Guide to Uncertainty Propagation and Error Analysis

Stony Brook Introductory Physics Labs

Contents

0.1 Organization of This Guide

This guide is split into two parts: the main text (with numbered sections) and the appendices (with lettered sections). For the introductory labs (PHY121/122 and PHY133/134), you should know the material from the main text.

Sections 2.3 and 2.4 are both optional, but recommended. Section 2.3 provides some motivation that may help the formulas to stick in your mind, but you can do the calculations without this motivation. Section 2.4 works through a few nice examples in very explicit detail, but some of you may not require this level of detail.

The appendices are purely supplementary material. You may appreciate them in later lab classes, or to further your understanding. They also may be referenced in the lab manuals on an as-needed basis.

A final note, for your convenience: this PDF is hyperlinked. Click on a section header in the table of contents to go to that section. Similarly, clicking on a section or equation number in the text will take you to that section or equation.

1 Introduction

Error is one of the most important features of doing quantitative analysis.

Suppose you were doing a lab to check whether $F = ma$. Your measured value of F is 0.99N, and your measured value of ma is 1.02N.

Is this good? Bad? Who knows? I mean, it's close, so perhaps you can say it agrees. If, on the other hand, you did a super-precise measurement, then maybe you can actually say that $F = ma$ is only *approximately* true.

Whether or not you confirm or deny Newton's second law depends on the precision of your experiment! So, in order to do our quantiative analysis properly and come to a conclusion as to whether it is correct or not, we need to know our errors.

1.1 Random vs. Systemic Error

There are two kinds of error that one deals with in the lab: **random error** and **systemic error**.

Random error, per the name, are statistical effects that push your data every which way from the "true" result, with no consistency to its effects. These will result in your data being right on average, but any individual data point will be off. This results in *uncertainty*. Random error won't make you wrong (usually), but will limit how much you can say. Data that doesn't have much random error is called *precise*.

Systemic errors are effects that come from aspects of your system that you have not considered. These result in *bias*, as well as more complicated effects. That is to say, systemic error will "push" your data in some *consistent, reliable* way. It won't (necessarily) increase the scatter of your data, so it won't make your data look "all over the place." Rather, it takes the nice straight line your data should make, and turns it into a line of a different slope (or another curve altogether).

1.2 Quantifying Error: Absolute vs. Relative Error

There are two general ways to talk about error: absolute error and relative error.

If you have a quantity X that you are measuring, the **absolute error** in X, denoted as σ_X or $\sigma(X)$, is simply the amount by which you are (or could be) off. For instance, if you measure something to be 5.89m, then you might reasonably say that your uncertainty is 1cm - that is to say, you could have been off by a

centimeter.

You could also report this same uncertainty as a **relative error**, denoted as $\sigma_{\text{rel}}(X)$. This is when you compare the size of your error to the size of the original quantity.[1] The formula for relative error is:

$$\sigma_{\text{rel}}X = \frac{\sigma_X}{|X|} \tag{1}$$

Thus, in the above example, your 1cm uncertainty on your 5.89m measurement would turn into a relative error of 0.0016. In other words, the absolute uncertainty on your measurement is 0.0016 times your original measurement.

You may be more familiar with the idea of relative error from the closely related *percent error*, which is just 100 times larger.

Both relative and absolute errors have advantages and disadvantages as methods of quantifying your error.

Absolute error is straightforward and easier to understand. They are also necessary for direct comparison: if I ask you whether 5cm $\pm$ 1mm agrees with a theoretical value of 4.8cm, that's easy to tell; if I ask whether 5cm with a relative uncertainty of 0.2 agrees with the same theoretical value, that's not nearly as obvious.

As such, when you state a quantity in your lab report, it should typically be written $X \pm \sigma_X$ - that is to say, you talk about uncertainties with the absolute error when reporting them. (This will also be the kind of error that these manuals will be referring to if they don't explicitly say "relative error.")

Relative errors give a sense of how "big" an error is in a general sense. That is to say, if I tell you I have a 5% error, you have an immediate sense of how much I'm off by; if I tell you I have a 1cm error, whether that's big or small depends on what I'm measuring.

