A few brief notes about introducing key math concepts

Zoya Gavrilov ©

July 24, 2012

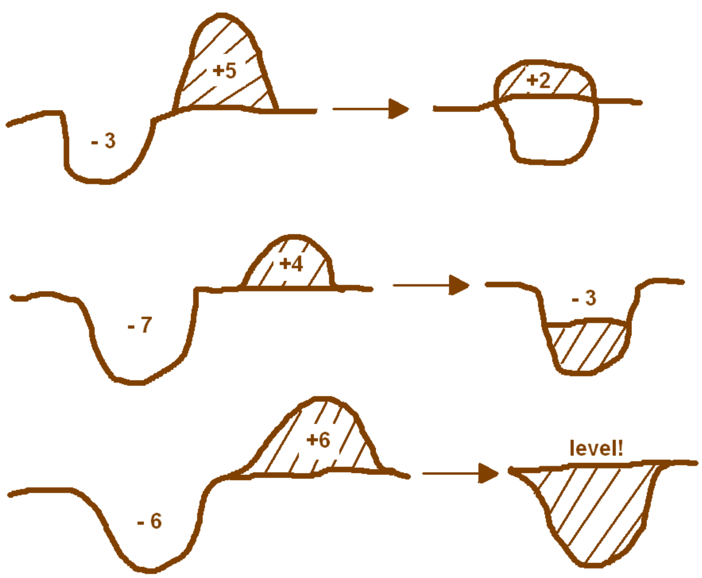
1 Preface

This is a collection of notes I put together to remind myself of good ways to introduce given concepts, and the types of analogies that help clarify ideas. This is a very rough guide that I will continue to update and add to. It may help serve as a resource to other teachers/tutors struggling to find the best ways to get across to their students. The target age group is mostly high school (with some secondary-school material), though unfortunately, many students can make it to university without knowing (not to mention having a strong foundation in) some of these key concepts. For those students this guide can serve as a review or else speedy introduction to these concepts. This document is mostly targetted at the teachers, not the students - it provides some ideas for effective ways of introducing concepts, rather than serving as an explanation of those concepts (the latter requiring more detail, more repetition, and many more examples). One final note: I don't shy away from terminology when explaining concepts to my students - I believe that the earlier they are exposed to them, the sooner they learn to use them correctly – mathematical terminology (usually) serves to formalise, not confuse.

2 Negative numbers

Disclaimer: this is a concept taught in North American schools at the grade 7/8 level... In European countries this understanding comes about in grade 2/3... Why such a discrepancy? The approach to teaching math in most North American schools is very different ('math is mysterious') - the approach is very rule-based and non-intuitive. Here's one example of a concept that should be intuitive, and yet many young students unfortunately struggle with it. With a clear enough explanation, it can very well be introduced at a much earlier age... As can all the rest of the mathematical building blocks, which will follow suit.

A good way to think about negative numbers is as holes or ditches (in the ground), relative to positive numbers, which are like mounds or hills (on the ground). Adding positive and negative numbers is thus nothing more than filling ditches and building mounds. For instance, if we add two numbers, one negative and the other positive, where the negative number is bigger in magnitude, we can imagine starting with a ditch, but not having enough material (from the mound) to fill it with. Thus we still end up with a ditch, but it'll be shallower. What if we start with a mound and subtract a number of the same magnitude? Then what we've done is levelled the mound to the ground (we end up with 0 as the result). What if the number we're subracting is even bigger? Then we do more than level the mound with the ground, we actually end up digging a ditch. Analogously, if we start with a ditch (a negative number) and add to it a positive number of the same magnitude, then we end up filling the whole ditch (thus levelling the ground). If we add a larger positive number, then we fill the ditch and have material left over to build a mound.



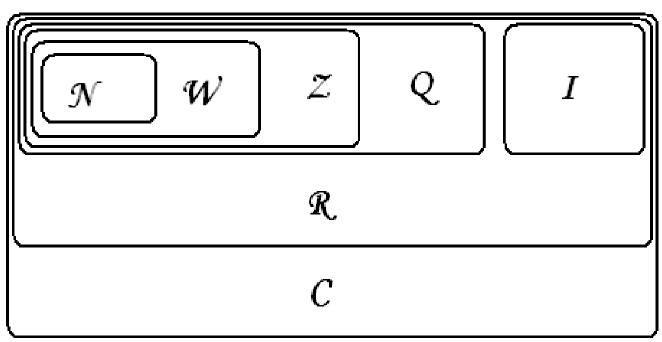
Thus, to figure out which situation we're in, we have to determine which is bigger: the ditch or the mound? Given two numbers, a positive one and a negative one, we compare their magnitudes (the sizes of the numbers themselves, disregarding the signs). If the negative one is greater in magnitude, then our ditch is deeper than the mound we have to fill it with, so we expect to end up with a ditch. If the positive number is greater in magnitude, then our mound is large enough to fill the ditch with some material left over to build a mound. How deep is the resulting ditch or else how high is the resulting mound? Just the difference between the magnitudes of the two numbers. Forget about the signs for now, and just figure out how much bigger the biggest number is of the pair than the smallest.

