Variables, Functions, Equations and Graphs

Q1: What does "calculus" mean?

The word "calculus" has the same root as "calcium" – this is because the Romans used little pebbles, generally of limestone, arrayed on boards to do their reckoning. During the middle ages, a "calculus" came to mean any method of reckoning or problem solving.

The subject of this course was once called "the calculus of infinitesimals", and was one calculus among many – such as the "calculus of probabilities". It has proven to be so useful that nowadays it is simply called "the calculus", although one does occasionally run into references to other calculi.

Calculus is part of a branch of mathematics known as analysis. And the things one analyses in analysis are functions, and the solutions to equations.

Q2: What are functions?

A function is simply a rule that assigns a member of an "output" set, usually called the **range**, to each member of a given set of possible "inputs", usually called the **domain**.

In this class, the domain and range will both generally be subsets of the real numbers, but later in the course we will consider the complex numbers as well. There is no restriction; the input and output sets can be any sets at all.

What is important though is that **exactly one output is assigned to each input**. Any relation between inputs and outputs that doesn't satisfy this requirement is just not a function.

Q3: O.K., a function is a rule assigning outputs to inputs – but how do we specify the rule?

There are many ways to specify functions. If the set of possible inputs – the domain – is a small enough finite set, one could just **list** the outputs associated to a given input.

For example, here is a function f with domain $\{0, 1, 2, 3\}$ and range $\{0, 2, 4, 6\}$ given explicitly by

$$f(0) = 0 \qquad f(1) = 2 \qquad f(2) = 4 \qquad f(3) = 6$$

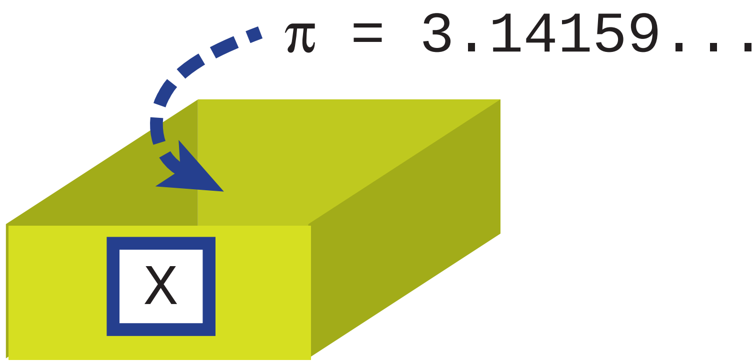
The input is shown inside the parentheses, and the output is specified to the right of the equal sign.

Q4: What if my domain is infinite? If there are infinitely many inputs, I would need an infinitely long list.

If there are infinitely many inputs, one definitely can't use a list. Instead, one has to specify **what it is that the function "does" to its inputs.**

You can see what the function that we specified by a list just above is doing – it just doubles the input. That makes sense for any real number as input, and by specifying this "action" of the function we can avoid an infinite list! To do this, we turn to **variables** and **operations**. A real variable is simply a named container for a real number. Think of it as a box, with a label on front, and a real number inside. We specify variables by name – the label on the box.

This picture represents a variable x, with the current contained value being $\pi = 3.14169...$



It is standard to use letters like x, y and z for the names of variables. There is actually an interesting story behind the use of x as the standard name for a variable.

An **operation** is something you can do the value; i.e, the real number, inside the box. For instance, you can double it, or you can square it. When we write $2x$, this means "take the value in the box labeled x, and double it. When we write x^2, this means "take the value in the box labeled x, and square it. And so on.

This way of thinking about variables and operations is familiar to you if you have done any computer programming, especially in a language like C. In this setting, a variable has a name and a type. The name is associated with an address in memory, and the type tells how big a block of memory, starting from the specified address, is used to hold the value. When you call the variable latter in the program, you get the value stored in that block of memory.

We can now specify a function f by giving its domain – say, all of the real numbers – and the sequence of operations it performs on a given variable x, which could be holding any of the values in the domain. For example,

$$f(x) = 2x \qquad \text{or} \qquad f(x) = x^2 \qquad \text{or} \qquad f(x) = 2(1 + x^2)$$

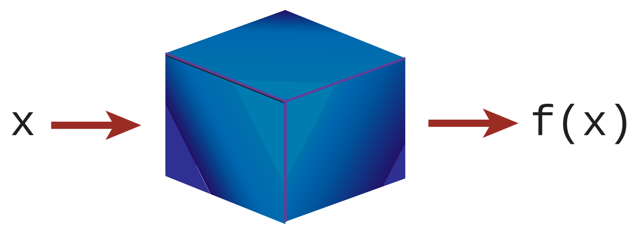
The last example was the only one involving more than a single operation. Here, there were three: first we squared the value in x, then added 1 to it, and then doubled that.

