Derivatives

Part I. Precalculus

1 Functions

The definition of "function" is very "weak". Since *any* rule mapping an input to an output is a function, there are very strange examples out there. Some sophisticated tricks are needed to produce the really weird ones, but a simple example (and it's not really pathological) is

$$f(x) = \begin{cases} 1 & \text{if } x \text{ is irrational} \\ 0 & \text{if } x \text{ is rational} \end{cases}$$

It's essentially meaningless to draw the graph of f.

In practice, we want our functions to have a little more structure than just being "functions", so we can do meaningful work with them. From one side, we study specific explicitly given functions that allow more or less easy manipulation (polynomials, exponential and logarithmic functions, trigonometric functions, and so on). From the other, we try to identify features that would allow us to work profitably with more general functions. Fact of the matter is that many functions in real applications arise as solutions to problems where we can deduce many properties of the solution, but not a really explicit formula. Even when a formula is available, though, except for very simple cases, we cannot say much about the function right away. Here is where "differential calculus" enters the picture.

Since polynomials are reasonably simple to study (at least if the degree is not too high), the first goal we aim at is to define classes of functions that behave "almost like" a polynomial, allowing us to get information on their behavior by finding polynomials of low degree that behave (at least over a small range) very much like them.

2 Irrational Numbers

Implicit in the discussion above is that in almost all cases, we don't have the possibility of providing an explicit *exact* answer to our questions, but that we will make do with suitable *approximate* answers, which, for the most part, will be more than adequate.

We already should know that this is the case – and this awareness dates back to ancient Greece. In fact, consider the simple geometric problem of finding the length of the diagonal of a square whose side is of length 1. The Pythagorean Theorem tells us that the diagonal has length $\sqrt{2}$. But $\sqrt{2}$ is only a symbol: how many inches is that? If there was a fraction whose square equals 2 the answer would be straightforward, but there is no such fraction

> *Proof*: suppose $\frac{p^2}{q^2} = 2$, where p and q are integers. Suppose we have already reduced the fraction $\frac{p}{q}$ to lowest terms, so that p and q have no common factors (except the trivial factor 1). Since we have $p^2 = 2q^2$, p^2 is even, and so, necessarily, will be p. Hence $p = 2k$ for some k. Hence, $p^2 = 4k^2$, and
>
> $$q^2 = \frac{p^2}{2} = \frac{4k^2}{2} = 2k^2$$
>
> But this means that q^2, and hence q, is even too. Now, p and q would have 2 as a common factor, which we had excluded. Hence, there are no such p and q.
>
> It is easy to transfer the proof to roots of any order, of any integer, that are not exact roots. In other words, for any integer m, and any n, either $\sqrt[n]{m}$ is an integer, or it is irrational (cannot be expressed as a fraction).

Of course, we are also familiar with the celebrated example of π, the ratio of the circumference to the diameter of a circle. Approximate methods to determine π have been known at least since the time of Archimedes, but it is a deep result from the 18th Century that π is irrational.

So, what is the length of the diagonal of our square? We cannot give an absolutely exact number of inches, but, then, we don't really need to. In fact, we can approximate $\sqrt{2}$ as much as we wish. For example, stating that $\sqrt{2} \approx 1.4142$ means that the rational number 1.4142 is such that $1.4142^2 < 2$, but $1.4143^2 > 2$. So, using this value, for a side of 1 inch, we can approximate the diagonal up to 1/10,000th of an inch. That is probably enough for most cases, but if it wasn't, we can push the approximation further. For example, $\sqrt{2} \approx 1.414213562$ and we are down to the billionth of an inch, and we could go on, if we wanted to. Since any instrument we might use will have a finite precision, even if this could be a fabulously great one, we won't be limited by our inability to pin down $\sqrt{2}$ exactly.

One way of thinking of this elusive number is then as the "limiting number" towards which the sequence of approximations

$$1.4, 1.41, 1.414, 1.4142, 1.41421, \ldots \tag{1}$$

is tending, without ever reaching it – but getting closer and closer, in fact, as close as we wish, provided we have the patience to proceed far enough.

