Orders of Approximation

Aaron Szasz

To understand complicated problems in the world, whether in the sciences or in the world of public policy, it is often important to make approximations so that a difficult problem becomes more feasible to solve. One approach to making such approximations and simplifications, a technique that mathematicians and physicists in particular rely on very heavily, is to consider different orders of approximation.

A The general idea and some examples

The basic idea of "orders of approximation" is that a difficult and complex problem can be made easier by considering only the most important information. Less important parts of the problem, things that have a much smaller impact on the solution, can be neglected completely until the most important and fundamental part of the problem has already been solved; these additional, less important considerations will then just produce a small modification to the result.

We use the jargon "orders of approximation" to describe this separation into the more and less important parts of a problem. The *lower order* considerations are more important, while *higher order* ones are less important. In particular, we usually rank them starting with "zeroth order" for information that is absolutely critical to finding even an approximate solution to the problem. "First order" contributions are also generally important for really solving a problem, but are much less important than the zeroth order part, and finding the *first order correction* is useless if you don't already know the zeroth order solution. "Second order" contributions are usually small details, and higher orders above second order can often be ignored.

This will all be much clearer with some examples. One that I think illustrates the idea well is planning a trip across the United States. In particular, suppose you are planning to go from your house in Berkeley to your friend's apartment in New York City. The zeroth order part of your plan is your mode of transportation: will you fly or drive? Without knowing this most important piece of information, doing any further planning of your travel route is completely useless! Once you have the zeroth order solution, you can consider the first order contribution: if flying, then which airport will you leave from (SFO, Oakland, or SJC) and which one will you arrive at (Kennedy, LaGuardia, or Newark); if driving, which major highways will you use for most of your trip, and in which cities (or national parks, etc) will you stop to sleep? Finally, you can consider the second order contribution to your route: if flying, how you get to or from the airports, or if driving, which smaller roads you will use to go between the various highways and also at which particular motels you will stay in each city where you decided to stop.

As you can see from this example, each higher order solution is a *refinement* of the previous order's solution, and the higher order contributions can only be determined once the lower order contributions are already known.

In general, very vague statements are often expressing an implicit zeroth order approximation. For instance, if someone asks you about how many undergrads versus grad students there are at Berkeley, you might say "most Berkeley students are undergrads." That is a zeroth order approximation—you didn't need to specify the exact numbers, yet you still captured the most important information needed to answer the question.

What if you needed to be more precise in your answer? From Wikipedia, you can learn that in Fall 2014, there were a total of 27126 undergraduates at Berkeley, out of 37581 total students, which makes 72.18%. If instead of saying that "most" Berkeley students are undergrads, you said that "about 72%" are undergrads, that would be a higher order approximation. In this example it is harder to precisely delineate different orders of approximation: is the first order approximation that about 3/4 are undergrads, or that 72% are undergrads? Unfortunately, there's no definitive answer to this question, but you can always keep in mind the rule of thumb that lower orders are more important than higher orders.

Here is a third example. Suppose you were planning to move to a new apartment next year, and you calculate that your rent will increase by $80 per month, your utility bills will decrease by $5 per month, and you will have a one-time cost of $50 for renting a van for the move. Let's try to classify these into different orders, ranked by their relative importance.

First of all, $80 per month seems to be about 10% of rent for a typical studio apartment in Berkeley, in other words, it is *much smaller*. So to *zeroth order*, your total yearly expenditures on housing would remain about the same. To *first order*, your costs will increase because your rent is going up. Compared with the change in rent, both the change in utility bills and the one-time moving costs are small, so those fall into the category of *second order* corrections.

A final important note is that the same information can fall into different orders of approximation depending on the situation, for the simple reason that what information is important depends on the precise question you are asking. For this, let's think about Asimov's example of the shape of the earth. If you plan to fly an airplane west from San Francisco until you get back to San Francisco again, to zeroth order you must assume that the earth is round! Higher order corrections are not really important. On the other hand, if you are planning to drive across the United States from San Francisco to New York, your zeroth order approximation about the shape of the earth should be that the earth is flat. (In other words, a machine designed to travel primarily horizontally, such as a car, will be sufficient for your trip.) Furthermore, your first order approximation will not be the curvature of the earth but rather the existence of mountains and rivers and lakes.