In particular, relative errors can be compared across different sources of errors: if one quantity has a relative uncertainty of 0.2 and another of 0.003, it's easy to see that the contribution of the second to your final amount of error is negligible. If I tell you one has an error of 1cm and the other an error of 0.2s, you have no sense of which is the more relevant contribution. Therefore, relative errors are the better things to look at if you want to compare errors of different varieties.

[1] If your "original quantity" is zero (or, more generally, you are trying to measure something which should be equal to zero), then you shouldn't use relative error. As a more concise rule of thumb: if your relative error is larger than 1, you should just use absolute error.

Finally, a note on units: absolute errors will have the same units as the original quantity,[2] so a time measured in seconds will have an uncertainty measured in seconds, etc.; therefore, they will only be unitless if the original quantity is unitless. Relative errors are *always* unitless.[3]

1.3 Comparing Quantities with Uncertainties

Once you have an uncertainty, you can use it to compare quantities. In this class, we will use two methods to do so.

The first method is used if only one quantity has uncertainty. This is typical if one is comparing a calculated quantity from lab to a "theoretical value."[4] In this case, one simply sees whether the quantity without uncertainty (the "theoretical value") lies within the uncertainty range of the experimental value.

That is to say: if one is comparing $A \pm \sigma_A$ to B, one simply sees whether or not it is the case that $A - \sigma_A < B < A + \sigma_A$.

The second method is used if one has two quantities with uncertainty. This is usually for comparing different measurements to each other. In this case, one sees whether the uncertainty ranges overlap at all.[5] If these ranges overlap, then we say the two values are consistent - both *could* be correct to within uncertainty simultaneously.

More concisely: to compare $A \pm \sigma_A$ to $B \pm \sigma_B$, we see whether we have both that $A + \sigma_A > B - \sigma_B$ and that $A - \sigma_A < B + \sigma_B$.

2 Random Error

2.1 Basic Uncertainties

In order to get a sense of what our uncertainty will be in our final result, we first should know what uncertainty comes from our measurements themselves. How we do this depends on the nature of our uncertainty.

The first thing we have to do is to determine what the dominant source of our uncertainty is. Is it the limits of our measuring apparatus? Is it the physical arrangement of our device?

---

[2] Actually, you could change the modifier if you wanted to - for instance, 1.2cm $\pm$ 1mm - but there's rarely a reason to do so.

[3] This is a part of what makes them directly comparable.

[4] These theoretical values typically also have uncertainty, but such uncertainties are small enough compared to the uncertainties in our experiments as to be neglected.

[5] For the sake of introductory labs, this is how we will do comparisons. In general, a slightly better method is to see if the difference (with uncertainties propagated) is consistent with zero (in the sense of a theoretical value). This accounts more accurately for uncorrelated errors.

For instance: if you're measuring the size of a string, the limit could generally be taken to be the markings on the ruler. However, if you were measuring the distance between two houses down the street, there would likely be significant uncertainty from the size of the houses themselves, because they aren't "point houses."

Your precision won't be 1mm there, even if you measure with a super-precise ruler, because the ruler isn't the limit of your uncertainty. Even if you specified an exact position (like "center of mass of the house"), you're unlikely to be able to know what that position is to within 1mm.

Then, we need to figure out how to determine this uncertainty. There are a variety of simple ways to do this:

- For a digital device which directly outputs a reading (like a digital scale), you can take the uncertainty to be given by the last digit the device outputs. For instance, if it reads 4.18, then the uncertainty would be 0.01 (in whatever units the scale reads).
  - Exception: if the reading from the device is significantly fluctuating, you should take the fluctuations to be the typical uncertainty in the reading, instead.
  - Some devices have multiple settings which will change what that last digit is. You should use the last digit from the setting you were using when you made the measurement. (As such, you should also, as a rule, use the setting with the smallest last digit that you can, because this will minimize your uncertainty from that device.)[6]
- For something you are reading by eye, a typical rule is "half the smallest division." This applies *unless* you think there is a reason it should be larger: say, you're measuring a long distance, and are having a hard time getting things to line up right, or the aforementioned house example.
- For pre-built devices (like a set of mass weights), there are usually a standard set of tolerances[7] given by the manufacturer, which tells you how much to expect they could be off by (at most). Such tolerances should be specified in your lab manual or by your TA, where relevant.