3 Sets and set notation

Often we want to write down a collection of numbers together, so we can talk about them as a group and describe some property/properties they share. In mathematical notion, we put such a collection of numbers between braces, and we call this a **set**. If we can list out all the numbers in a set, or else make it clear what is in the set via the listing out of a few elements that follow a clear pattern, we do so, like this: $\{5, 37, 99\}$ (the set of size 3 containing only the numbers 5, 37 and 99), $\{1, 2, 3, ...\}$ (the set containing all positive integers), $\{2, 4, 6, ...\}$ (the set containing all even positive integers starting at 2). Also see: 7.1.

3.1 Number sets

A few main number sets keep recurring, so we have special symbols set aside for them. The nesting of these number sets closely follows our growing/evolving understanding of numbers. The natural numbers are the numbers with which we learn to count: $\mathbb{N} = \{1, 2, 3, ...\}$. Then we learn the concept of the number 0, where the set which contains the natural numbers along with 0 is the set of whole numbers $\mathbb{W} = \{0, 1, 2, 3, ...\}$. Next come the integers, which include the natural numbers, their negative counterparts, and zero: $\mathbb{Z} = \{..., -3, -2, -1, 0, 1, 2, 3, ...\}$. The set of rational numbers are all numbers that can be represented as a fraction, which includes the integers, since any integer is a fraction with 1 in the denominator. Examples of numbers in this set are: $\frac{1}{2}$, $\frac{3}{5}$, 0.4 (which can be represented as the fraction $\frac{4}{10}$), $0.\overline{3}$ (which can be represented as $\frac{1}{3}$). Note, any finite decimal expansion, or else any repeating decimal expansion, are part of this set. Not part of this set are all decimal numbers that can not be written as a fraction, for instance π, e (with infinite decimal expansions that don't repeat), as well as radicals that don't simplify like $\sqrt{2}$. These numbers are found in the set $\mathbb{I}$ of irrational numbers. Together, rationals and irrationals form the set of reals $\mathbb{R}$, while the reals are in turn a subset of complex numbers $\mathbb{C}$ (consisting of real and imaginary parts). A venn diagram explaining how the sets are related among themselves is pictured below:



3.1.1 A note on repeating decimal expansions

When we want to write the number $0.\overline{3}$, what we really mean is that we have an infinite expansion of 3s after the decimal: 0.333.... Similarly, $0.\overline{45}$ is equivalent to 0.454545..., but $0.4\overline{5}$ is equivalent to 0.45555.... In other words, the bar over the numbers after the decimal indicates which part repeats. The neat thing is that any repeating decimal expansion can be expressed as a fraction.

How? Here's the recipe:

- separate out the repeating (a) and non-repeating (b) parts (e.g. $0.4\overline{5} = 0.4 + 0.0\overline{5}$)
- construct the denominator for (a): write down a 9 for each repeating digit, and follow this with as many zeros as there are zeros after the decimal in (a) (e.g. for $0.0\overline{5}$, only one digit is repeating, which makes for one 9, and only one zero occurs before 5, which translates to one 0; thus, the denominator is 90).
- combine the fractions that represent (a) and (b) (e.g. $0.4 = \frac{4}{10}$, $0.0\overline{5} = \frac{5}{90}$, so the result is $\frac{4}{10} + \frac{5}{90} = \frac{41}{90}$)

Thus, all numbers with repeating decimal expansions are in the set of rational numbers.

3.2 Notation

Often, it is possible to more succinctly and precisely write down a set of numbers, with the aid of the main number sets introduced in the previous section. For instance, we can write down the set of numbers $\{2, 4, 6, ...\}$ as $\{2k|k \in \mathbb{N}\}$. What does this notation mean? It reads: "the set of numbers of the form $2k$ where k is a natural number". This symbol: "$|$" can be read as "where" or "such that". This symbol: "$\in$" can be read as "element of (the set)" or "is a member of (the set)". Hence, this mathematical notation reads just like a natural English sentence but is much more compact. A simple change will help us similarly rewrite the set $\{0, 2, 4, 6, ...\}$ as $\{2k|k \in \mathbb{W}\}$. What about $\{1, 3, 5, ...\}$? Easy! $\{2k + 1|k \in \mathbb{W}\}$. Note that when $k = 0$, $2k + 1 = 1$; when $k = 1$, $2k + 1 = 3$, etc. Hence, to make sure we are including the numbers we want to be including, we can thus test out (by plugging in various values for k) which numbers will be part of the set. We can also describe the set of all positive real numbers as $\{x|x > 0\}$. However, because this set is often used, it has a more compact form, written most often as either: $\mathbb{R}^+$ or as $\mathbb{R}_{>0}$.