To summarize, we can help tackle hard problems by first determining the most important part of the problem (the zeroth order) and then adding additional information (first, second, and higher order corrections) to refine the solution. As pointed out by the webcomic XKCD, physicists like myself may sometimes rely a little too much on this technique, but in fact many problems in the real world really are amenable to this type of analysis. When faced with a difficult question that seems unapproachable, this idea should be one of the tools you consider using.

B (OPTIONAL) Motivation for the idea: mathematics

For me, as a physicist, this idea is actually most clearly understood through the lens of mathematics. In the following section I develop the idea of orders of approximation in the context of approximating the values of a function. Some calculus appears for the sake of completeness, but it's not actually necessary to understand my argument so please don't be deterred by it. This section is completely optional, and you will never be tested on it or expected to know it for this class, but if it sounds interesting to you, please read on - I hope you enjoy it!

Consider the function $f(x) = e^x$, the exponential function. A graph of this function is shown in Figure 1. Suppose you wanted to know the value of this function when $x = 0.01$. Well, that's actually pretty hard to calculate by hand. If $y = e^{0.01}$, that means $y^{100} = e$. To find that by hand,

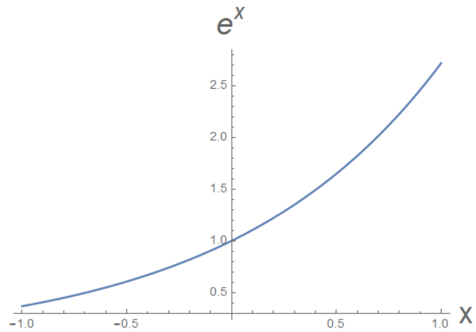

Figure 1: Plot of $f(x) = e^x$

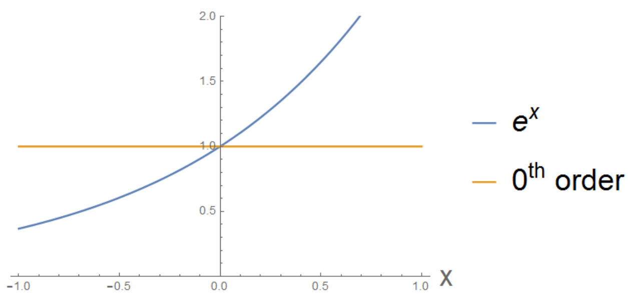

Figure 2: The blue line is $f(x) = e^x$. The orange line is the 0^{th} order approximation.

you could try some value for y, raise it to the hundredth power, and then compare the result to e. If it was too big, you could try a smaller value for y, and vice-versa. Eventually you would find a good approximation to the value of y. This is pretty hard even if you know the value of e, which is itself quite hard to calculate or remember.

But there's a better method, which requires very little work! In particular, the value of x we are looking at is very close to 0, so e^x is close to e^0. (Technically, this is true because $f(x) = e^x$ is a continuous function, but you don't need to worry about that!) So to *lowest order* we can make the approximation that when x is close to 0, e^x is approximately e^0, which is just 1. That's so much easier! This saved us a lot of effort because the function was *much* easier to calculate at one special value of x, which let us avoid doing the more difficult general calculation. It's a pretty good approximation, too—look at Figure 2. You can see that when x is between about -0.1 and 0.1, the error in this approximation is less than about 10%.