However, some errors are not easily estimated. For instance, how precisely can you hit the stop button on a stopwatch? (This will vary from person to person, and you probably haven't measured that for yourself!)

[6] A higher sensitivity might lead to you seeing fluctuations on a sensitive reading. This is actually a *good* thing - you are not unnecessarily increasing your uncertainty by rounding.

[7] E.g., for our masses, see the NIST documentation defining standards for masses: `https://www.gpo.gov/fdsys/pkg/GOVPUB-C13-50fecb383812d067b82bca54d84af943/pdf/GOVPUB-C13-50fecb383812d067b82bca54d84af943.pdf`

In this case, there is a statistically rigorous[8] way to determine the uncertainty.

We first do the measurement of the same quantity, X, some number of (statistically-independent) times N. Call these different measurements x_1, x_2, ..., x_N. Then, we take the average of the results, $\overline{X}$:

$$\overline{X} = \frac{x_1 + x_2 + \ldots + x_N}{N} \tag{2}$$

This is then our best estimate of the relevant measurement. The uncertainty of this average is:

$$\sigma_{\overline{X}} = \sqrt{\frac{(x_1 - \overline{X})^2 + (x_2 - \overline{X})^2 + \ldots + (x_N - \overline{X})^2}{N(N-1)}} \tag{3}$$

If you want the uncertainty of an individual measurement x_i, you instead use the formula:

$$\sigma_{x_i} = \sqrt{\frac{(x_1 - \overline{X})^2 + (x_2 - \overline{X})^2 + \ldots + (x_N - \overline{X})^2}{N-1}} \tag{4}$$

This is handy if you're trying to use your measurements to estimate the uncertainty of some future measurement of the same type. For instance, you could measure the uncertainty in hitting a stopwatch on one set of measurements, then apply that uncertainty to another set of measurements where reaction time is a factor.

Note, by the way, that the two quantities are closely related:[9]

$$\sigma_{x_i} = \sqrt{N}\sigma_{\overline{X}} \tag{5}$$

Note also how these quantities scale when taking many measurements: σ_{x_i} approaches a constant (the sum in the numerator has N terms and you divide by approximately N), whereas $\sigma_{\overline{X}}$ is inversely proportional to $\sqrt{N}$. This is a general statistical idea that is valuable to have in your intuition.

At the end of the day, though, most of these error estimates are just that: estimates. Professional physicists (and scientists in general) spend a good deal of time getting statistically-rigorous uncertainty measurements; for this lab, these rough estimates will suffice. Use your best judgement.

[8] With appropriate mathematical assumptions, of course. Technically, we assume that these measurements are *independent* and *identically distributed* - that is to say, that they don't influence each other and all function identically (with the same probability distribution of measurements, in principle).

[9] Formally, you can prove this result - with appropriate assumptions on the individual measurements - using just the addition propagation formula that we will derive in the next section. If you are mathematically inclined, take a look!

2.2 Uncertainty Propagation Formulas

Usually, you don't directly measure your final result, though. Usually, you measure some different quantities and *calculate* the final result. E.g.: if we want to know the area of something, we'll measure some lengths and apply geometry formulas, not directly measure area.

Therefore, we are obligated to ask: once you have these basic measurement uncertainties, how do you determine the uncertainty in your final results? Somehow, you need to turn uncertainties in these quantities into uncertainties in your final *calculated* results.

The first step to doing so is to identify how you calculated your quantity. For instance, suppose you calculate the area of a rectangle, $A = lw$, from the length and width of the rectangle. Then, of course, you calculated A as the product of l and w.

