

$$\binom{k}{p_1\ p_2\ \cdots\ p_r} = \frac{k!}{p_1!p_2!\cdots p_r!}, \quad p_1+p_2+\cdots+p_r = k;$$

the **multinomial expansion** is

$$(\Delta\mathbf{x}\cdot\nabla)^k = \sum_{p_1+\cdots+p_r=k} \binom{k}{p_1\ \cdots\ p_r} (\Delta x_1)^{p_1}\cdots(\Delta x_r)^{p_r} \frac{\partial^k}{\partial x_1^{p_1}\cdots\partial x_r^{p_r}}.$$

Taylor's formula for functions of r variables

Theorem 3.18 (Taylor). *If $f(\mathbf{x})$ has $n+1$ continuous derivatives on an open set containing the line segment from $\mathbf{a}$ to $\mathbf{a}+\Delta\mathbf{x}$, then*

$$f(\mathbf{a}+\Delta\mathbf{x}) = \sum_{k=0}^{n} \frac{1}{k!}(\Delta\mathbf{x}\cdot\nabla)^k f(\mathbf{a}) + R_{n,\mathbf{a}}(\Delta\mathbf{x}),$$

where $R_{n,\mathbf{a}}(\Delta\mathbf{x}) = \dfrac{1}{n!}\displaystyle\int_0^1 (\Delta\mathbf{x}\cdot\nabla)^{n+1} f(\mathbf{a}+t\Delta\mathbf{x})(1-t)^n\,dt.$ □

Forms of the remainder

The remainder $R_{n,\mathbf{a}}(\Delta\mathbf{x})$ can be rewritten in different forms, just as in the one-variable case. The proofs are the same as the one-variable versions.

Corollary 3.19 (Lagrange's form of the remainder) *For each $\Delta\mathbf{x}\approx\mathbf{0}$, there is a $\theta = \theta(\Delta\mathbf{x})$ with $0\le\theta\le 1$ for which*

$$R_{n,\mathbf{a}}(\Delta\mathbf{x}) = \frac{1}{(n+1)!}(\Delta\mathbf{x}\cdot\nabla)^{n+1} f(\mathbf{a}+\theta\Delta\mathbf{x}). \quad \square$$

The next corollary asserts that the remainder for the Taylor polynomial of degree n is approximately the highest-degree homogeneous part of the Taylor polynomial of degree $n+1$.

Corollary 3.20 (Generalized microscope equation) *When $\Delta\mathbf{x} \approx \mathbf{0}$,*

$$R_{n,\mathbf{a}}(\Delta\mathbf{x}) \approx \frac{1}{(n+1)!}(\Delta\mathbf{x}\cdot\nabla)^{n+1} f(\mathbf{a}). \qquad \square$$

Also as in the one-variable case, the Taylor polynomial provides the "best fit" to a function near a given point, among all polynomials of the same degree. To see this, we use the same device we employed in the one-variable case, namely, the order of vanishing of the remainder. The definitions are analogous. Let

$$\mathbf{t} = (t_1,\ldots,t_r), \quad \|\mathbf{t}\| = \sqrt{t_1^2+\cdots+t_r^2},$$

and suppose $z = \varphi(\mathbf{t})$ is a real-valued function that vanishes at the origin: $\varphi(\mathbf{0}) = 0$. You can extend these definitions to the case where φ vanishes at an arbitrary point $\mathbf{a}$, as on pages 89–90.

Order of vanishing of a multivariable function

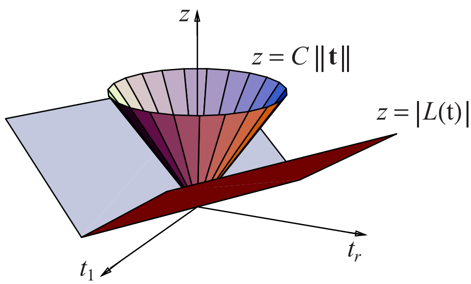
Definition 3.12 *We say the function* $\boldsymbol{\varphi}(\mathbf{t})$ ***vanishes to order greater than p****, and write* $\boldsymbol{\varphi}(\mathbf{t}) = \boldsymbol{o}(\boldsymbol{p})$*, if*

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{\varphi(\mathbf{t})}{\|\mathbf{t}\|^p} = 0.$$

Definition 3.13 *We say* $\boldsymbol{\varphi}(\mathbf{t})$ ***vanishes to order at least p****, and write* $\boldsymbol{\varphi}(\mathbf{t}) = \boldsymbol{O}(\boldsymbol{p})$*, if there are positive constants δ, C for which $|\varphi(\mathbf{t})| \le C\|\mathbf{t}\|^p$ when $\|\mathbf{t}\| < \delta$. Otherwise, we say* $\boldsymbol{\varphi}(\mathbf{t})$ ***fails to vanish to order p****, and write* $\boldsymbol{\varphi}(\mathbf{t}) \neq \boldsymbol{O}(\boldsymbol{p})$.