In formal, mathematical terms, this is expressed by considering the successive numbers in (1), let's call them $a_1 = 1.4, a_2 = 1.41, \ldots$, and stating that, for any value ε that we may choose (for example, $\varepsilon = 10^{-5}$), we can push our calculation so far that we will reach an n, such that from that point on ("for any $m \geq n$),

$$\left| a_m - \sqrt{2} \right| < \varepsilon$$

For example, from the values listed above, we see that $a_5 = 1.41421$, and, indeed,

$$\left| 1.41421 - \sqrt{2} \right| < 10^{-5}$$

"for any $m \geq 5$, $\left| a_m - \sqrt{2} \right| < 10^{-5}$".

Most mathematical results are very similar to this: rather than expressing exact equalities, they will allow for precise control of our approximations. The foremost tool for this is differential calculus.

3 Studying Polynomials

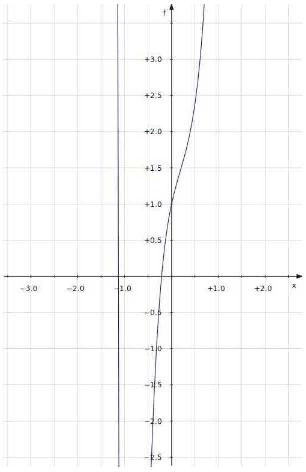
Even though polynomials are very simple functions (for example, given an exact value for x, we can compute the exact corresponding value of the polynomial via simple products and sums[1]), it is not instantaneous to figure out how the graph will look like near a given value of x, as soon as the degree is not low. We can however find out using a simple trick.

3.1 A Polynomial Near $x = 0$

This fact is well known: for a number a, such that $|a| < 1$, the higher the power a^n, the smaller its absolute value. This feature accentuates the closer a is to 0. For example, if $a = 10^{-1}$, $a^n = 10^{-n}$, and if $a = 10^{-3}$, $a^n = 10^{-3n}$. Hence, if x is close to 0, each of its successive powers becomes quickly negligible. So, if $|x|$ is sufficiently small, x^2 will be "invisible" when compared to x. Hence, a polynomial like

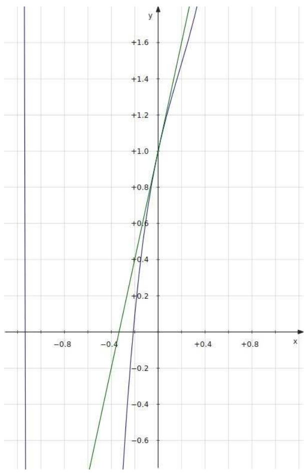
$$p(x) = 7x^{10} - 4x^9 + 3x^8 - 2x^7 + x^6 + 6x^5 - 9x^4 + 12x^3 - 5x^2 + 3x + 1$$

[1] So, for $p(x) = -3x^5 + 2x^4 - \frac{x^3}{3} + \frac{2}{3}x^2 + 5x - 1$, $p(2) = -3 \cdot 2^5 + 2 \cdot 2^4 - \frac{2^3}{3} + \frac{2}{3} \cdot 2^2 + 5 \cdot 2 - 1$

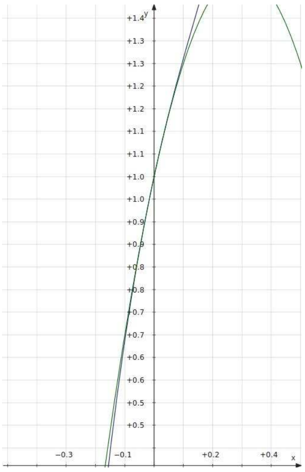


we can say

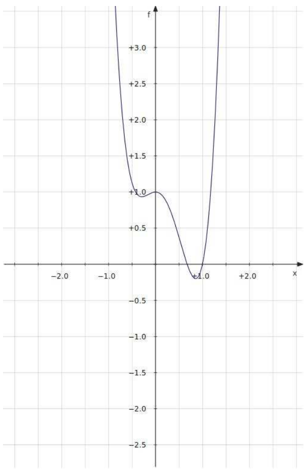
- $p(0) = 1$ (that's easy)
- If $|x|$ is really close to 0, $p(x) \approx 1 + 3x$. This means that the graph of p will look almost like the graph of the straight line $1 + 3x$. Geometrically, this means that the tangent line to the graph at $x = 0$ has slope 3, and the graph crosses the y axis going from below 1 to above 1 – it is increasing near $x = 0$, and it has approximate slope 3



