10

Local approximation in one variable

10.1 The derivative as a function

When considering a differentiable function, say f, by definition we can compute the derivative of f at each point of its domain. We can then think of a function that takes as input some value x and returns the derivative of f at x. We denote this function f'.

An important consequence of this way of thinking is that we can iterate this process to yield the second $f^{(2)}$, third $f^{(3)}$ and so on derivatives of f. In general, we will denote the n-th derivative of f as $f^{(n)}$.

There are a number of examples where a legitimate differentiable function f yields a derivative function f' that is itself non-differentiable or even non-continuous. Just for illustration, the following function

$$V(x) = \begin{cases} x^2 \sin 1/x & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

known as Volterra's function, is differentiable everywhere, but it is discontinuous at a bunch of points known as the Smith-Volterra-Cantor set.

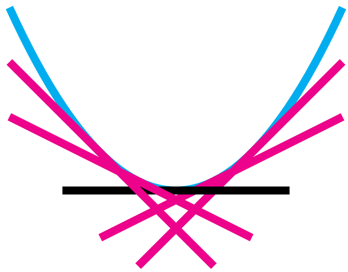
For the purpose of this course, we will consider only a class of functions known as *smooth* which admit as many higher-order derivatives as needed in the domain of the function.

10.2 Local quadratic approximation

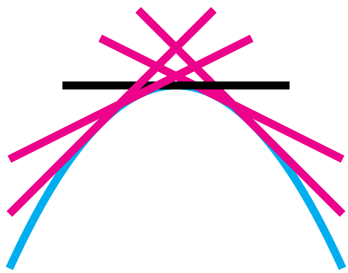
The importance of the higher-order derivatives of f is that they convey more nuances about the local shape of the function about the point of interest.

Let's consider the second derivative of f at a point a, say $f^{(2)}(a)$. What is it telling us? This is the local growth rate of the function f' at a, i.e. at what rate is the local growth rate of f changing about a. The situation in a point p where $f^{(2)}(p) > 0$ can look very different compared with the situation in another point q where $f^{(2)}(q) < 0$, even if $f'(p) = f'(q)$.

In the first case, with $f^{(2)}(p) > 0$, we may encounter a situation like the following one:



Whilst in the second case, with $f^{(2)}(q) < 0$, we may encounter a completely different situation, even if both functions have derivative equal to zero at the point of interest:



These examples suggest that a better approximation could be attained if only we could incorporate also the information conveyed by the second derivative of f. Using an affine functions we have no more room left to carry more information about the local shape of f at a certain point a.

We must use another type of function that gives us some more flexibility to satisfy the additional requirement that the second derivative of the approximation must be equal to the second derivative of f. In other words, we look for an as simple function as possible, say $Q(x)$ that satifies the following three requirements:

1. $Q(a) = f(a)$
2. $Q'(a) = f'(a)$
3. $Q^{(2)}(a) = f^{(2)}(a)$

Can we figure out a solution to the problem of the following form?

$$Q(x) = k_0 + k_1(x - a) + k_2(x - a)^2$$

Using the first requirement we get $k_0 = f(a)$. The derivative of $Q(x)$ is $Q'(x) = k_1 + 2k_2(x - a)$, therefore using the second requirement we conclude that $k_1 = f'(a)$. Finally, we can see that the second derivative of $Q(x)$ is $Q^{(2)}(a) = 2k_2$, therefore using the third requirement we conclude that $k_2 = \frac{1}{2}f^{(2)}(a)$. Our candidate function

$$Q(x) = f(a) + f'(a)(x - a) + 1/2 f^{(2)}(x - a)^2$$

is very much like the affine approximation, albeit with an additional term of degree 2. Does this additional quadratic term make the approximation better than the affine approximation in any way? The answer is yes, in a sense that is made precise in the following proposition, which we leave without proof:

Proposition 35. *If f is a smooth function at a, then*

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

In other words, the approximation error of the quadratic approximation is guaranteed to converge towards zero locally at a faster than $(x-a)^2$.