A linear function vanishes at least to first order

For example, any linear function $z = L(\mathbf{t}) = m_1t_1 + \cdots + m_rt_r$ vanishes at least to order 1: $|L(\mathbf{t})| \le C\|\mathbf{t}\|$, for some C. The graph of $z = C\|\mathbf{t}\|$ is a cone, whereas the graph of $z = |L(\mathbf{t})|$ resembles a (hyper)plane that has been folded upward along the set where $L(\mathbf{t}) = 0$. The cone can always be elongated enough (by increasing C) to make the folded hyperplane lie below it.



Corollary 3.21 $R_{n,\mathbf{a}}(\Delta\mathbf{x}) = O(n+1)$.

Proof. The proof of Taylor's theorem (Theorem 3.17) used the one-variable function

$$F(s) = f(\mathbf{a}+s\mathbf{u}) = f(\mathbf{a}+\Delta\mathbf{x}),$$

where $\mathbf{u}$ is a unit vector, $s > 0$, and $s\mathbf{u} = \Delta\mathbf{x}$. (The theorem was stated and proved when $\Delta\mathbf{x}$ was 2-dimensional, but nothing needs to be changed in higher dimensions.)

In the proof, Taylor's formula for $F(s)$ at $s=0$ became Taylor's formula for $f(\mathbf{x})$ at $\mathbf{x}=\mathbf{a}$. In particular, the remainder $R_{n,\mathbf{a}}(\Delta\mathbf{x})$ was just the remainder $R_{n,0}(s)$ for $F(s)$. We know $R_{n,0}(s)$ vanishes at least to order $n+1$ in s, so there are positive constants δ and C for which $|R_{n,0}(s)| \le C|s|^{n+1}$ when $|s|<\delta$. But $s=\|\Delta\mathbf{x}\|$, so

$$|R_{n,\mathbf{a}}(\Delta\mathbf{x})| = |R_{n,0}(s)| \le C|s|^{n+1} = C\|\Delta\mathbf{x}\|^{n+1}$$

when $\|\Delta\mathbf{x}\|<\delta$. □

The Taylor polynomial provides the "best fit"

Thus, $f(\mathbf{a}+\Delta\mathbf{x})$ agrees with its Taylor polynomial $P_{n,\mathbf{a}}(\Delta\mathbf{x})$ at least up to order $n+1$ in $\Delta\mathbf{x}$. There is no other polynomial of degree n for which this is true; according to the following theorem (which mimics the one-variable case), the agreement is always of lower order.

Theorem 3.22. *Suppose $Q(\Delta\mathbf{x})$ is a polynomial of degree n that differs from the Taylor polynomial $P_{n,\mathbf{a}}(\Delta\mathbf{x})$ at least in some term of degree $k\le n$; then*

$$f(\mathbf{a}+\Delta\mathbf{x}) - Q(\Delta\mathbf{x}) \ne O(k+1).$$

Proof. We can use the idea of the proof of the last corollary. Write $\Delta\mathbf{x}=s\mathbf{u}$ for a suitable $s>0$ and unit vector $\mathbf{u}$. Then the one-variable function $q(s)=Q(\Delta\mathbf{x})=Q(s\mathbf{u})$ is a polynomial of degree n.

Let $p_{n,0}(s)=P_{n,\mathbf{a}}(s\mathbf{u})$; then $p_{n,0}(s)$ is the Taylor polynomial of degree n for $F(s)=f(\mathbf{a}+s\mathbf{u})$. Therefore, $p_{n,0}(s)$ and $q(s)$ differ at least in the term of degree k, implying

$$F(s)-q(s)\ne O(k+1)$$

(as functions of s) by Theorem 3.14. Because $F(s)=f(\mathbf{a}+\Delta\mathbf{x})$ and $q(s)=Q(\Delta\mathbf{x})$, we have

$$f(\mathbf{a}+\Delta\mathbf{x}) - Q(\Delta\mathbf{x}) \ne O(k+1)$$

as functions of $\Delta\mathbf{x}$. □

The Taylor polynomial of certain products

In certain circumstances, it is possible to construct the Taylor polynomial more directly, without evaluating a multitude of partial derviatives. For example, if $f(x,y)$ is a product in which the variables are separated, $f(x,y)=g(x)\,h(y)$, we can just multiply together the Taylor polynomials for the individual factors g and h. To illustrate, we construct the 4th-degree polynomial for $\mathrm{e}^x\cos y$ at $(x,y)=(0,0)$ from

$$\mathrm{e}^x = 1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\frac{x^4}{4!}+O(5), \quad \cos y = 1-\frac{y^2}{2}+\frac{y^4}{4!}+O(6).$$