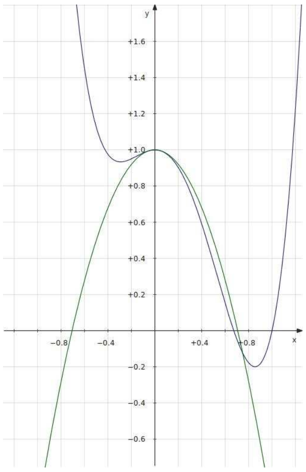
- If $|x|$ is very close to 0, $p(x) \approx 1 + 3x - 5x^2$. This means that it will look, near 0, much like the parabola $y = -5x^2 + 3x + 1$. This is easily seen to be a parabola with a graph that is open down, and a vertex at some $x > 0$. In other words, the graph is increasing near $x = 0$, but it is doing so at a slowing rate, just like the graph of our approximating parabola.



This can be pushed further, but you get the idea. An even more interesting case is when there is no linear term. For example, if $q(x) = 4x^4 - 3x^3 - 2x^2 + 1$,



we still have $q(0) = 1$, but now the first interesting observation we can make is that, for $|x|$ close enough to 0, $q(x) \approx 1 - 2x^2$. This is a parabola, open down, with vertex at $(0, 1)$. This is a maximum point for the parabola, and this allows us to say that this is also a maximum for our polynomial.



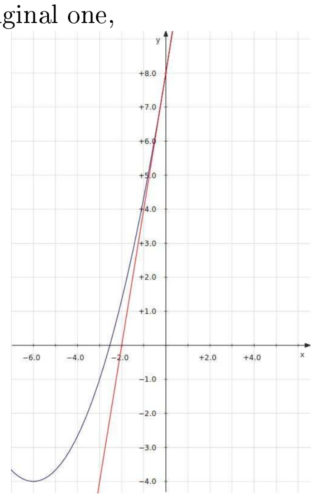
3.2 Polynomials At Any x

What if we want to learn about our polynomial near a non zero value of x? The argument above does not work, of course. However, we can take advantage of a trick we learned in Algebra and Precalculus: we can *shift our graph horizontally so that the value we are interested in gets shifted to* 0!

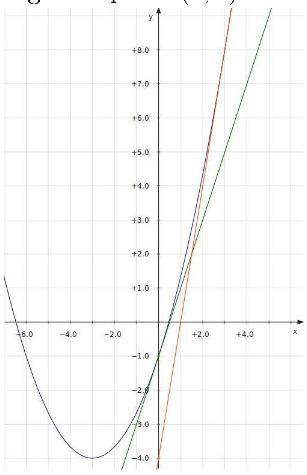
As a very simple example, think of the question "how does the tangent line to $f(x) = \frac{x^2}{3} + 2x - 1$ at $x = 3$ look like?". By the argument above, we know that the tangent line at $x = 0$ is $y = 2x - 1$, but what about $x = 3$? Well let's shift our curve to the right by 3 units: this way $x = 3$ becomes $x = 0$. As we know this is performed by substituting $x + 3$ in place of x:

$$f(x+3) = \frac{(x+3)^2}{3} + 2 \cdot (x+3) - 1 = \frac{x^2}{3} + 4x + 8$$

The tangent at $x = 0$ of this function is the same as the one at $x = 3$ of our original one,



since we simply shifted it sideways: it will be a line with slope 4, going through the point $(3, 8)$: that's $y - 8 = 4\,(x - 3) \Rightarrow y = 4x - 4$



4 How To Go Beyond Polynomials?

All the above is pure Algebra. But how could we extend this idea to functions like e^x, or $\tan x$, or even more complicated ones? Well, that was solved by the genius of Newton and Leibniz, back in the 17th Century, and is the topic of most of our class. In fact, one way to look at derivatives is as coefficients of polynomials that are "close" to our function.