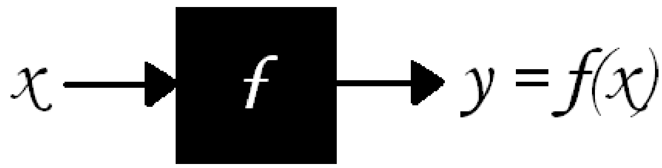
4 Functions

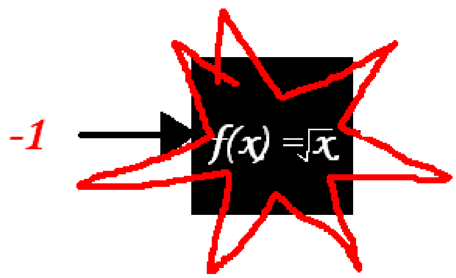
4.1 Domain and Range

Consider a mathematical function: it is a black box that accepts input, operates on it, and returns a result. Here is a picture:

What is y? It is the result of calling the function with the input argument x. In other words, it is what the box spits out at us after we put x in.

Now consider all the things that we can put into the box, and all the things that we can get back out from it. It turns out that there are certain inputs, which when fed into the box, cause it to blow up. For instance, consider the black box representating the square root function, and how it behaves when given a negative number (we are operating in the real domain).





It blows up! It can't handle negative numbers (the square root of a negative number does not exist as far as real numbers are concerned). If we wanted to similarly "blow-up" the box that corresponds to the function $f(x) = \frac{1}{x}$, we could feed it $x = 0$.
In other words, for each function box, there are many inputs on which our box will give us a valid output, and there are other inputs on which our box will blow up. We say that all inputs which the box can handle are in the **domain** of f, and the ones that make the box blow up are *not* in the domain of f. Additionally the **range** of a function consists of all values y which the box can spit out at us.

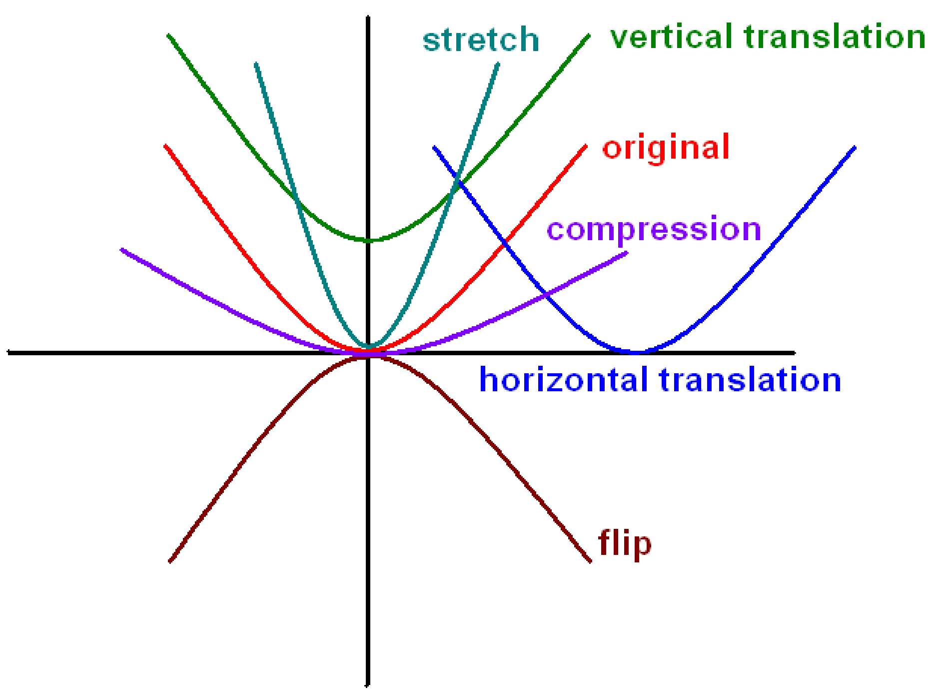
Consider again $f(x) = \sqrt{x}$. Is $x < 0$ (all negative values of x) in the domain of f? No. Above, we determined that negative values of x will cause our function box to blow up. What about $x = 0$? Sure, because $f(0) = \sqrt{0} = 0$, so our box, when fed 0, spits out 0. What about $x > 0$? Yes, since the square root function is defined for all positive values of x. So, the domain of f is all $x > 0$. This can be written more precisely as: $dom(f) = \{x|x \geq 0\}$ (see the section on sets and set notation). Now what about the range of this function? Can the result of a square root operation ever give us a negative answer? No. Can we get 0? Yes, because plugging in 0 will give us 0. Hence, the range of our function f also includes all positive real numbers, written: $range(f) = \{x|x \geq 0\}$.

4.2 Functions as Sets

Now that you've read about sets, it makes sense to think about a function as a set of tuples - i.e. a set of points that satisfy the function. For instance, the function $f(x) = x^2$ can be represented as the set: $(x, x^2)|x \in \mathbb{R}$ - note that we've also thus included information about the domain of the function. Sometimes one representation of a function is easier to work with than the other, but it is important to understand that they convey the same concept.