Now just distribute the terms of $\cos y$ over the terms of e^x:

$$\begin{aligned}
e^x \cos y &= \left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+O(5)\right)\times 1 \\
&\quad -\left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+O(5)\right)\times \frac{y^2}{2} \\
&\quad +\left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+O(5)\right)\times \frac{y^4}{24} \\
&\quad +\left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}+O(5)\right)\times O(6) \\
&= 1+x+\frac{x^2}{2}+\frac{x^3}{6}+\frac{x^4}{24}-\frac{y^2}{2}-\frac{xy^2}{2}-\frac{x^2y^2}{4}+\frac{y^4}{24}+O(5) \\
&= 1+x+\frac{x^2-y^2}{2!}+\frac{x^3-3xy^2}{3!}+\frac{x^4-6x^2y^2+y^4}{4!}+O(5).
\end{aligned}$$

All of the individual products (e.g., $x^3y^2/12$) that do not appear explicitly in the last two lines have been absorbed into the symbol $O(5)$ because they vanish at least to order 5. You should check that this agrees with the definition of $P_{4,(0,0)}(x,y)$ for $e^x \cos y$; see Exercise 3.25.

Taylor polynomials for vector-valued functions

The last possibility we need to consider is a nonlinear map $\mathbf{f} : V^p \to \mathbb{R}^q$, where V^p is an open set in $\mathbb{R}^p$,

$$\mathbf{f} : \begin{cases} x_1 = f_1(v_1, \ldots, v_p), \\ x_2 = f_2(v_1, \ldots, v_p), \\ \quad\vdots \\ x_q = f_q(v_1, \ldots, v_p). \end{cases}$$

This is a vector-valued function, $\mathbf{x} = \mathbf{f}(\mathbf{v})$, and the Taylor polynomial of degree n at $\mathbf{v} = \mathbf{a}$ for $\mathbf{f}$ is just the vector of Taylor polynomials for the individual component functions $f_i(\mathbf{v})$. That is, let $P_{i;n,\mathbf{a}}(\Delta\mathbf{v})$ be the Taylor polynomial of degree n for $f_i(\mathbf{v})$ at $\mathbf{v} = \mathbf{a}$. Then the polynomial map $\mathbf{P}_{n,\mathbf{a}} : \mathbb{R}^p \to \mathbb{R}^q$,

$$\mathbf{P}_{n,\mathbf{a}} : \begin{cases} x_1 = P_{1;n,\mathbf{a}}(\Delta v_1, \ldots, \Delta v_p), \\ x_2 = P_{2;n,\mathbf{a}}(\Delta v_1, \ldots, \Delta v_p), \\ \quad\vdots \\ x_q = P_{q;n,\mathbf{a}}(\Delta v_1, \ldots, \Delta v_p), \end{cases}$$

is the Taylor polynomial for the map $\mathbf{f}$ at the point $\mathbf{v} = \mathbf{a}$. Likewise, the remainder is the vector of the corresponding remainder functions $x_i = R_{i;n,\mathbf{a}}(\Delta\mathbf{v})$. It is the map $\mathbf{R}_{n,\mathbf{a}} : V^p_{\mathbf{0}} \to \mathbb{R}^q$, where $V^p_{\mathbf{0}}$ is a suitable open neighborhood of $\mathbf{0}$:

$$\mathbf{R}_{n,\mathbf{a}}:\begin{cases} x_1 = R_{1;n,\mathbf{a}}(\Delta v_1,\dots,\Delta v_p),\\ x_2 = R_{2;n,\mathbf{a}}(\Delta v_1,\dots,\Delta v_p),\\ \quad\vdots \\ x_q = R_{q;n,\mathbf{a}}(\Delta v_1,\dots,\Delta v_p). \end{cases}$$

Taylor's formula with remainder

Taylor's formula then holds for the maps themselves:

$$\mathbf{f}(\mathbf{a}+\Delta\mathbf{v}) = \mathbf{P}_{n,\mathbf{a}}(\Delta\mathbf{v}) + \mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v}).$$

We can even describe the order of vanishing of the remainder. Suppose $\mathbf{\Phi}: T^p \to \mathbb{R}^q$ is a vector-valued function that vanishes at the origin: $\mathbf{\Phi}(\mathbf{0}) = \mathbf{0}$.