Part II. Calculus

5 Notations for derivatives of a function

The most common notations for the derivative of a function f are

- f'
- Df convenient: allows for, say, $D\frac{x^2}{\sin x}$, possibly more convenient than $\left(\frac{x^2}{\sin x}\right)'$
- $\frac{df}{dx}$ very common, somewhat misleading (derivatives are not quotients), but hugely convenient. It can be made rigorous in two ways (at least). One is exotic, and we will not address that (there are good texts available on the web for that). The other is obtained by defining the "differential" of a function at a point a as the linear approximation;

$$f(x) = f(a) + df(a) + \text{small error} \tag{2}$$

so that

$$df(a) = f'(a)\,(x-a) \tag{3}$$

and, of course, it follows that

$$dx = x - a$$

so that

$$\frac{df(a)}{dx} = \frac{f'(a)(x-a)}{x-a} = f'(a)$$

Note: the "small error" in (2) is "small" in the following precise sense:

$$f\,(x) - f\,(a) = f'\,(a)\,(x-a) + \varepsilon$$

where, as $x \to a$, $\frac{\varepsilon}{x-a} \to 0$ ("ε goes to zero faster than $x - a$"). As we'll see next, (2) and (3) provide an alternate, very useful, definition of "first derivative".

6 Polynomials and "Taylor's Formula"

We already had a method to find the slope of the tangent at a point for a polynomial. Let's see how it is the same as taking the first derivative on an example.

Suppose you have

$$p(x) = 2x^3 - 3x^2 + x - 1$$

and want the slope of the tangent at $x = -1$. You may compute

$$p'(-1) = 6 \cdot (-1)^2 - 6 \cdot (-1) + 1 = 13$$

or you could

1. Shift the graph to the right by 1, and check the new linear term:

$$2\,(x-1)^3 - 3\,(x-1)^2 + (x-1) - 1 = 6x + 6x + x + \ldots = 13x + \ldots$$

2. Or (even better), we could rewrite the polynomial in powers of $x + 1 = x - (-1)$, by substituting $x = (x+1) - 1$:

$$2\,((x+1)-1)^3 - 3\,((x+1)-1)^2 + ((x+1)-1) - 1 =$$

$$2\left((x+1)^3 - 3\,(x+1)^2 + 3\,(x+1) - 1\right) - 3\left((x+1)^2 - 2\,(x+1) - 1\right) + (x+1) =$$

$$2\,(x+1)^3 - 9\,(x+1)^2 + 13\,(x+1) + 2$$

and, expanding the powers, the term linear in $x + 1$ has, indeed, 13 as coefficient.

This points to an alternate definition of derivative, that has long range implications:

1. First derivative of f at a: a number α, such that
$$f(x) - f(a) = \alpha\,(x-a) + \varepsilon \tag{4}$$
where ε is small, in the sense we have mentioned: it gets smaller as $x \to a$ faster than $x - a$ – for example this works if $\varepsilon \approx k\,(x-a)^2$, as is the case for polynomials.

2. Second derivative at a is (assuming we found α above) a number β such that
$$f(x) - f(a) = \alpha\,(x-a) + \frac{\beta}{2}\,(x-a)^2 + \delta \tag{5}$$
where δ is small in a similar sense: it gets smaller faster than $(x-a)^2$. The $\frac{1}{2}$ factor has to do with reconciling with the "usual" definition of second derivative.

3. You can imagine how it proceeds from here. The only issue is the extra factor which, for the nth derivative turns out to be $\frac{1}{n!}$

Remark: We can see that polynomials have derivatives of all orders (by the way, if the degree is n, all derivatives from the $n+1$-th on are zero). On the other hand, "most" functions don't have any derivative at all, in that their graphs do not have tangents. However, all "elementary" functions (those that we know by name) also have derivatives of all orders, except where they are not continuous, or at otherwise exceptional points (e.g., $|x|$ has no derivative at $x = 0$).

7 Review of Important Theorems

7.1 Product and Quotient Rule

It's easy to prove both:

$$\frac{f(x+h)g(x+h) - f(x)g(x)}{h} = \frac{f(x+h)g(x+h) - f(x+h)g(x)}{h} +$$
$$+ \frac{f(x+h)g(x) - f(x)g(x)}{h} \to f'(x)\,g(x) + f(x)\,g'(x)$$

We first check the rule for the derivative of the reciprocal,

$$\frac{\frac{1}{f(x+h)} - \frac{1}{f(x)}}{h} = \frac{f(x) - f(x+h)}{hf(x+h)f(x)} \to -\frac{f'(x)}{f^2(x)}$$

Hence,

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)}{g(x)} - f(x)\frac{g'(x)}{g^2(x)} = \frac{f'(x)\,g(x) - f(x)\,g'(x)}{g^2(x)}$$