5 Graph transformations

A few quick points about graph transformations and how to think about them: as an example, consider a quadratic: $y = x^2$. What happens when we add a positive term to the right hand side of the equation $y = x^2 + q$ $(q > 0)$? The y-coordinate of every point will translate up by q units. This is a *vertical translation* of the graph. Accordingly, if $q < 0$, every point will move down by q units. What if we add a term under the function $y = (x - p)^2$? This transformation affects the x-coordinate, and is termed a *horizontal translation*. How does the graph translate? Not to the left (as your intuition might first tell you), but to the right by p units! Why? Consider where the original graph $y = x^2$ had its vertex: at $(0, 0)$. What about the new graph $y = (x - p)^2$? Note that when $x = p$, $y = 0$, so the vertex is $(p, 0)$. The vertex moved to the right by p units, and similarly will the other points. Now what happens when we add a multiplicative factor $y = ax^2$? If $a > 1$, then every point will move vertically by the factor a (not a constant amount, as in a vertical translation, but a multiplicative amount). In other words, points with large y-coordinates, will have much larger y-coordinates, so the whole graph will be sharper (steeper). This is called a *vertical stretch*. What about for $0 < a < 1$? We decrease the y-coordinates of points but keep their signs, so all points come closer to the x-axis. This is a *vertical compression*. And if $a < 0$? Well, we first consider $-a$ and follow the transformations above, and then we just flip the graph over the x-axis.



6 Quadratic Functions

Disclaimer: from my experience tutoring, this is the most badly taught unit in high-school mathematics (and correspondingly, the most poorly understood). This is highly unfortunate since the concepts that come up in this unit are a stepping stone to understanding higher-level material like optimization. Without a solid foundation at this point, students are likely to be struggling later, as they'll have nothing to build on, and will have trouble connecting ideas. To avoid offering up yet another textbook-like unit on quadratic functions, below I describe the best order (from my experience) in which to introduce concepts to students. The following needs to be highly supplemented with examples.

The quadratic function is a powerful function that can be used to model the relationship between variables in many real-world problems. In the 2-dimensional case, students are familiar with problems of the following type: an object is thrown, and traces out a quadratic trajectory. Students are then asked a variety of questions including how high the object was off the ground when it was thrown, what the highest point was, when the object will hit the ground, etc. Height is plotted as a function of time. This relationship between a physical problem and its mathematical counterpart makes sense to students because they've seen the arc traced out by objects as they fly through the air before landing. What is less intuitive is how the quadratic function can be used to model more abstract relationships: like profit as a function of commodity price, or area of enclosure as a function of side length (of the enclosing shape).

Here's a simple example that can be used to lead in to more: consider a piece of wire of length P. What are the dimensions of the rectangle with maximal area that can be constructed out of this length of wire? Call one side x; thus the other side is of length $\frac{P-2x}{2}$. What is the area of this rectangle? $\frac{1}{2}x(P-2x)$. Consider what happens when we make x as big as possible... the biggest we can make x is $\frac{P}{2}$ (since we have two sides of the same length). What happens to the other two sides? They shrink to 0. Hence the area of our rectangle is 0 - no good. Alright, what about if we make side x as small as possible, effectively shrinking it to 0? This case is symmetric to the last, and we once again end up with an area of 0. What have we accomplished? We have discovered when $\frac{1}{2}x(P-2x) = 0$, which is equivalent to finding the x-intercepts of the function $A(x) = \frac{1}{2}x(P-2x)$ where x is side length, and $A(x)$ is the area of the resulting rectangle (under fixed P). In fact the form that the quadratic has been expressed in is *intercept form* (which makes it easy to read off the intercepts). We can express this quadratic in *standard form* (by expanding): $A(x) = -x^2 + \frac{P}{2}x$. This form makes explicit a few things: first, we can see that this is in fact quadratic (the highest-order term has power 2); second, using our knowledge of graph transformations, we note that a negative in front of x^2 implies that the graph is upside-down (an arc shape). Now that we've established that our graph is a quadratic, we know that the vertex occurs halfway between the two intercepts. This means that the vertex occurs at $x = \frac{P}{4}$. But this is exactly when our rectangle is a square! (4 sides, each of length $\frac{P}{4}$). We have thus shown that a square gives us maximal area. Another way we can derive the same result is my explicitely computing *vertex form* from standard form:

$$\begin{aligned} A(x) &= -x^2 + \frac{P}{2}x && (1)\\ &= -(x^2 - \frac{P}{2}x) && (2)\\ &= -(x^2 - \frac{P}{2}x + (\frac{P}{4})^2 - (\frac{P}{4})^2) && (3)\\ &= -(x - \frac{P}{4})^2 + \frac{P^2}{16} && (4) \end{aligned}$$