Order of vanishing of a vector-valued function

Definition 3.14 *We say the function* $\mathbf{\Phi}(\mathbf{t})$ ***vanishes to order greater than p**, and write* $\mathbf{\Phi}(\mathbf{t}) = \boldsymbol{o}(\boldsymbol{p})$, *if*

$$\lim_{\mathbf{t}\to\mathbf{0}} \frac{\|\mathbf{\Phi}(\mathbf{t})\|}{\|\mathbf{t}\|^p} = 0.$$

Definition 3.15 *We say* $\mathbf{\Phi}(\mathbf{t})$ ***vanishes at least to order p**, and write* $\mathbf{\Phi}(\mathbf{t}) = \boldsymbol{O}(\boldsymbol{p})$, *if there are positive constants* δ, C *for which* $\|\mathbf{\Phi}(\mathbf{t})\| \le C\|\mathbf{t}\|^p$ *when* $\|\mathbf{t}\| < \delta$. *Otherwise, we say* $\mathbf{\Phi}(\mathbf{t})$ ***fails to vanish to order p**, and write* $\mathbf{\Phi}(\mathbf{t}) \neq \boldsymbol{O}(\boldsymbol{p})$.

Theorem 3.23. $\mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v}) = \boldsymbol{O}(n+1)$.

Proof. We must show there are positive numbers δ, C for which

$$\|\mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v})\| \le C\|\Delta\mathbf{v}\|^{n+1} \quad \text{when} \quad \|\Delta\mathbf{v}\| < \delta.$$

Each component of $\mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v})$ is just a real-valued function, so we know it vanishes at least to order $n+1$ (Taylor's theorem for real-valued functions, Theorem 3.18, and Corollary 3.21). Hence, for each $i = 1, \dots, q$, there are positive numbers δ_i, C_i for which

$$|R_{i;n,\mathbf{a}}(\Delta\mathbf{v})| \le C_i\|\Delta\mathbf{v}\|^{n+1} \quad \text{when} \quad \|\Delta\mathbf{v}\| < \delta_i.$$

All the inequalities remain true when we take $\|\Delta\mathbf{v}\| < \delta$, where δ is the smallest of $\delta_1, \dots, \delta_q$.

For the magnitude of the vector-valued function $\mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v})$ we have

$$\begin{aligned} \|\mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v})\|^2 &= |R_{1;n,\mathbf{a}}(\Delta\mathbf{v})|^2 + \cdots + |R_{q;n,\mathbf{a}}(\Delta\mathbf{v})|^2 \\ &\le C_1^2\|\Delta\mathbf{v}\|^{2(n+1)} + \cdots + C_q^2\|\Delta\mathbf{v}\|^{2(n+1)} \end{aligned}$$

when $\|\Delta\mathbf{v}\| < \delta$. Therefore, if we set $C = \sqrt{C_1^2 + \cdots C_q^2}$, then

$$\|\mathbf{R}_{n,\mathbf{a}}(\Delta\mathbf{v})\| \le C\|\Delta\mathbf{v}\|^{n+1}.$$

□

The Taylor polynomial map of degree 1

For our future work, the Taylor polynomial map of degree 1 is the most important. In terms of components, it is

$$\mathbf{P}_{1,\mathbf{a}}:\begin{cases} x_1 = f_1(\mathbf{a}) + \dfrac{\partial f_1}{\partial v_1}(\mathbf{a})\,\Delta v_1 + \cdots + \dfrac{\partial f_1}{\partial v_p}(\mathbf{a})\,\Delta v_p, \\ x_2 = f_2(\mathbf{a}) + \dfrac{\partial f_2}{\partial v_1}(\mathbf{a})\,\Delta v_1 + \cdots + \dfrac{\partial f_2}{\partial v_p}(\mathbf{a})\,\Delta v_p, \\ \quad\vdots \\ x_q = f_q(\mathbf{a}) + \dfrac{\partial f_q}{\partial v_1}(\mathbf{a})\,\Delta v_1 + \cdots + \dfrac{\partial f_q}{\partial v_p}(\mathbf{a})\,\Delta v_p. \end{cases}$$

The derivative of $\mathbf{f}$

The initial constant terms are the components of the vector $\mathbf{f}(\mathbf{a})$. The remaining terms are linear in $\Delta v_1, \dots, \Delta v_p$; they are naturally represented by a linear map acting on the vector $\Delta\mathbf{v}$:

$$\mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{v}) = \begin{pmatrix} \dfrac{\partial f_1}{\partial v_1}(\mathbf{a}) & \cdots & \dfrac{\partial f_1}{\partial v_p}(\mathbf{a}) \\ \vdots & & \vdots \\ \dfrac{\partial f_q}{\partial v_1}(\mathbf{a}) & \cdots & \dfrac{\partial f_q}{\partial v_p}(\mathbf{a}) \end{pmatrix} \begin{pmatrix} \Delta v_1 \\ \vdots \\ \Delta v_p \end{pmatrix}.$$