7.2 Chain Rule

Also easy: let's use the alternate definition of derivative:

$$f\left(g\left(x+h\right)\right)-f\left(g\left(x\right)\right)=\frac{d}{dx}f\left(g\left(x\right)\right)h+\varepsilon$$

but

$$f\left(g\left(x+h\right)\right)=f\left(g\left(x\right)+g'\left(x\right)h+\varepsilon'\right)=f\left(g\left(x\right)\right)+f'\left(g\left(x\right)\right)\left(g'\left(x\right)h+\varepsilon'\right)+\varepsilon"$$

Neglecting all the ε's, we have that

$$\frac{d}{dx}f\left(g\left(x\right)\right)h=f'\left(g\left(x\right)\right)g'\left(x\right)h$$

that is, the chain rule.

Note also how, if $f\left(x\right)=ax+b, g(x)=cx+d$, $f\left(g\left(x\right)\right)=ag(x)+b=acx+ad+b$, so the linear part is indeed the product. Since differentials are the linear approximation to functions, the chain rule says that the differential of the composition of two functions is the composition of the differentials! This simple observation will expand to provide better insight when you move on to more advanced calculus, as in managing several functions of several variables at once.

8 Derivatives of Transcendental Functions

8.1 Exponentials

Let $f(x)=e^x$. We have that

$$\frac{e^{x+h}-e^x}{h}=e^x\frac{e^h-1}{h}=e^xf'(0)$$

Now, the book defines e as that number such that $f'(0)=1$. On the other hand, if we think of $e=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$, we can produce an *intuitive* reason why De^x is 1 for $x=0$: think of the linear interpolation[2] at $x=0$ for very large n, as the initial slope is indeed 1:

$$\left(1+\frac{x}{n}\right)^n=1+n\cdot\frac{x}{n}+\frac{n(n-1)}{2}\cdot\frac{x^2}{n^2}+\ldots=1+x+\frac{n^2-n}{n^2}\cdot\frac{x^2}{2}+\ldots\approx 1+x+\frac{x^2}{2}+\ldots$$

as $n\to\infty$. Incidentally, referring to our previous discussion of second derivatives (formula (5)), the second derivative of e^x at $x=0$ is equal to 1 as well (and so are all the others).

Then, since for any $a>0$, $a=e^{\ln a}$, $a^x=e^{x\ln a}$, and

$$\frac{da^x}{dx}=a^x\ln a \tag{6}$$

You can see how much more convenient base e is, compared to any other base.

[2] This comment refers to the intuitive discussion of how the number e appears as a limiting case for things like interest compounded continuously, or population growth.

9 Derivation of Inverse Functions (Application to Logarithms)

From the chain rule and $f\left(f^{-1}(x)\right) = x$, we have

$$\frac{d}{dx} f\left(f^{-1}(x)\right) = f'\left(f^{-1}(x)\right) \frac{df^{-1}}{dx} = 1$$

$$\frac{df^{-1}(x)}{dx} = \frac{1}{f'\left(f^{-1}(x)\right)} \tag{7}$$

Now, let $f(x) = e^x$, $f^{-1}(x) = \ln x$. Hence,

$$\frac{d \ln x}{dx} = \frac{1}{e^{\ln x}} = \frac{1}{x}$$

Since, for any $a > 0, a \neq 1$, $\log_a x = \frac{\ln x}{\ln a}$, we have

$$\frac{d \log_a x}{dx} = \frac{1}{x \ln a} \tag{8}$$

Equations (6) and (8) show why it makes little sense to consider any base different from e whenever we are doing calculus – which is what we do in almost any scientific application involving exponential or logarithmic functions.

Remark: From (7) we see that for the inverse function of f to be differentiable, we need $f'\left(f^{-1}(x)\right) \neq 0$. We'll soon see that this condition is not only necessary, but also sufficient, in the sense that if a function has continuous derivative and $f'(a) \neq 0$, then we can define an inverse function (that turns out to be also differentiable) f^{-1} with a domain containing at least a small interval around a.

