Taylor's Theorem

Recall that when a function f is differentiable at a number c, then we can approximate f near c with a linear function that has both the same value and derivative as f does at c:

$$f(x) \approx L(x) = f(c) + f'(c)\,(x - c);$$

here, L is a linear function, $L(c) = f(c)$, and $L'(c) = f'(c)$. This is actually only the beginning (well, slightly after the beginning) of a whole sequence of approximations, each (typically) better than the one before it:

$$\begin{aligned}
f(x) &\approx P_0(x) = f(c);\\
f(x) &\approx P_1(x) = f(c) + f'(c)\,(x - c);\\
f(x) &\approx P_2(x) = f(c) + f'(c)\,(x - c) + \frac{1}{2}\,f''(c)\,(x - c)^2;\\
f(x) &\approx P_3(x) = f(c) + f'(c)\,(x - c) + \frac{1}{2}\,f''(c)\,(x - c)^2 + \frac{1}{6}\,f'''(c)\,(x - c)^3;\\
&\ \ \vdots
\end{aligned}$$

(The function that used to be called L is now called P_1.) The general form of this is

$$f(x) \approx P_k(x) = \sum_{n=0}^{k} \frac{1}{n!} f^{(n)}(c)(x - c)^n.$$

(Recall that $f^{(n)}$ is the nth derivative of f.) Of course, f must be differentiable at c at least k times for P_k to make sense.

The function P_k is the **Taylor polynomial** of f at c of order k. The Taylor polynomial of f at 0 of order k is also called the **Maclaurin polynomial** of f of order k. This terminology is standard (except for some variations in the phrase 'of order' that you may see); however, the notation P_k is *not* standard (and in principle it ought to mention f and c as well as k). Strictly speaking, Taylor polynomials are polynomial *functions* rather than polynomials as such (which are simply algebraic expressions without any variable picked out); otherwise, you'd have to mention the variable x as well.

Notice that a Taylor polynomial P_k of order k is a polynomial function of degree at most k. (The degree is normally exactly k, but it's smaller if $f^{(k)}(c)$ happens to be 0.) Also, the nth derivative of P_k at c agrees with that of f, if $n \leq k$; that is,

$$P_k^{(n)}(c) = f^{(n)}(c)$$

if $n \leq k$. (On the other hand, if $n > k$, then $P_k^{(n)}(c) = 0$, which is always the case for a higher-order derivative of a polynomial function when the order of the derivative is greater than the degree of the polynomial.) The Taylor polynomial of f at c of order k is the *only* polynomial function of degree at most k whose derivatives at c of order up to k agree with those of f.

Since polynomials are easy to work with, it's convenient to make approximations like these. But in practice, it's also important to know *how good* the approximations are. Since these approximations are based on the behaviour of f at c, we can really only expect them to be good when $x \approx c$. So one way to say that these approximations work is to say that $P_k(x)$ approaches $f(x)$ (or more formally that the error of the approximation, $|P_k(x) - f(x)|$, approaches 0) as x approaches c. This is true for $k = 0$ if f is continuous at c, and for $k > 0$ if f is differentiable k times at c. But in fact, the higher-order Taylor polynomials satisfy a stronger condition:

$$\lim_{x \to c} \frac{|P_k(x) - f(x)|}{|x - c|^k} = 0,$$

which is called (one version of) **Taylor's Theorem**. As x approaches c, of course $|x - c|$ approaches zero, so dividing by $|x - c|$ would tend to make a positive quantity larger. So P_k is such a good approximation to f that the error not only approaches zero but still approaches zero even after dividing by $|x - c|$ multiple times.

When investigating these questions, it's helpful to change perspective slightly. Write R_k for $f - P_k$, the Taylor **remainder** of f at c of order k. Then the statement above, showing what a good approximation P_k is, becomes

$$\lim_{x \to c} \frac{|R_k(x)|}{|x - c|^k} = 0.$$

This is good to know, but it may not really be enough; it tells us that moving x close to c will make the approximation better, and very quickly; roughly, when x is already close to c, then moving it twice as close will make the approximation 2^k times better, or you can make the approximation one decimal digit more accurate by moving x only $\sqrt[k]{10}$ times as close. However, this doesn't tell us how accurate the approximation was to start with, nor how close x has to be for this method of improving the approximation to start working.