Definition 3.16 *The* ***derivative*** *of the map* $\mathbf{f} : V^p \to \mathbb{R}^q$ *at* $\mathbf{a}$ *is the linear map* $\mathrm{d}\mathbf{f}_{\mathbf{a}} : \mathbb{R}^p \to \mathbb{R}^q$ *that is represented by the* $q \times p$ *matrix with components* $\partial f_i/\partial v_j(\mathbf{a})$.

In terms of the derivative, the Taylor polynomial map of degree 1 for $\mathbf{f}$ at $\mathbf{a}$ is

$$\mathbf{P}_{1,\mathbf{a}}(\Delta\mathbf{v}) = \mathbf{f}(\mathbf{a}) + \mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{v}),$$

and Taylor's formula is

$$\mathbf{f}(\mathbf{a} + \Delta\mathbf{v}) = \mathbf{f}(\mathbf{a}) + \mathrm{d}\mathbf{f}_{\mathbf{a}}(\Delta\mathbf{v}) + \boldsymbol{O}(2).$$

In the next chapter we study in detail how $\mathrm{d}\mathbf{f}_{\mathbf{a}}$ approximates $\mathbf{f}$ near $\mathbf{a}$.

Examples

Here are two examples of Taylor approximations to maps. The first map is the polar coordinate change

$$\mathbf{f}:\begin{cases} x = r\cos\theta, \\ y = r\sin\theta. \end{cases}$$

At the point $(r,\theta) = (r_0, 0)$ (so $\Delta r = r - r_0$, $\Delta\theta = \theta$), the Taylor polynomial map of degree 3 is

$$\mathbf{f}:\begin{cases} x = r_0 + \Delta r - \dfrac{r_0}{2}(\Delta\theta)^2 - \dfrac{1}{2}(\Delta r)(\Delta\theta)^2 + O(4), \\ y = r_0\,\Delta\theta + (\Delta r)(\Delta\theta) - \dfrac{r_0}{6}(\Delta\theta)^3 + O(4). \end{cases}$$

Notice that the polynomial terms are just the products of $r = r_0 + \Delta r$ with the familiar Taylor polynomials for the cosine and sine functions.

The second map is

$$\mathbf{g}: \begin{cases} x = u^3 - 3uv^2, \\ y = 3u^2v - v^3. \end{cases}$$

The derivative of $\mathbf{g}$ at the point $\mathbf{a} = (a,b)$ is given by the matrix

$$d\mathbf{g}_{(a,b)} = \begin{pmatrix} 3u^2 - 3v^2 & -6uv \\ 6uv & 3u^2 - 3v^2 \end{pmatrix}\Bigg|_{(u,v)=(a,b)} = \begin{pmatrix} 3a^2 - 3b^2 & -6ab \\ 6ab & 3a^2 - 3b^2 \end{pmatrix}.$$

The determinant of $d\mathbf{g}_{\mathbf{a}}$ has the simple form

$$\det(d\mathbf{g}_{\mathbf{a}}) = (3a^2 - 3b^2)^2 + (6ab)^2 = 9a^4 - 18a^2b^2 + 9b^4 + 36a^2b^2 = 9(a^2+b^2)^2;$$

thus $d\mathbf{g}_{\mathbf{a}}$ is invertible for all $\mathbf{a} \neq \mathbf{0}$.

Exercises

3.1. Determine the mean value of each of the following functions on the given domain.

a. $f(x) = x^n$ on $[0,1]$.
b. $f(x) = \sin x$ on $[0,\pi]$.
c. $f(x,y) = x^2 + y^2$ on $x^2 + y^2 \leq 1$. (Suggestion: Use polar coordinates.)

3.2. a. Let R be the rectangle $[a,b] \times [c,d]$ in the (x,y)-plane. Determine the coordinates (ξ,η) of the point at the center of R.

b. Show that the mean value of $f(x,y) = \alpha x + \beta y + \gamma$ on R is the value $f(\xi,\eta) = \alpha\xi + \beta\eta + \gamma$ of f at the center of the rectangle.

3.3. Show that the mean value of a linear function on a circular disk in the (x,y)-plane is its value at the center of the disk.

3.4. Find c in $[0,1]$ so that $\displaystyle\int_0^1 x^n\,dx = c^n$.

3.5. Find c in $[0,\pi]$ so that $\displaystyle\int_0^\pi \sin x\,dx = \pi \sin c$.

3.6. Assume $0 < a < b$. Find c for which

$$\int_a^b x^n\,dx = c^n(b-a),$$

and show that c lies between a and b.