The second step is to break this calculation into a sequence of elementary steps. $A = lw$ is already there: you're just multiplying two things. However, now let's suppose you're calculating the uncertainty in the area of a trapezoid, $A = \frac{1}{2}(b_1 + b_2)h$, where you calculated the area of the trapezoid based on the two bases and the height. Then, you can break that formula into performing the following sequence of steps:

1. Add b_1 to b_2.
2. Multiply by h.
3. Muliply by $\frac{1}{2}$.

Each of those corresponds to a single computation. (Each such "simple step" will have a separate formula we'll give in a minute.)

The final step is to propagate error for each step. To do so, we'll need some special formulas that tell us how this is done; for instance, how to determine the error in $A + B$ based on the errors in A and B individually.

The following formulae indicate how to do this, given two experimental quantities[10] A and B and constants (i.e., numbers without error[11] c and n:

$$\sigma(A + c) = \sigma_A \tag{6}$$

$$\sigma(cA) = |c|\sigma_A \tag{7}$$

[10] Different experimental quantities, by the way - you can't use, e.g., the addition formula when A and B are the same!

[11] These include mathematical constants, like 2 or π, as well as physical quantities whose errors are so small we can ignore them, such as a theoretical value of g.

$$\sigma(A \pm B) = \sqrt{(\sigma_A)^2 + (\sigma_B)^2} \tag{8}$$

$$\sigma(A \times B) = |AB| \sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2} \tag{9}$$

$$\sigma\left(\frac{A}{B}\right) = \left|\frac{A}{B}\right| \sqrt{\left(\frac{\sigma_A}{A}\right)^2 + \left(\frac{\sigma_B}{B}\right)^2} \tag{10}$$

$$\sigma(A^n) = |n||A|^{n-1}\sigma_A \tag{11}$$

$$\sigma\left(\frac{1}{A}\right) = \frac{\sigma_A}{A^2} \tag{12}$$

Some of these can be more conveniently expressed in terms of relative errors:

$$\sigma_{\text{rel}}(cA) = \sigma_{\text{rel}} A \tag{13}$$

$$\sigma_{\text{rel}}(A \times B) = \sigma_{\text{rel}}\left(\frac{A}{B}\right) = \sqrt{(\sigma_{\text{rel}} A)^2 + (\sigma_{\text{rel}} B)^2} \tag{14}$$

$$\sigma_{\text{rel}}(A^n) = |n|\sigma_{\text{rel}} A \tag{15}$$

$$\sigma_{\text{rel}}\left(\frac{1}{A}\right) = \sigma_{\text{rel}} A \tag{16}$$

You may recognize that several of them have the same squaring, adding, and square-rooting pattern of different quantities. If we take the quantity $\sqrt{x^2 + y^2}$, that is called "adding x and y *in quadrature*." Using that terminology, we can express the above formulas in plain English:

- When you add a constant to a quantity with error, the absolute error remains the same (6).
- When you multiply a quantity with error by a constant, the relative error remains the same (13).
- When you add or subtract two quantities with error, you add the absolute errors in quadrature to get the absolute error of the sum (8).
- When you multiply or divide quantities with error, you add the relative errors in quadrature to get the relative error of the product or quotient (14).
- When you take a power (without error) of a quantity with error, you multiply the relative error by the absolute value of the exponent (15).

- When you take one over a quantity with error, the relative error remains the same (16).

If you're adding and subtracting a bunch of things, or multiplying and dividing a bunch of things, the formulas (8) and (14) extend basically like you would expect:

$$\sigma(A_1 \pm A_2 \pm A_3 \pm \ldots) = \sqrt{\sigma_{A_1}^2 + \sigma_{A_2}^2 + \sigma_{A_3}^2 + \ldots} \tag{17}$$

$$\sigma_{\text{rel}}\left(\frac{A_1 \times A_2 \times \ldots}{B_1 \times B_2 \times \ldots}\right) = \sqrt{(\sigma_{\text{rel}} A_1)^2 + (\sigma_{\text{rel}} A_2)^2 + \ldots + (\sigma_{\text{rel}} B_1)^2 + \ldots} \tag{18}$$

You may find the above formulas more convenient than breaking such things into a whole bunch of individual multiplications.