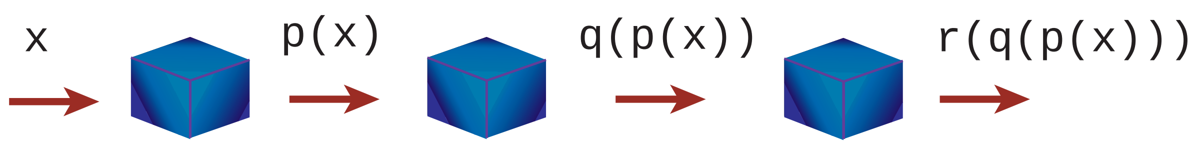
Q4: Can't I just think of the action of f in the last example as one bigger operation?

Yes, you can. But actually it wil turn out to be very useful in general to think of functions as an assembly line:

We can think of this function f as an assembly line: A box – x– comes in at the left containing some value, it is opened and sent to three successive stations, where operations are applied to it. Let the three stations be given by the elementary functions $p(x)$, $q(x)$ and $r(x)$ where

$$p(x) = x^2$$
$$q(x) = 2 * x$$
$$r(x) = 1 + x$$





A basic strategy in the analysis of functions – in all of science for that matter – is to take complicated functions apart into simpler pieces, and then to analyse them in terms of these simpler constituents.

We will do this again and again here. The divide–and-conquer principle is fundamental to mathematics and much else besides warfare.

The assembly line analogy leads us straight into another important notion – the composition of functions. If f and g are two functions, and the domain of g contains the range of f, then we get a new function $g \circ f$, called "g composed with f" by defining

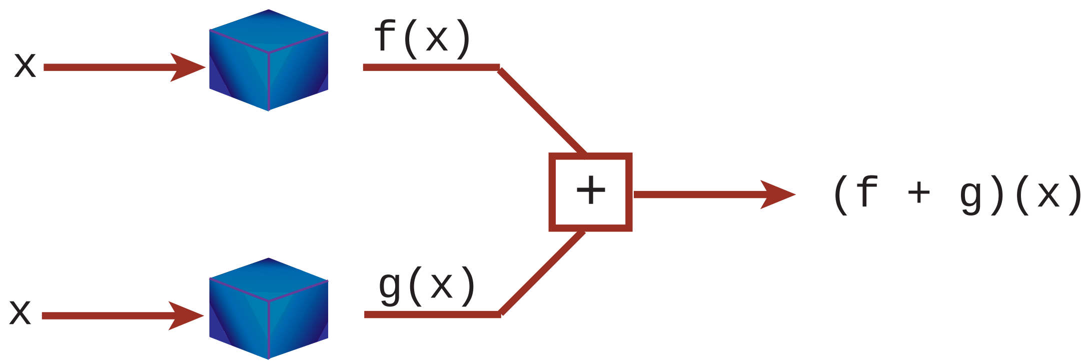
$$g \circ f(x) = g(f(x)) \ .$$

This means "take the value in x, put it through f, and then take what comes out, and put it through g".

This only makes sense if the output values of f are possible input values for g, so the requirement that the domain of g contains the range of f is crucial. However, in many cases the domain of g will contain all real numbers, and there is no problem. Or, if there is a problem, the domain of f can be restricted so it doesn't produce any output values outside the domain of g.

Q5: Is composition the only way to build complicated functions out of simple ones?

No, composition is just one way to combine a pair of functions to produce a new function. There are others, more closely related to the familiar arithmetic operations we can perform on numbers. For instance, we can add up the outputs of two functions. The net result can be viewed as a new, more complicated function.