We can get better results if f is differentiable one more time ($k + 1$ times, not just k times) and near c (not just at c). This strong version of Taylor's Theorem says that

$$R_k(x) = \int_{t=0}^{1} (1 - t)^k f^{(k+1)}(c - ct + xt)(x - c)^{k+1} \, \mathrm{d}t,$$

as long as f is continuously differentiable $k + 1$ times (at least) between c and x. (The integral here may exist even if f is not *continuously* differentiable $k + 1$ times, but then the value of this integral might not be the remainder.) To be more explicit, here is the statement for the first few values of k:

$$\begin{aligned} f(x) &= f(c) + \int_{t=0}^{1} f'(c - ct + xt)(x - c) \, \mathrm{d}t \\ &= f(c) + f'(c)(x - c) + \int_{t=0}^{1} (1 - t) f''(c - ct + xt)(x - c)^2 \, \mathrm{d}t \\ &= f(c) + f'(c)(x - c) + \frac{1}{2} f''(c)(x - c)^2 + \frac{1}{2} \int_{t=0}^{1} (1 - t)^2 f'''(c - ct + xt)(x - c)^3 \, \mathrm{d}t \\ &\vdots \end{aligned}$$

These statements may be proved by repeated application of integration by parts (and the Fundamental Theorem of Calculus, which is why $f^{(k+1)}$ must not only exist but also be continuous). For example, to prove the first one, start on the right-hand side and integrate by parts using $u = 1$ (so $\mathrm{d}u = 0$) and $v = f(c - ct + xt)$ (so $\mathrm{d}v = f'(c - ct + xt)(x - c) \, \mathrm{d}t$):

$$\begin{aligned} f(c) + \int_{t=0}^{1} 1 f'(c - ct + xt)(x - c) \, \mathrm{d}t &= f(c) + \Big(1 f(c - ct + xt)\Big)\Big|_{t=0}^{1} - \int_{t=0}^{1} f(c - ct + xt) 0 \\ &= f(c) + f\Big(c - c(1) + x(1)\Big) - f\Big(c - c(0) + x(0)\Big) - \int_{t=0}^{1} 0 \\ &= f(c) + f(x) - f(c) - 0 = f(x). \end{aligned}$$

In general, you can prove each statement using $u = (1 - t)^k \big/ k!$ and $v = f^{(k)}(c - ct + xt)(x - c)^k$, integrating by parts, simplifying, and applying the previous statement.

For purposes of approximation, it's useless to actually work out the integral that appears here; if you knew the exact value of $f^{(k+1)}$ at all of the points between c and x, then you could probably just evaluate f at x directly. However, if there is a value M_k such that you know that $f^{(k+1)}$ never has an absolute value greater than M at any point between c and x, then you can use M_k to get a bound on the remainder:

$$|R_k| \leq \frac{M_k}{(k + 1)!} |x - c|^{k+1}.$$

The reason for this is that we know that R_k is exactly the integral that appeared in the full version of the theorem, and we can bound its absolute value using the bound on its integrand:

$$\begin{aligned}|R_k| &= \left|\frac{1}{k!}\int_{t=0}^{1}(1-t)^k f^{(k+1)}(c-ct+xt)h^{k+1}\,\mathrm{d}t\right| \le \frac{1}{k!}\int_{t=0}^{1}(1-t)^k \left|f^{(k+1)}(c-ct+xt)\right| |x-c|^{k+1}\,\mathrm{d}t \\ &\le \frac{1}{k!}\int_{t=0}^{1}(1-t)^k M_k\,|x-c|^{k+1}\,\mathrm{d}t = \frac{M_k}{k!}\,|x-c|^{k+1}\int_{t=0}^{1}(1-t)^k\,\mathrm{d}t \\ &= \frac{M_k}{k!}\,|x-c|^{k+1}\frac{1}{k+1} = \frac{M_k}{(k+1)!}\,|x-c|^{k+1}.\end{aligned}$$

To be more specific:

$$|R_0| = |f(x)-f(c)| \le M_0\,|x-c|$$

if $|f'|$ is never greater than M_0 between c and x,

$$|R_1| = |f(x)-(f(c)+f'(c)(x-c))| \le \frac{1}{2}M_1\,|x-c|^2$$

if $|f''|$ is never greater than M_1 between c and x,

$$|R_2| = \left|f(x)-\left(f(c)+f'(c)h+\frac{1}{2}f''(c)(x-c)^2\right)\right| \le \frac{1}{6}M_2\,|x-c|^3$$

if $|f'''|$ is never greater than M_2 between c and x, etc.

Finally, of course, we can extend from polynomials to power series and get the **Taylor series** of f at c:

$$P_\infty(x) = \sum_{n=0}^{\infty}\frac{1}{n!}f^{(n)}(c)(x-c)^n.$$

(When $c=0$, this is the **Maclaurin series** of f.) This power series exists as long as f is infinitely differentiable at c, that is as long as f has derivatives of all orders at c. However, there are no theorems guaranteeing that this series converges, nor that it's anything like $f(x)$ when it does converge. We say that f is **analytic** if this series converges to $f(x)$ at least on some interval around c. Any function built out of the usual operations is analytic, as long as it's infinitely differentiable, but there are piecewise-defined functions for which this series fails to converge anywhere near c and others for which it converges to something other than $f(x)$.