Consider what information this gives us. We're looking for the maximum of $-(x - \frac{P}{4})^2 + \frac{P^2}{16}$. Note that $-(x - \frac{P}{4})^2 \leq 0$, so we would maximize the whole expression by setting $x - \frac{P}{4} = 0$ - in other words, by setting $x = \frac{P}{4}$ (which is the same answer we got before). Now what is the value of $A(x)$ at the maximum? Well if $x - \frac{P}{4} = 0$, it follows that $-(x - \frac{P}{4})^2 + \frac{P^2}{16} = \frac{P^2}{16}$. Thus, vertex form (ideally named) makes it easy for us to determine the vertex. More generally, if $f(x) = -a(x - p)^2 + q$, then f attains its maximum of q precisely when $x = p$. In other words, the vertex of this quadratic is (p, q). Correspondingly, the function of the form $f(x) = a(x - p)^2 + q$ has a minimum of q when $x = p$ (same logic of figuring out that $a(x - p)^2 + q$ is minimized when $a(x - p)^2$ is minimized which in turn happens when $x - p = 0$).

Here, we can step back and talk about some of the things we have learned through the simple example above: we have considered 3 forms in which we can express a quadratic: standard form, intercept form, and vertex form. Each form makes something about the graph of the quadratic more explicit, but all 3 forms carry the exact same information and are perfectly equivalent. This also implies that for each max-min problem, there is a multitude of approaches, depending on which form you choose and how you go about manipulating it. This is an important point: realizing that there is a multitude of ways to get at the same solution (and in fact, working out the solution using a multitude of methods) will only improve the understanding of how everything pieces together.

At this point, what we can do is start with the most general quadratic equation in standard form, convert it to vertex form (and thus, in the process, learn how to convert an arbitrary quadratic equation in standard form into vertex form), and by carrying out a few further manipulations, end up with the quadratic formula. Then we may go full circle by showing how obtaining the roots using the quadratic formula can help us write down intercept form.

Staring with standard form:

$$f(x) = ax^2 + bx + c \tag{5}$$
$$f(x) = a(x^2 + \frac{b}{a}x) + c \tag{6}$$
$$f(x) = a(x^2 + \frac{b}{a}x + (\frac{b}{2a})^2 - (\frac{b}{2a})^2) + c \tag{7}$$
$$f(x) = a(x^2 + \frac{b}{a}x + (\frac{b}{2a})^2) - a(\frac{b}{2a})^2 + c \tag{8}$$
$$f(x) = a(x + \frac{b}{2a})^2 - (\frac{b^2}{4a}) + c \tag{9}$$
$$\tag{10}$$

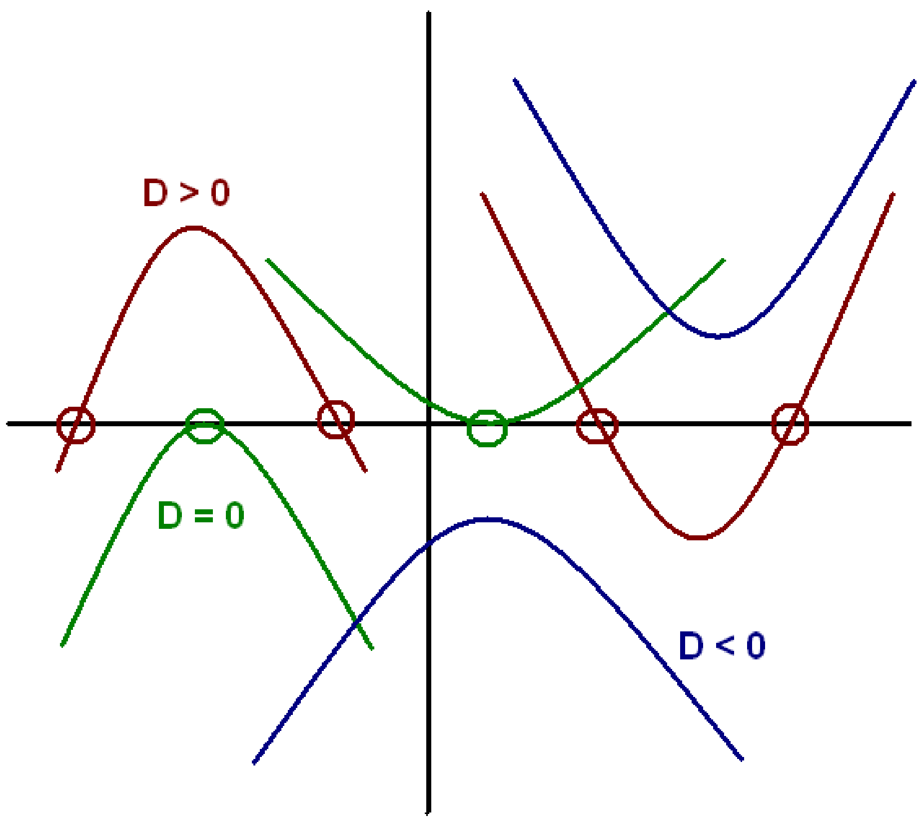
Note that at this point we have obtained vertex form, with vertex: $(-\frac{b}{2a}, -(\frac{b^2}{4a}) + c)$