More generally, if f and g are two functions with the same domain, we define new functions by

$$\begin{aligned}(f+g)(x) &= f(x)+g(x)\\ (f-g)(x) &= f(x)-g(x)\\ fg(x) &= f(x)g(x)\end{aligned}$$

If moreover 0 is not in the range of g, we define

$$\frac{f}{g}(x) = \frac{f(x)}{g(x)}\ .$$

In this way some very complicated functions can be built out of very simple building blocks. For example, with

$$f(x) = x+1 \qquad g(x) = 1/x \qquad h(x) = x^2 \text{and} \qquad j(x) = \sin(x)$$

$$\frac{1}{1+\sin^2(x+1)} + \sin^2(x^2) = g\circ f\circ h\circ j\circ f + h\circ j\circ h$$

Q6: O.K., so now I know what functions are, and I understand that being able to take them apart into simple pieces is supposed to help me analyze them,

but exactly what sort of analysis will I be doing? Suppose I have a function to analyze. Which specific questions will we be asking about it?

There are many kinds of questions, and it is not possible to list them all now. But here are two examples that we can talk about at this time.

(1) Given a function f, is there an input x_0 so that

$$f(x_0) \geq f(x)$$

for all other x in the domain of f? If so, what is $f(x_0)$, the larges output, and what are all of the input values that produce it?

(2) Given a function f and a value a, are there any input values x so that $f(x) = a$? That is, are there any solutions of the equation $f(x) = a$? If so, what are they?

The first of these is an **optimization problem**, as would be the corresponding question about smallest values. If the function has a finite domain, and is given in the form of an explicit list, as in our first example, then the problem is solved simply by running down the list. But if the domain is infinite, we cannot use a list. We must instead analyse the operations, or assembly line steps, out of which the function is built. Calculus provides methods for this.

Likewise with the second problem. If f is given by a list, just look down the list of output values and see if you see a. In the special case $a = 0$, the solutions are called **roots** of f. We can always reduce to this case by defining a new function $g(x) = f(x) - a$. Then solutions of $f(x) = a$ are roots of g. Calculus provides powerful methods for finding roots.

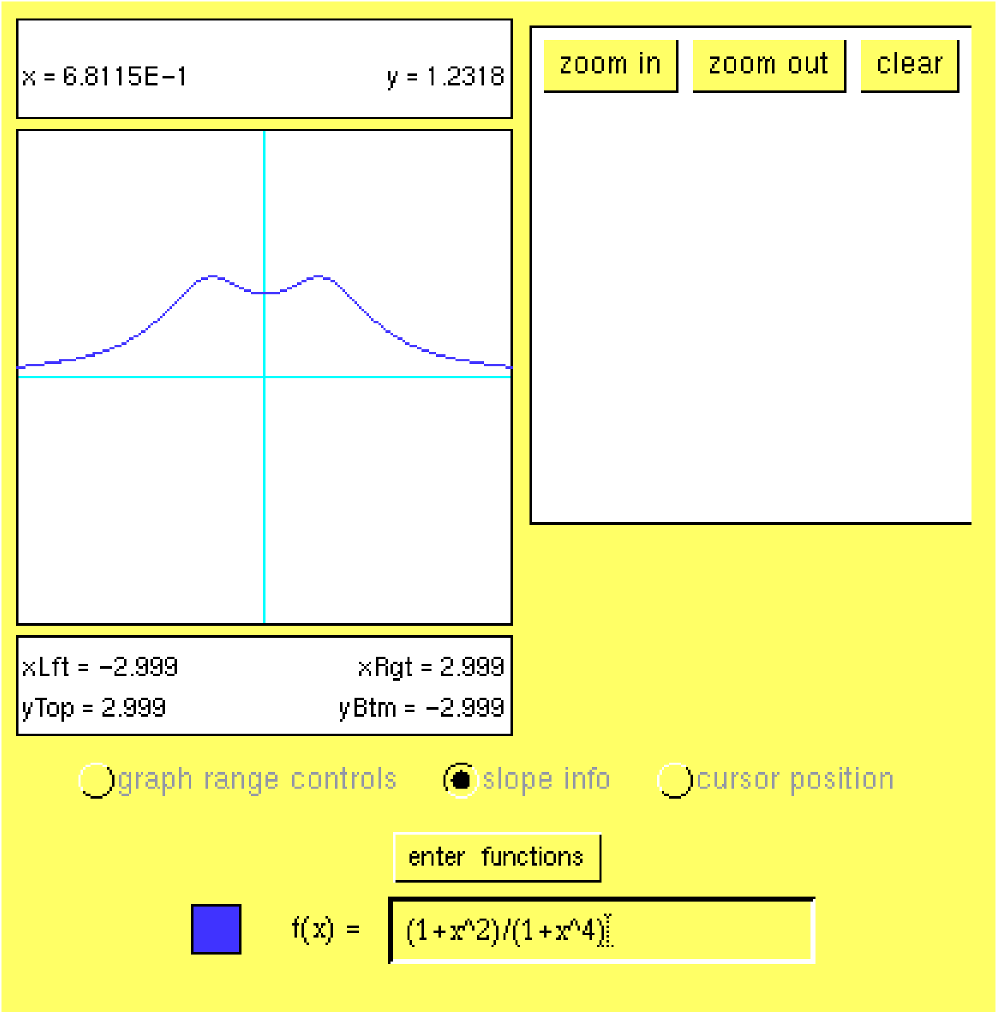
Q7: What kinds of methods will we use – besides taking functions apart into simple pieces?