3.7. Find a point (α,β) in the unit disk for which

$$\iint\limits_{x^2+y^2\leq 1} (x^2+y^2)\,dx\,dy = \pi(\alpha^2+\beta^2).$$

3.8. a. Let $f(x) = \sqrt{r^2 - x^2}$, $-r \leq x \leq r$. Show that

$$\int_{-r}^{r} f(x)\,dx = 2r\,f(c),$$

where $c = \pm r\sqrt{1-\pi^2/16}$. (Suggestion: Evaluate the integral and use that value to find c.)

b. Sketch the graph $y = f(x)$ on the interval $-r \leq x \leq r$. In the sketch, mark one of the points c and sketch the horizontal graph $y = f(c)$ to make a rectangle over the interval $[-r,r]$. Does this rectangle appear to contain the same area as the semicircular graph of $f(x)$ over $[-r,r]$? Is this what you expect?

3.9. a. Prove the generalized integral law of the mean when the condition $g(x) \geq 0$ on $[a,b]$ has been changed to $g(x) \leq 0$.

b. Suppose $f(x) = g(x) = x$. Show that there is no c in the interval $[-1,1]$ for which

$$\int_{-1}^{1} f(x)g(x)\,dx = f(c)\int_{-1}^{1} g(x)\,dx.$$

Why does this not contradict the generalized integral law of the mean?

3.10. Prove the law of the mean for double integrals (Theorem 3.7). Why does the domain D have to be connected?

3.11. a. Obtain estimates for the numerical values of $\sqrt{102}$ and $\sqrt{120}$ using the Taylor polynomials $P_{n,100}(\Delta x)$ of degree $n = 2$, 3, 4, and 5 for $f(x) = \sqrt{x}$ at $a = 100$. Use these values to verify that, for each n, the error estimating $\sqrt{102}$ is only about $1/10^{n+1}$ times the size of the error estimating $\sqrt{120}$.

b. For each $n = 2$, 3, 4, and 5, sketch the graph of the remainder function $y = R_{n,100}(\Delta x) = \sqrt{100+\Delta x} - P_{n,100}(\Delta x)$ on the interval $-1 \leq \Delta x \leq 1$. How does your graph indicate that $R_{n,100}(\Delta x) = O(n+1)$?

c. For each $n = 2$, 3, 4, and 5, there is a C_n for which

$$R_{n,100}(\Delta x) = \sqrt{100+\Delta x} - P_{n,100}(\Delta x) \approx C_n(\Delta x)^{n+1}.$$

Determine C_n and sketch $R_{n,100}(\Delta x)$ and $C_n(\Delta x)^{n+1}$ together on the same axes to indicate that $R_{n,100}(\Delta x) = O(n+1)$. In each case, take $-1 \leq \Delta x \leq 1$.

3.12. a. Construct the Taylor polynomials $P_{n,1}(\Delta x)$ centered at $a = 1$ for $f(x) = \ln x$; take $n = 1$, 2, 3, and 4.

b. Obtain estimates for $\ln 1.02$ and $\ln 1.2$ using the four different Taylor polynomials you found in part (a), and determine the error in each of these estimates. Is the error that $P_{n,1}$ makes for $\ln 1.02$ only about $1/10^{n+1}$ times the size of the error the same polynomial makes for $\ln 1.2$? Explain.

c. For each $n = 1, 2, 3$, and 4, sketch the graph of the function $y = R_{n,1}(\Delta x) = \ln(1+\Delta x) - P_{n,1}(\Delta x)$ on the interval $-0.3 \le \Delta x \le 0.3$. Does your graph demonstrate that $R_{n,1}(\Delta x) = O(n+1)$? How, or why not?

3.13. Prove the induction step

$$\begin{aligned} R_{k,a}(\Delta x) &= \frac{(\Delta x)^{k+1}}{k!}\int_0^1 f^{(k+1)}(a+t\Delta x)(1-t)^k\,dt \\ &= \frac{f^{(k+1)}(a)}{(k+1)!}(\Delta x)^{k+1} + \frac{(\Delta x)^{k+2}}{(k+1)!}\int_0^1 f^{(k+2)}(a+t\Delta x)(1-t)^{k+1}\,dt \\ &= \frac{f^{(k+1)}(a)}{(k+1)!}(\Delta x)^{k+1} + R_{k+1,a}(\Delta x) \end{aligned}$$

in the proof of Taylor's theorem.