You can even take formula (18) and extend it to combinations of products and powers. Taking a canonical example rather than a fully general case for simplicity, we can get formulas like:

$$\sigma_{\text{rel}}\left(\frac{cA^n B^m}{C^k}\right) = \sqrt{(n\sigma_{\text{rel}}(A))^2 + (m\sigma_{\text{rel}}(B))^2 + (k\sigma_{\text{rel}}(C))^2} \tag{19}$$

When you don't have any operations other than multiplication, division, and (constant) powers, then equations like that are all you need.[12]

2.3 Where do those formulas come from?

Having formulas is nice, but it's even nicer to understand those formulas. While this won't be a complete explanation, nor will it describe how to take errors of more complicated formulas (see Appendix A for that), it should suffice to give you some idea of why the formulas are the way that they are.

Let's begin with adding two things. Let's suppose we make two measurements of the quantities A and B. Let's say A and B are the actual values, but we measure $A + \delta A$ and $B + \delta B$ - that is to say, δA and δB are the amounts we are off by. (We don't actually know these numbers, but they do exist.)

Now, when we take $A + B$, we can ask how far are we off there; i.e., what is $\delta(A + B)$:

$$\delta(A + B) = ((A + \delta A) + (B + \delta B)) - (A + B) = \delta A + \delta B$$

[12] As a point of advice: if you learn how to make and use formulas like that, it will make your life in this course much easier in the long run. Some later labs involve propagations with many multiplications at once, and recognizing the logic behind equation (19) will let you immediately know the answer rather than doing out a whole bunch of algebra.

This is true for an individual measurement, but that's not the same thing as (8). Why?

Well, because sometimes, δA and δB will have opposite signs. Sometimes, our errors will cancel. Just adding uncertainties doesn't account for that possibility.

To cut short a longer explanation,[13] if you directly add (or subtract) errors in two different quantities, like the formula above, you add the corresponding uncertainties in quadrature. This accounts for the possibility of cancelling, mathematically speaking.

Alright. If we have that, then we do get (8). Now, let's look at the multiplication formula, (9).

Let's do the same procedure: take $A+\delta A$ and $B+\delta B$ as our measurements, and calculate $\delta(AB)$:

$$\delta(AB) = (A+\delta A)(B+\delta B) - AB = A\delta B + B\delta A + \delta A\delta B$$

Now, we're going to make a simplification: we're going to assume our errors are relatively "small."[14] If δA and δB are both small, then their product is *really* small, and so we can ignore it. Therefore, approximately, we find:

$$\delta(AB) \approx A\delta B + B\delta A$$

Alright, now we're going apply the same principle as above: we have two errors in different quantities added together, so we're going to add in quadrature to get the uncertainty propagation formula. This gives us:

$$\sigma(AB) = \sqrt{(A\sigma_B)^2 + (B\sigma_A)^2}$$

Dividing the entire equation by AB and putting that inside the square-root then gives us:

[13] We technically quantify uncertainty in the following way: we take the average value of the *square* of our error, and square root that: $\sigma_x = \sqrt{\overline{(\delta x)^2}}$. If we square both sides of $\delta(A+B) = \delta A + \delta B$, we get

$$(\delta(A+B))^2 = \delta A^2 + \delta B^2 + 2\delta A \cdot \delta B$$

Now, take the average value of both sides, and you find

$$(\sigma(A+B))^2 = \sigma_A^2 + \sigma_B^2 + 2\overline{\delta A \delta B}$$

We now make an assumption, known as *statistical independence*: we assume $\overline{\delta A \delta B} = 0$, which gives us (8). Practically speaking, this assumption is true if A and B depend entirely on separate measurements; to make things easy, in this lab, we'll just always pretend that it's true. If it is not true, the quantities A and B are said to be *correlated*, and that makes error analysis significantly harder.