Analysis is part of "figuring things out", and figuring things out is quite literally part of analysis. **Drawing figures and using geometrical insight** will basic to our strategy of analysis.

We can bring in a geometric perspective by turning to the **graphs** of our functions. The range, or some piece of it, is conventionally drawn on the vertical axis, and the domain, or some piece of it, on the horizontal. For each point in the domain of f, draw the vertical line though that input value on the horizontal axis. Then draw the horizontal line through the corresponding output value on the vertical axis. These two lines meet at a single point, which is a point on the graph of f. The graph of f consists of all the points that are obtained in this way.

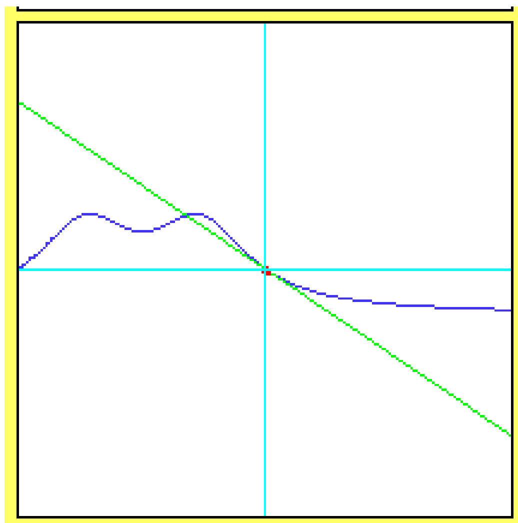
As a subset of the plane, a graph can be drawn on a sheet of paper, or a computer screen. One way of drawing a graph is to compute the points on it for a large but finite collection of input values, and then to "connect the dots". This is tedious to do by hand, but easy on a computer.

The Course Notes web page contains a link to a java applet that you can use to draw graph and "zoom in" on details of the graph. Here is what you see if you ask it to graph $f(x) = (1 + x^2)/(1 + x^4)$

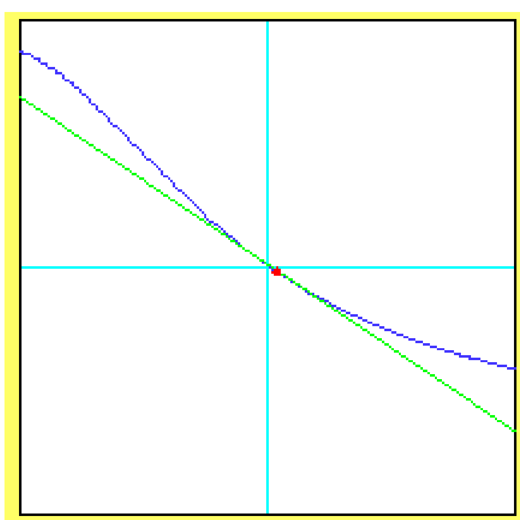


The applet is easy to use; instructions are on the applet page. If you "click" on a point in the graph, you see a line drawn in that "fits" the curve "as well as it can" at that point. This is the tangent line, and we'll be doing a lot of work with tangent lines later in the course.