Now, to solve for intercepts we set $f(x) = 0$

$$a(x + \frac{b}{2a})^2 - (\frac{b^2}{4a}) + c = 0 \tag{11}$$
$$a(x + \frac{b}{2a})^2 = (\frac{b^2}{4a}) - c \tag{12}$$
$$(x + \frac{b}{2a})^2 = \frac{(\frac{b^2}{4a}) - c}{a} \tag{13}$$
$$(x + \frac{b}{2a})^2 = \frac{b^2 - 4ac}{4a^2} \tag{14}$$
$$x + \frac{b}{2a} = \pm\sqrt{\frac{b^2 - 4ac}{4a^2}} \tag{15}$$
$$x + \frac{b}{2a} = \frac{\pm\sqrt{b^2 - 4ac}}{2a} \tag{16}$$
$$x = \frac{\pm\sqrt{b^2 - 4ac}}{2a} - \frac{b}{2a} \tag{17}$$
$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{18}$$
$$\tag{19}$$

Thus, we end up with the quadratic formula. What does the quadratic formula give us? The roots/solutions of the equations - the x-intercepts. We can take these x-intercepts and reconstruct intercept form as follows: $f(x) = (x - \frac{-b+\sqrt{b^2-4ac}}{2a})(x - \frac{-b-\sqrt{b^2-4ac}}{2a})$. Note that we can also now go to vertex form, since the x-coordinate of the vertex occurs halfway between the two intercepts, while the y-coordinate of the vertex can be obtained by plugging the x-coordinate of the vertex back into $f(x)$. In such a way, we can convert from any one form to any other form! (of course, all of this would be much easier if we had numbers to work with – but if we can do the general case, then we're all-powerful).

Now that we've derived the quadratic formula, it is time to ask what this can tell us about our problem. What does it mean if our quadratic formula yields two solutions? We have two intercepts, so our graph crosses the x-axis twice. When will this happen? Well, we have that $\pm$ built into our formula, so that must take care of it... but in which case? What happens when $b^2 - 4ac = 0$? Well, in that case $\sqrt{b^2 - 4ac} = 0$, and so $\frac{-b \pm \sqrt{b^2-4ac}}{2a} = \frac{-b}{2a}$. We have only one solution (our quadratic function "bounces" at one point on the x-axis). When will we have two? It must be when $b^2 - 4ac > 0$. And what happens when $b^2 - 4ac < 0$? Well in that case, we'd be having to take the root of a negative number. We would say that we have "no real solutions" (actually, we have two "imaginary solutions"). What does this case looks like? Well, our quadratic function won't cross the x-axis. Because the value of $b^2 - 4ac$ has so much to tell us about how many roots a quadratic equation has, there is a special name for it - *Discriminant*, and it is denoted $D = b^2 - 4ac$. The diagram below demonstrates these cases.



Now, let's get back to solving max/min problems. Usually, the problem requires us either to construct/sketch a quadratic with given properties (i.e. we'd be given some of: curvature, intercepts, vertex, points that lie on the quadratic, etc. - or we'd be given equivalent information, from which we could obtain exactly what we need). Otherwise, we would likely be given a quadratic function that models some relationship between variables, and asked to answer a variety of questions regarding this relationship (as goes the typical example with a ball thrown off a cliff). How do we go about parsing such a problem? First, we must figure out which variables are on the x and y axes (which is the independent variable, and which is the dependent variable). Then, we can sketch a very rough graph - at the very least specifying whether the quadratic is right-side-up or upside-down (recall that the sign of x^2 will give us this information). In the former case, we are solving a minimization problem (why? because we have a unique minimum point), and in the latter case, a maximization problem. At this point, we should check that this matches the intuition of the problem: does it make sense for the problem to have a unique min/max? Is this what we're looking for? Once that's established, we can use the form of the quadratic given to obtain all the information required of us (perhaps we may have to carry out a few transformations along the way). There is really just a set list of questions we'd usually be asked (for instance, using the ball example):

1. When does the ball reach its maximum height? Well, that's just the x-coordinate of the vertex

2. What is the maximum height? Well, that's the y-coordinate of the vertex (note: we can either convert to vertex form to answer this and the previous question; or, we can first find the intercepts, using the quadratic formula, and then obtain the vertex by finding the point on the quadratic halfway between the intercepts)

3. When will the ball hit the ground? Set the dependent variable (in this case, height) to 0, and solve for time. Really, we're looking for the intercepts - in fact, the second intercept (the first intercept will be either at time 0 or at negative time, which is not defined)

4. At what height was the ball originally? Setting the independent variable (time) to 0, gives us the height of the ball at time 0 - which is precisely when it was thrown.