[14] More mathematically precisely: we assume that the relative errors are much less than 1.

$$\frac{\sigma(AB)}{AB} = \sqrt{\frac{A^2(\sigma_B)^2}{(AB)^2} + \frac{B^2(\sigma_A)^2}{(AB)^2}}$$

This is the same as equation (14) for multiplication, from which you can get (9) by multiplying by AB.

The rest follow from similar rules - if you really want to get where they come from, you can work through some of them yourself. The multiplication by a constant rule is straightforward. The power rule you can do for positive integer powers using the binomial theorem (and remember that errors squared, cubed, etc. are "very small" and you can ignore them).

For non-integer powers and for the division formula, you'll need some calculus, unfortunately - see Appendix A.3 if you want to see how to get those.

2.4 Uncertainty Propagation Examples

Let's now work through some examples so you can see how these things work. These examples are all borrowed from high-school math, so there'll be no physics here; everything will be calculations you should recognize.

2.4.1 Simplest Example

Let me begin with the simplest possible kind of example: a propagation that exactly matches one of our basic uncertainty formulas.

Let's suppose you measure the length and width of a rectangle, and want to measure the area. The formula for this, is of course, $A = lw$, so we've identified how we calculated area: by multiplying length and width.

Therefore, since we are doing a product, we use the multiplication formula, (9). Since everything is positive, we can ignore the absolute values.

We are multiplying l and w, so we replace every A in (9) by an l and every B by a w. Note that the A in formula (9) is a "dummy variable" - it has nothing to do with the area!

This gives us the formula:

$$\sigma(lw) = lw\sqrt{\left(\frac{\sigma_l}{l}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2}$$

Since $A = lw$, the left-hand side of that equation *is* just σ_A, and so gives us our uncertainty.

As an alternative approach, we could have used (14), which gives us:

$$\sigma_{\text{rel}}(A) = \sqrt{\sigma_{\text{rel}}(l)^2 + \sigma_{\text{rel}}(w)^2}$$

This simplifies (after applying the definition of relative error) to the equivalent formula:

$$\sigma_A = A\sqrt{\left(\frac{\sigma_l}{l}\right)^2 + \left(\frac{\sigma_w}{w}\right)^2}$$

2.4.2 Non-Trivial Example

Now, let's see what happens when we have something that doesn't match one of our basic uncertainty formulas.

Let's now suppose that we've measured the diameter d of a circle and want to know its area. The formula for the area of a circle is $A = \pi r^2$, but we didn't measure radius, we measured diameter.

There are two ways to deal with this. One is simply to modify our formula to $A = \frac{\pi}{4}d^2$, and propagate through that formula. That works, but we'll take a slightly different (but equivalent) approach for the sake of this example.

First, we compute the radius in terms of the diameter, as $r = d/2$. So we just need to reference our formula for dividing by a constant... oops, we don't have one!

However, we have one for *multiplying* by a constant, so we just write our calculation in a slightly different way: $r = \frac{1}{2}d$. There: now we know which formula to use, formula (7), because we're multiplying our number with uncertainty (d) by a number without uncertainty ($\frac{1}{2}$).

So, as we did previously, we take formula (7), and make the appropriate substitutions: we replace c with $\frac{1}{2}$, and A with d. This gets us:

$$\sigma(\frac{1}{2}d) = \frac{1}{2}\sigma_d$$

Since $r = \frac{1}{2}d$, we thus have $\sigma_r = \frac{\sigma_d}{2}$. So we've done one step of our uncertainty propagation.

Now, to return to our original problem: finding the area of the circle, now using the formula $A = \pi r^2$ to calculate in terms of r.

Again, we are left without any one of our formulas representing the operation we are doing. However, we are actually doing two things: squaring r, and then multiplying that square by π.