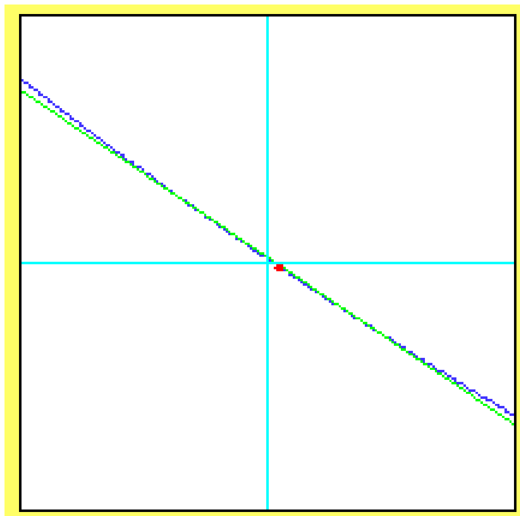
The thing to notice now though is that the graph becomes much simpler if you "zoom in" on small pieces of the graph. Drag themouse to move the red dot to the center (on the applet page itself; this is just a picture!) and click the zoom button. The picture changes to



Now press the "zoom in" button a few times, and you see which already looks a lot simpler!



After a few more times you see:



You begin to see a line! Go to the applet page and try this out for several functions and points.

This is the absolutely central point in this whole course: **Up close, the graphs of most reasonable functions look like lines.**

The escape clause about "most reasonable functions" is necessary. Try for example

$$f(x) = |x|$$

which you would type in the applet as

$$\text{abs}[x]$$

at $(0, 0)$.

As you zoom in around this point, the graph doesn't even change; it keeps its kink forever. But at any other point, it does become linear if we zoom in.

Q7: It's nice that the functions look like lines up close, but what's so good about lines?

Lines are very simple – they are described by simple equations. Simplicity is not just good – its great!.

Any line that is not exactly vertical has a finite slope and y-intercept, and is the graph of a linear function

$$f(x) = mx + b\ .$$

Here m is the slope, and b is the y-intercept. These two numbers m and b completely specify the linear function f, and the corresponding line that is its graph.

Another way to specify a line is to give a point (x_0, y_0) on the line and the slope m. The function f is given by

$$f(x) = m(x - x_0) + y_0 \ .$$

Notice that $f(x_0) = y_0$, as must be the case. Also

$$m(x - x_0) + y_0 = mx + (y_0 - mx_0)$$

so that the y intercept b is $y_0 - mx_0$.

Now, consider any point (x_0, y_0) on the graph of f – i.e., any point (x_0, y_0) with $y_0 = f(x_0)$. Let m be the "zoomed in slope" – i.e., the slope of the line we see when we zoom in on (x_0, y_0). Then the graph of

$$f(x) = m(x - x_0) + y_0$$

is a line which passes through (x_0, y_0) and "fits" the graph of f itself there as much as possible. This line is called the tangent line to the graph of f at (x_0, y_0). The problem of finding the formula for this line – which amounts to the problem of finding the "zoomed in slope" m since x_0 and y_0 are known – is called, naturally enough, **"the tangent line problem"**. It is the central problem of this course.

Q8: What's so important about this tangent line problem?

It is not at all obvious that this should be such a central problem. The greeks missed this point completely. So let's recapitulate. We've seen that up close most functions – at most points, anyhow – look like linear functions, and linear functions are very simple. This simplicity "in the small" is the guiding light in the infinitessimal calculus. To see how one could use this simplicity in the small to analyse a problem, lets look at an example. Let's try to answer question (1) for the function

$$f(x) = \frac{1 + x^2}{1 + x^4}$$

on the domain $-2 \le x \le 2$. The graph appears a bit back in this section of notes.

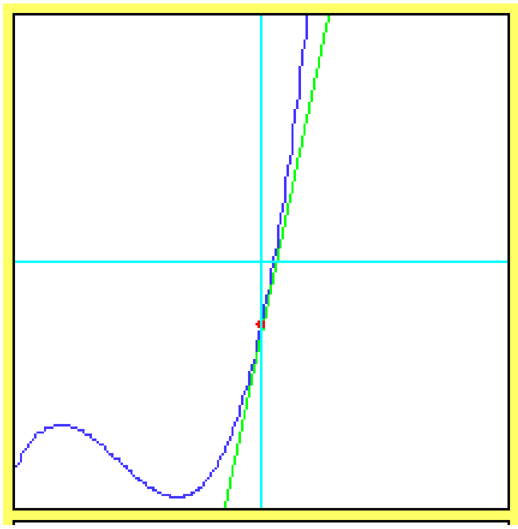
You can see two "humps" near $x = \pm 1$ where f takes on the value 1. Zoom in on $(1, 1)$ in the applet until you see a line. What is its slope?