3.14. Prove l'Hôpital's rule in the following form. Suppose $f(a) = f'(a) = \cdots = f^{(n-1)}(a) = 0$, $g(a) = g'(a) = \cdots = g^{(n-1)}(a) = 0$, and either $f^{(n)}(a) \neq 0$ or $g^{(n)}(a) \neq 0$ (or both); then

$$\lim_{x\to a}\frac{f(x)}{g(x)} = \begin{cases} \infty & \text{if } g^{(n)}(a) = 0, \\ \dfrac{f^{(n)}(a)}{g^{(n)}(a)} & \text{otherwise.} \end{cases}$$

(Suggestion: Use Taylor's formula with Lagrange's form of the remainder, for both $f(a+\Delta x)$ and $g(a+\Delta x)$.)

3.15. Let $\varphi(t) = t^\alpha$, $1 < \alpha < 2$. Show that $\varphi(t) = o(1)$ but $\varphi(t) \neq O(2)$.

3.16. Use the fact that e^x grows faster than any positive power of x (i.e., $x^p/e^x \to 0$ as $x \to +\infty$ for any $p > 0$) to show that $\psi(u) = \exp(-1/|u|)$ vanishes to order greater than p for any $p > 0$. It follows that we can define $\psi(0) = 0$. Sketch the graph of $t = \psi(u)$ and determine the image of ψ on the t-axis.

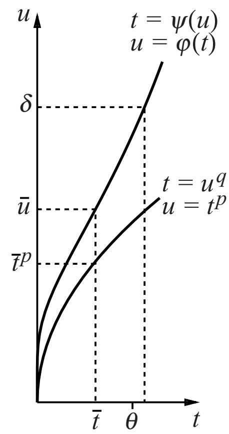
3.17. Show that the condition $\varphi(t) = o(p)$ can be expressed in the following way. Given any $\varepsilon > 0$, there is a $\delta > 0$ so that

$$|t| < \delta \implies |\varphi(t)| \le \varepsilon |t|^p.$$

(Suggestion: The fact that $\varphi(t)/t^p \to 0$ as $t \to 0$ means that $|\varphi(t)/t^p|$ can be made smaller than any preassigned $\varepsilon > 0$ by making $|t|$ sufficiently small, i.e., by making $|t| < \delta$ for some suitable δ.)

Note: Although this formulation of "little oh" may seem more cumbersome, it has the advantage of avoiding using a quotient, a useful feature in some of our later work (cf. p. 133). This formulation is also more like our definition of "big oh."

3.18. Let $\varphi(t) = -1/\ln t$, $0 < t < 1$, $\varphi(0) = 0$; the graph of φ appears in the margin. The goal of this exercise is to show that $\varphi(|t|) = -1/\ln|t|$ vanishes to order less than p for any $p > 0$. This is true if, for a given $p > 0$ and for every θ sufficiently small, $t^p < \varphi(t)$ for every $0 < t < \theta$ (the contrapositive of Exercise 3.17).



a. Show that $t = \psi(u)$ (Exercise 3.16) is invertible on $u \geq 0$, and show that $u = \varphi(t)$ is the inverse.

b. Fix $p > 0$, set $q = 1/p$, and choose $\delta > 0$ so that $|u| < \delta \implies \psi(u) < |u|^q$ (as provided by Exercise 3.17 with any $\varepsilon < 1$). Now take any $0 < \theta < \psi(\delta)$ and let $0 < \bar{t} < \theta$ be arbitrary. Set $\bar{u} = \varphi(\bar{t})$ and show that $\bar{t}^p < \bar{u} = \varphi(\bar{t})$.

3.19. Let $f(x) = \sqrt{x}$; show that

$$f^{(k+1)}(x) = \frac{\pm 1 \cdot 3 \cdots (2k-1)}{2^{k+1} x^{k+1/2}}.$$

3.20. In the microscope equation, $\Delta y \approx f'(a)\,\Delta x$, the nature of the approximation is unclear. Show that the microscope equation has the more explicit form $\Delta y = f'(a)\,\Delta x + O(2)$; in words, "$\Delta y$ agrees with $f'(a)\,\Delta x$ at least up to order 2 in Δx."