10 Trigonometric Functions

Let's recall:

$$\frac{\sin(x+h) - \sin x}{h} = \frac{\sin x \cos h + \cos x \sin h - \sin x}{h} =$$

$$\frac{\sin x (\cos h - 1)}{h} + \cos x \frac{\sin h}{h}$$

Now, we know that

$$\lim_{x \to 0} \frac{\cos h - 1}{h} = 0 \tag{9}$$

(see the remark at the end of this section), while

$$\lim_{x \to 0} \frac{\sin x}{x} = 1$$

Hence,

$$D \sin x = \cos x$$

Similarly,

$$\frac{\cos(x+h) - \cos x}{h} = \frac{\cos x \cos h - \sin x \sin h - \cos x}{h} =$$

$$= \cos x \frac{\cos h - 1}{h} - \sin x \frac{\sin h}{h} \to -\sin x$$

just as before. To find the derivative of the tangent, we may use the quotient rule:

$$D \frac{\sin x}{\cos x} = \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}$$

Remark: Here is a proof of (9):

$$\frac{\cos^2 x - 1}{x^2} = \frac{\cos^2 x - \cos^2 x - \sin^2 x}{x^2} = -\frac{\sin^2 x}{x^2} \to -1$$

as $x \to 0$. Hence,

$$\frac{(\cos x - 1)(\cos x + 1)}{x^2} = (\cos x + 1) \frac{\cos x - 1}{x^2} \to -1$$

and since $\cos x + 1 \to 2$

$$\frac{\cos x - 1}{x^2} \to -\frac{1}{2}$$

Besides showing that $\cos x \approx 1 - \frac{x^2}{2}$ for small x (a very interesting fact in itself), this also proves that

$$\frac{\cos x - 1}{x} = x \cdot \frac{\cos x - 1}{x^2} \to 0 \cdot \left(-\frac{1}{2}\right) = 0$$

11 Implicit Derivatives

There is a huge hidden fact behind this result. Superficially, this is simply an application of the chain rule, with lots of assumptions. The simple case is

$$f(x) + g(y) = c$$

so that (assuming g is invertible)

$$y = g^{-1}(c - f(x)) \tag{10}$$

Now, combining the chain rule and the inverse function rule, we can compute $\frac{dy}{dx}$. However, there is a faster way. Since we know from (10) that we may consider y as a function of x, we may write the identity

$$f(x) + g(y(x)) = c$$

for all x, and deriving this, we get (chain rule)

$$f'(x) + g'(y(x))\, y'(x) = 0$$

$$y'(x) = -\frac{f'(x)}{g'(y(x))}$$

This is a special case of a much more interesting result. We have to agree to think of "partial derivatives". That's when $f(x,y)$ is derived with respect to x, respectively y, *by keeping the other variable constant.* Notations are

- f'_x, f'_y
- $D_x f, D_y f$
- $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$

For technical reasons, we need to assume that these derivatives are continuous functions. In this case, it goes like this:

> Given a function $f(x,y)$, and a (non empty) set defined by $f(x,y) = c$, where c is a constant, *suppose that there is a function $y = g(x)$, such that $f(x, g(x)) = c$, then, via the chain rule,*

$$\frac{\partial f(x, g(x))}{\partial x} + \frac{\partial f(x, g(x))}{\partial y} \cdot \frac{dg(x)}{dx} = 0 \tag{11}$$

$$\frac{dg(x)}{dx} = -\frac{\frac{\partial f(x,g(x))}{\partial x}}{\frac{\partial f(x,g(x))}{\partial y}} \tag{12}$$

Here's a quick idea for proof of (11):

$$\frac{f(x+h, g(x+h)) - f(x, g(x))}{h} = \frac{f - f(x, g(x+h)) + f(x, g(x+h)) - f(x, g(x))}{h}$$

and the rest proceeds just as for the proof of the chain rule.

The beauty of this result is that if (12) makes sense, then the reverse is true! That is, if $\frac{\partial f}{\partial y}$ is continuous near a point $(\bar{x}, \bar{y})$, such that $f(\bar{x}, \bar{y}) = c$, and $\frac{\partial f}{\partial y} \neq 0$ is continuous at $(\bar{x}, \bar{y})$ (in a sense that needs to made precise), then there is a function defined at least for x close to $\bar{x}$, such that

$$f(x, g(x)) = c$$

for all x for which g is defined (this is known as the *Implicit Function Theorem*, a basic result in multivariate calculus). Though it may not be obvious, this theorem (for functions of more than one variable) is strictly germane to the the result we mentioned in section 9, which is the one-dimensional version of the so-called *Inverse Function Theorem.*