What if we want to do better than this approximation? Can we find a calculation that is still easy to do but which gives a more accurate answer than just saying the value is approximately 1? This is where the aforementioned calculus comes in. (Again, if the calculus doesn't help your understanding, you can just ignore it and look at the result, below.) We can construct a Taylor series approximation to our function, which says that when x is small, then

$$f(x) \approx f(0) + x\left[\frac{d}{dx}f(x)\right]_{x=0} + \frac{x^2}{2}\left[\frac{d^2}{dx^2}f(x)\right]_{x=0} + \cdots$$

The approximation we used above, $e^x \approx e^0 = 1$, is the first term in this Taylor series. It is called the 0^{th} order approximation. To get a somewhat better approximation, we can add on the next term, which in the case of $f(x) = e^x$ is just x. This gives us our *first order* approximation,

$$e^x \approx 1 + x$$

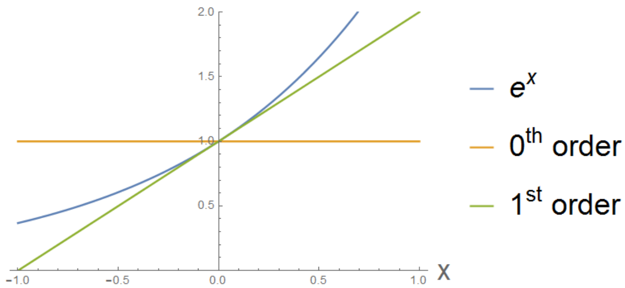

Figure 3: Compare the first order approximation (green) with the original function (blue) and the zeroth order approximation (orange).

when x is small. This new approximation is shown compared with both the original function and the zeroth order approximation in Figure 3. As you can see, the first order approximation is better than the lowest order (zeroth order) approximation, but when x is very close to 0, the zeroth order approximation still gives you most of the answer.

Let's be quantitative about measuring their relative importance. We'll go back to our example of $x = 0.01$. In that case, the $e^x \approx 1.01005$. The zeroth order approximation is $e^x \approx 1$, so the error is 0.01005, or as a percentage of the actual value, about 1%. The first order approximation is $e^x \approx 1 + x = 1.01$, so the error is 0.00005 or 5×10^{-5}, corresponding to a relative error of less than one hundredth of a percent.

The point of all this is that we were able to perform a very difficult calculation ($e^{0.01}$) with very little effort ($1 + 0.01$) and got a very good approximation (0.005% error). By dividing the calculation into different orders of approximation, we give ourselves power to control the tradeoffs we are making between computational effort and accuracy. We could keep going to higher order approximations, each time adding a little bit of difficulty to the calculation but improving the accuracy and precision of our estimate.

This ranking of different orders of approximation continues to be useful so long as the zeroth order value is more important than the first order correction, which is more important than the second order correction, and so forth. In other words, the zeroth order approximation should be mostly correct, while the first order approximation gets rid of most of the error that remains, etc. Now look back at Figure 3, and looks at $x = 1$. It clearly doesn't work here, when the zeroth order approximation is off by more than 50% of the actual value. In that case, we need to reevaluate what counts as zeroth order, first order, etc. In calculus, this is done by a generalization of the Taylor series to approximate functions at values of x that are not near 0. If $x \approx a$, then

$$f(x) \approx f(a) + (x - a)\left[\frac{d}{dx}f(x)\right]_{x=a} + \frac{(x-a)^2}{2}\left[\frac{d^2}{dx^2}f(x)\right]_{x=a} + \cdots$$

Using this method to redo our orders of approximation near $x = 1$, we get the zeroth and first order approximations as shown in Figure 4. When approximating the same basic thing (the function $f(x) = e^x$), what counts as a zeroth order approximation or a first order approximation may vary depending on the exact question we want to answer.

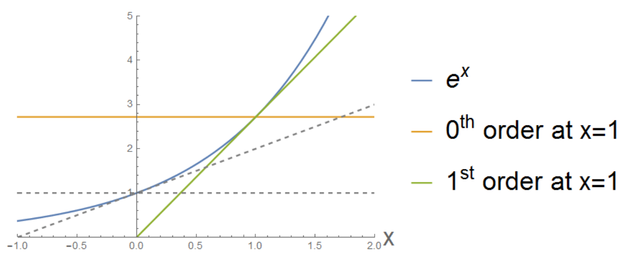

Figure 4: Zeroth and first order approximations for $x \approx 1$.