3.21. Adapt the proof of Theorem 3.14 to prove Corollary 3.15.

3.22. The purpose of this exercise is to show that the extent to which a polynomial approximates a given function near a given point depends on the extent to which it matches the Taylor polynomial constructed at that point (cf. Theorem 3.14 and Corollary 3.15).

a. Show that $P(x) = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$ is the Taylor polynomial of degree 3 at $x = 0$ for the function e^x.

b. Sketch the graph of $y = R(x) = e^x - P(x)$ in a small neighborhood of $x = 0$ to demonstrate that $R(x) = O(4)$, as required by Taylor's theorem.

c. Sketch the graph of $y = V_1(x) = e^x - (1 + x + \frac{1}{2}x^2 + \frac{1}{5}x^3)$ in a small neighborhood of $x = 0$. Determine the value of p for which $V_1(x) = O(p)$ and $V_1(x) \neq O(p+1)$.

d. Sketch the graph of $y = V_2(x) = e^x - (1 + x + \frac{1}{3}x^2 + \frac{1}{6}x^3)$ in a small neighborhood of $x = 0$. Determine the value of p for which $V_2(x) = O(p)$ and $V_2(x) \neq O(p+1)$.

e. Sketch the graph of $y = V_3(x) = e^x - (1 + 1.1x + \frac{1}{2}x^2 + \frac{1}{6}x^3)$ in a small neighborhood of $x = 0$. Determine the value of p for which $V_3(x) = O(p)$ and $V_3(x) \neq O(p+1)$.

f. Sketch the graph of $y = V_4(x) = e^x - (1.1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3)$ in a small neighborhood of $x = 0$. Determine the value of p for which $V_4(x) = O(p)$ and $V_4(x) \neq O(p+1)$.

3.23. Write the Taylor polynomial of degree 2 for the given function at the given point.

a. $e^x \sin y$ at $(0,0)$
b. $\cos x \cos y$ at $(0,\pi/2)$
c. $x^3 - 3x + y^2$ at $(-1,0)$
d. $\ln(x^2+y^2)$ at $(1,0)$
e. xyz at $(1,-2,4)$
f. $1-\cos\theta + \frac{1}{2}v^2$ at $(\pi,0)$

3.24. Write the Taylor polynomial of degree 4 for $(x^2+y^2)^2-(x^2+y^2)$ at the point $(x,y)=(1/2,1/2)$.

3.25. Show that the Taylor polynomial of degree 4 for $e^x \cos y$ at $(x,y)=(0,0)$, as obtained from the definition, agrees with the computation done on page 96.

3.26. Write out in words what "$O(p)\cdot O(q) = O(p+q)$" means, and prove it.

3.27. Construct the Taylor polynomial of degree 2 centered at the point $(\rho,\theta,\varphi) = (\rho_0,\pi/2,0)$ for the *spherical coordinate change*

$$\mathbf{s}: \begin{cases} x = \rho\cos\theta\cos\varphi, \\ y = \rho\sin\theta\cos\varphi, \\ z = \rho\sin\varphi; \end{cases} \qquad \begin{array}{c} -\pi \le \theta \le \pi, \\ -\pi/2 \le \varphi \le \pi/2. \end{array}$$

3.28. a. Suppose $\mathbf{L}:\mathbb{R}^p \to \mathbb{R}^q$ is linear; show $\mathbf{L}(\Delta\mathbf{u})$ vanishes at least to first order in $\Delta\mathbf{u}$. In fact, show there is a positive number C for which $\|\mathbf{L}(\Delta\mathbf{u})\| \le C\|\Delta\mathbf{u}\|$ for all $\Delta\mathbf{u}$.

b. The smallest number C for which this inequality holds is called the **norm** of the linear map $\mathbf{L}$, written $|||\mathbf{L}|||$. It follows that $\|\mathbf{L}(\Delta\mathbf{u})\| \le |||\mathbf{L}|||\,\|\Delta\mathbf{u}\|$ for all $\Delta\mathbf{u}$. Show that

$$|||\mathbf{L}||| = \max_{\|\Delta\mathbf{u}\|=1} \|\mathbf{L}(\Delta\mathbf{u})\|.$$

c. Suppose the linear map $\mathbf{L}:\mathbb{R}^p \to \mathbb{R}^p : \Delta\mathbf{u} \mapsto \Delta\mathbf{x}$ is invertible. Show that $\mathbf{L}$ and $\mathbf{L}^{-1}$vanish exactly to order 1 in the sense that there are bounding constants $0 < A_1 \le A_2$, $0 < B_1 \le B_2$ for which

$$A_1 \le \frac{\|\mathbf{L}(\Delta\mathbf{u})\|}{\|\Delta\mathbf{u}\|} \le A_2 \quad\text{and}\quad B_1 \le \frac{\|\mathbf{L}^{-1}(\Delta\mathbf{x})\|}{\|\Delta\mathbf{x}\|} \le B_2$$

for all $\Delta\mathbf{u}, \Delta\mathbf{x} \ne \mathbf{0}$. (This is an adaptation of Definition 3.4 to multivariable functions.)

d. Show that we can take $B_1 = 1/A_2$, $B_2 = 1/A_1$ in part (b).