5. At what time will the ball once again be at a height of K (where K is some positive value given to us)? Set height to be K and solve for time.

With regards to questions 3 and 4, an important point should be brought up: because we are dealing in this example with a real problem, we have to consider the domain of our function (see the section about domain and range of functions). In particular, negative time and height is not defined, so our quadratic will be a truncated one, defined only on the first quadrant. We can still manipulate the quadratic function as usual, but before answering the question we must check that our solutions lie within the domain and range of our function (section 4.1).

At this point, an important point must be mentioned about the difference between local and global optima. Note that in a general quadratic function, without extra restrictions on domain and range, for a concave function we have a single minimum - this is our global minimum. Other than that, the graph indefinitely grows in the positive direction, without ever achieving a global maximum.

If, however, we restrict the domain, so that for instance $M \leq x \leq K$ (for some constants M,K), then we will have a global maximum: we must check the value of the function at the endpoints to determine which it is. Note that if x is only bounded on one side (e.g. $x \geq M$, then we $x = M$ is likely a local maximum, but there is no global maximum). A local optima (max/min) occurs where there is a point that is smaller/larger than all the other points in its neighbourhood, but not necessarily smaller/larger than all other points in the function (if it is, then it is a global optima). For an extention of some of these ideas see section 7.2.

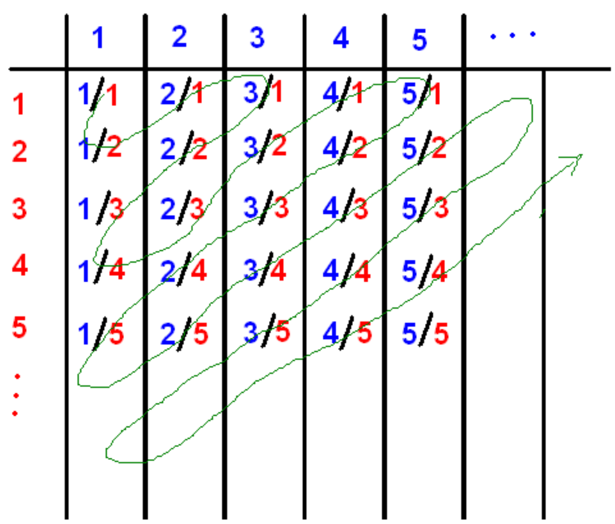
7 Appendix

7.1 Sets: Countability

We can compare the sizes of finite sets directly by counting the number of elements in each set (since these sets are finite, they each contain some fixed number of elements). But what about infinite sets? Can we say anything about the relative sizes of different sets with ininitely many elements? For instance, consider the two infinite sets: $A = \{1, 2, 3, ...\}$ and $B = \{2, 4, 6, ...\}$. Let's imagine you own set A and I own set B. Who has a bigger set? You would tell me "I have a bigger set! I have twice as many elements as you.". In fact, the surprising thing is that effectively, both sets are the same size! Why?
Here's what I'd answer: "Alright, let's start counting how many elements we have in our sets. For each element that you pull out of your set, I will pull one element out of mine. If I run out of elements faster than you, then you win - your set is bigger." So we start: you pull out a 1, I pull out a 2; you pull out a 2, I pull out a 4; you pull out a 3, I pull out a 6... Do you see a pattern? You pull out a 1233, I pull out a 2466... In fact, as long as you don't run out of numbers, I won't run out of numbers either (I just multiply whatever you have by 2). Thus if we can associate each element of one set with some element of another set, we may go on to say that the sets have the same size.
In set theory we talk about a set being **infinitely countable** if we can associate each element of the set with an element from the natural numbers $\mathbb{N}$. In other words, let's say we have some set S of numbers. If we can show that we can enumerate all the numbers in this set, that we can match each number to some natural number, so that we never run out of natural numbers, our set is inifinitely countable precisely because it's infinite and yet we can still go about counting the elements in some order.