You presumably found a horizontal line; i.e., zero slope. There are exactly three points in the graph at which the "zoomed in slope" is zero. And as we will see later, it is only at such points that we have to look for maxima and minima. So finding the "zoomed in slope" will be the key to solving optimization problems.

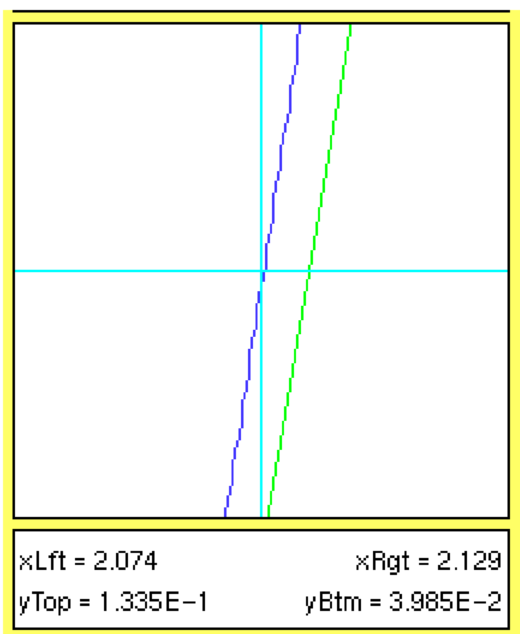
How about the second problem. Let's try to find the roots of

$$f(x) = x^3 - 2x - 5 \ .$$

Here is a graph centered on $(2, 0)$. The red dot is at $(2, -1)$, which is on the graph of f.



There is exactly one root, and it is not too far away from $x = 2$. (If you go to the applet itself, and move the mouse over a point in the graph, the applet tells you the coordinates of that point.) Also on the graph is the tangent line to the graph at $(2, -1)$. You can see that the tangent line and the graph are pretty close to each other near $(2, -1)$. The tangent line itself is the graph of $h(x) = 10x - 21$. (You'll have to take this on faith right now, but soon you'll have the means to easily work out that the "zoomed in slope" at $(2, -1)$ is 10.) Zooming in on the root, we see:



Now, since the graphs of f and h are close together near $x = 2$, the root of h must be close to the nearby root of f. The later is hard to find, since f is cubic, but it is east to find the root of h because it is linear: $h(x) = 0$ means $x = 21/10 = 2.1$ The next picture show the "zoomed in" view neear the root – that is where the blue line crosses. The tangent line is in green, and it crosses a bit to the right. The two crossings are close – as you can see, the left edge of the graph is at $x = 2.074$ and the right edge at $x = 2.129$.

Now, this point where the green line crosses is a root of h, not f. But since the graphs of f and h are close to one another in this region, we expect to get something small when we plug 2.1 into f. Indeed, $f(2.1) = 0.061$.

Now let's zoom in and look at the graph of f near $(2.1, 0.061)$. Here, the tangent line and the graph are really close. The tangent line at $(2.1, 0.061)$ is the graph of

$$h(x) = 11.23(x - 2.1) + 0.061$$

left in point–slope form this time. Again, it is easy to solve $h(x) = 0$; one gets 2.094568121 (keeping 9 decimal places). At this root of h, the value of f is 0.000185732. This procedure can be further iterated. We could compute the tangent line at each new point and go on. This procedure, due to Newton, yields the answer to any desired degree of accuracy in a few steps. In fact, the number of correct decimal places in the answer doubles with each iteration.

You might be dissatisfied with such an approach to finding roots since it doesn't give the exact answer. For quadratic equations we have the quadratic formula. Isn't there something like that here? Actually yes in this example (which is the one Newton used in his original discussion in *Methodus Fluxionum et Serierum infinitarum* written between 1664 and 1671) since it is cubic. There is even a formula for finding the roots of fourth order polynomials. But for fifth degree or higher, there simply isn't any such formula. This was proved by Galois in the early 19th century.

So in general, there isn't anything "better" in the sense of providing a closed form solution. However, things are not so bad either. If one works on a computer, one rarely needs anything more than IEEE double precision accuracy for floating point numbers. This keeps 16 accuarate digits, which can almost always be gotten by 4 or 5 iterations of Newton's method from a reasonable starting approximation (loosely speaking, one with one digit of accuracy).