10.3 Taylor's approximation

There is no reason why we cannot extend the argument that made possible the quadratic approximation with higher-order derivatives of f. Intuitively, the more higher-order derivatives we know, the more nuanced an approximation we can build. So we can extend our search to higher-degree polynomials. Concretely, if we know all the derivatives of f at a up to order n, can we find a polynomial function $T(x)$ that satisifies the following requirements?

1. $T(a) = f(a)$
2. $T'(a) = f'(a)$
3. $T^{(2)}(a) = f^{(2)}(a)$

   $\dots$

$n+1$. $T^{(n)}(a) = f^{(n)}(a)$

Repeating the same argument as before in an iterative way, we can see that our candidate polynomial is uniquely determined by the list of requirements:

$$T(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a)(x-a)^k$$

where we adopt the convention that $f^{(0)}(a) = f(a)$ and that

$0! = 1$.

In which sense do these additional higher-order terms make the approximation better than the lower degree ones? The answer is provided by the following result, which we leave without proof:

Theorem 4 (Taylor's theorem). *If f is a smooth function at a, then*

$$f(x) = \sum_{k=0}^{n} \frac{1}{k!} f^{(k)}(a)(x-a)^k + \mathfrak{o}((x-a)^n)$$

In other words, the approximation error of the Taylor polynomial of degree n is guaranteed to converge towards zero locally at a faster than $(x-a)^n$.

10.4 Finding local extrema

A differentiable function f is said to have a *critical point* in a if $f'(a) = 0$. The collection of critical points of a differentiable function are important, because they are the only candidates where the function can reach a maximum or a minimum. In general, using the approximation results above described, we can classify all the critical points in either of three categories: maxima, minima and inflection points.

Theorem 5 (Classification of critical points). *Let a be a critical point of a smooth function f. Let*

$$n = \min\{n \mid f^{(n)}(a) \neq 0\} \geq 2.$$

The following statements hold:

- *If n is even and $f^{(n)} < 0$, then a is a local maximum.*
- *If n is even and $f^{(n)} > 0$, then a is a local minimum.*
- *If n is odd, then a is an inflection point.*

Using Taylor's theorem in the situation of the theorem, f takes the following form:

$$f(x) = f(a) + \frac{1}{n!} f^{(n)}(a)(x-a)^n + \mathfrak{o}((x-a)^n)$$

If n is even and $f^{(n)}(a) < 0$, we should justify that there is an interval about a where $f(a)$ is the maximum value that $f(x)$ takes. This would mean that $f(x) - f(a) < 0$ for $x \neq a$ in the said interval. But using the expression of f we can prove that, indeed. Since

$$\psi(x) = f(x) - f(a) = \frac{1}{n!} f^{(n)}(a)(x-a)^n + \mathfrak{o}((x-a)^n)$$

we are bound to prove that $\psi(x) < 0$ in some interval (however small) about a. Since n is even, $(x-a)^n > 0$ for $x \neq a$, so dividing by $(x-a)^n$ would not change the sign:

$$\frac{\psi(x)}{(x-a)^n} = \frac{1}{n!} f^{(n)}(a) + \frac{\mathfrak{o}((x-a)^n)}{(x-a)^n}$$

But we know that the second term converges towards zero as we get closer to a

$$\lim_{x \to a} \frac{\mathfrak{o}((x-a)^n)}{(x-a)^n} = 0$$

meaning that for values x close enough to a, we must have

$$\frac{\psi(x)}{(x-a)^n} < 0,$$

but since the sign of $\psi(x)/(x-a)^n$ must be equal to the sign of $\psi(x) = f(x) - f(a)$ the conclusion follows. The other two cases of the theorem can be justified in exactly the same way.

10.5 TL;DR

The more higher-derivatives of f at a we know, the more nuanced our understanding of the local behviour of f about a. For smooth functions Taylor's polynomial is the key mathematical object that assembles all the higher-order derivatives in a polynomial function that provides good local approximations of f, the better the higher the degree of the polynomial. With this polynomial we generalize the idea of local affine approximation. With these results at hand, we can then go about studying the critical points of f into three possible categories: maxima, minima and inflection points.