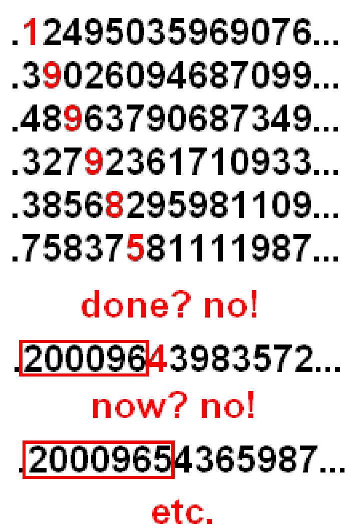
7.1.1 Countability of the rationals

Using the arguments from the previous section, we can see that if we need to prove that the rationals are infinitely countable, we would need to show that we can associate each number from the rationals with a number from the set of natural numbers without skipping any. Recall that if I can pull out an element from my set of rationals for each element you pull out of the set of natural numbers, then our sets are the same size. The natural numbers are infinite, and we define them as countable (we count with them)... so, if we can provide this mapping from the rationals, then we would have shown that the rationals are infinitely countable as well. How can we do this? Using a trick called diagonalization. Create a table (matrix): put the numbers 1 to ∞ as the top row, and the numbers 1 to ∞ as a column; then each square of the table will contain a fraction - with the numerator taken from the corresponding column entry, and the denominator taken from the corresponding row entry. Doing so for all squares of a table (an infinitely many), we would have covered (enumerated) all the rationals. Why? Because in such a way we construct all possible combinations of numerators and denominators. Good: now we're convinced we have in our table all the rationals. Now, how do we associate them with the natural numbers? Travel along the diagonals (see the diagram below), and start counting. We are guaranteed to reach all the numbers in such a way, and to never run out. Note: we don't have that guarantee if we start counting by rows (since the first row is infinite, when are we ever going to get to the second row?)



7.1.2 Uncountability of the reals

"Alright then", you say, "then can't I just come up with a clever trick for counting just about any set?" No! In fact, you may have to come up with a clever trick to show that a set is uncountable! Consider the real numbers. First let's start by assuming that we *can* count/list out all the reals without missing any. Put them all in a table, one real number per row - in any order you wish. Complete the decimal expansions so that there is an infinite number of numbers after the decimal (how? just tack on zeros at the end for fillers). Now you tell me: "ok, I have written down all the real numbers!". But "No!", I say, "I can construct a number that I can guarantee is not in your list!". How? I will construct my number one digit at a time: the first digit of my number can be any digit as long as it differs from the first digit of the first number in your list (by doing so, I've already made sure that my number is different from the first number in your list - since it differs by at least this one digit); the second digit of my number can be any digit as long as it differs from the second digit of the second number in your list (now my number is also different from the second number in your list); I continue to do so, making sure that the i^{th} digit of my number differs from the i^{th} digit of the i^{th} number in your list. Can you see that this guarantees that the number I construct will be different from *every* number in your list? You complain. "Hey, how do I know that your number isn't the $100,000,000^{th}$ number in my list?" Well, even if my number looks like your number based on almost all the other digits, by my construction, my number will differ from yours on the $100,000,000^{th}$ digit. See? There's no way around it! Ok, so you add this new number I created to your list. Now have you got them all? Nope! I can go ahead and use the same rule to construct yet another number that's not in your list! In fact, I can keep doing this an infinite number of times without running out of numbers! To be able to count the numbers in a set, you must have a way of listing (enumerating) them all. Since you can't do this for the reals, they are uncountable! Who can you thank for all these clever tricks? The German mathematician and inventor of set theory, Georg Cantor! (look him up! The history of mathematics is a fascinating thing...)



7.2 A few notes on optimization

In general, we may think of optimization as follows: we have a bunch of independent variables, and one dependent variable (for instance, success of a company, or rating of a movie). We want to be able to reason about how the dependent variable varies with regards to the values of the independent variables. We may want to find where we hit an optima (a max or a min value). We can imagine it this way: we are working with a manifold - an n-dimensional space (since we have n variables) with hills and ditches. We want to be able to find these hills and ditches (because, for instance, we want to discover what it is that makes a company successful, or a movie rating poor). Each hill and ditch looks like a little multi-dimensional quadratic function, so in some sense we can adopt the ideas from our problem-solving with 2D quadratics to this more general problem. However, the overall function we are working with is very complex, and often we can not express it in closed-form (i.e. we can not generate an equation for it). Instead, there are ways we can attempt to explore this complicated landscape in order to be able to say more about it (make some inferences, and begin to approximate it). The easiest way to imagine what we are doing is playing a game: we have a very uneven surface (with hills and ditches), and a little ball rolling over them. We tilt and shake our surface so the ball can roll around different parts of it. Notice that the ball always rolls down to the lowest point in its neighbourhood - it has found the local minima in this neighbourhood! This is exactly what we want: balls will keep rolling down the steepest parts of the landscape, until they can roll down no further (a simplistic explanation of what gradient descent is all about). The problem is that these balls may be trapped in local minima, whereas there may be some other part of the landscape which is the true, global minimum that we strive to get to. What can we do? Well, if we shake the surface hard enough, then we'll increase the chances that the ball will exit the ditch and enter a new one (more technically, what we are doing is randomly perturbing our system). We can also just restart, by randomly dropping the ball into a new starting location (random initialization) and going from there. If we keep doing this, perhaps we may eventually get to the global minimum - however, we have no guarantees. We can only measure our progress by determining if our ball has landed in a minimum that is smaller that the previous ones. With more and more trials we can get a better and better sense of what our landscape looks like. These types of problems (of optimization) are at the heart of machine learning, which is in turn at the heart of many real-world problems. Methods to find local optima, to approximate complex landscapes, to measure probabilities of obtaining good solutions using various techniques, are all very exciting areas of research.