We do each step separately. We begin by taking the square of r. This is a power, so we follow formula (11).[15] We take A and replace it by r, and n and replace it by 2, giving us:

$$\sigma(r^2) = |2||r|^{2-1}\sigma_r = 2r\sigma_r$$

Alright, now we want to multiply r^2 by π. If we look at r^2 as *one thing*, at which point this looks like formula (7). If we replace c by π and A by r^2 (again, treating the expression r^2 as one "thing" in our propagation), then formula (7) turns into:

$$\sigma(\pi r^2) = \pi\sigma_{r^2}$$

Now we have two formulas: we can just plug the first into the second by substitution! Doing so gives us our final formula for our uncertainty in area:

$$\sigma_A = \sigma(\pi r^2) = \pi\sigma_{r^2} = \pi(2r\sigma_r)$$

So that gives us our formula for the uncertainty of a circle in terms of our uncertainty in radius.[16] We can use the propagation from diameter to radius as an intermediary step, then propagate from radius to area.

It's worth specifically pointing out that we showed off *two* ways to break a calculation into steps in this example. (They're really the same, but they look a little different.)

In going from diameter to radius, we showed how to use an "intermediary calculation" to help you get from one quantity to another. You could explicitly write r and its uncertainty on your data table, do that calculation, then do the second calculation. This is one method of doing a complicated propagation.

Alternatively, you can do what we did in going from radius to area: we had two steps, so we did them out separately, then used algebra to combine them. We then got a "final" formula that we could plug in. This is also a perfectly acceptable way to do a complicated propagation.[17]

We could have used either of those methods for both steps, if we so chose; we just mixed-and-matched as convenient, because r is an intermediary quantity with a nice name (and physical intuition), whereas r^2 is not. The data tables are generally designed for this mix-and-match approach. If you like writing

[15] Note: even though this is also technically a multiplication, we can't just use the multiplication formula - it will give us a different result. The reasons behind this are a little subtle (although if you work through the logic in section 2.3, you may have some intuition), but basically, equation (9) only works if you're multiplying *different* things. If you're multiplying the same thing, remember to always use the power formula!

[16] You may have noticed that this result is the uncertainty in r times the perimeter of the circle. This is not a coincidence, but is also only tangentially related to error propagation.

[17] If you use this method, it is highly recommended that you do the algebra to simplify your formula before you type it into any calculator or spreadsheet program, where possible. This helps to avoid errors of incorrectly placed parentheses and the like.

everything out explicitly with the first method, you can always do the extra calculations off to the side somewhere[18] and combine them into your final answer.

2.4.3 Hard Example

Now, let's do one more example, for good measure. This is approximately equivalent in complexity to our most complicated propagations, so serves as a nice final case to look at.

Let's suppose you want to calculate the uncertainty in the area of a trapezoid for which we've measured the two bases and the height. Recall the formula for the area of a trapezoid:

$$A = \frac{1}{2}(b_1 + b_2)h$$

We want to break this into steps as we did in the previous example. As mentioned in section 2.2, you can split this calculation as follows:

1. Add b_1 to b_2.
2. Multiply by h.
3. Muliply by $\frac{1}{2}$.

Therefore, we calculate the uncertainty using formulas (8), (9), and (7), in order, to get the following result:

$$\sigma(b_1 + b_2) = \sqrt{\sigma_{b_1}^2 + \sigma_{b_2}^2}$$

$$\sigma((b_1 + b_2)h) = ((b_1 + b_2)h)\sqrt{\left(\frac{\sigma(b_1 + b_2)}{b_1 + b_2}\right) + \left(\frac{\sigma_h}{h}\right)^2}$$

$$\sigma_A = \sigma\left[\left(\frac{1}{2}\right)((b_1 + b_2)h)\right] = \frac{1}{2}\sigma((b_1 + b_2)h)$$

Combining all of these together (and noting that when we combine all the constants out front, they just calculate out to a factor of A), we get our final error propagation formula:

$$\sigma_A = A\sqrt{\left(\frac{\sqrt{\sigma_{b_1}^2 + \sigma_{b_2}^2}}{b_1 + b_2}\right)^2 + \left(\frac{\sigma_h}{h}\right)^2}$$

That explanation is deliberately a little briefer than previous sections - you should work through the details for yourself, and try to replicate that result. After all, in the lab, you'll have to figure out the formulas for yourself!

[18] If you do this, it is advisable to label these extra calculations so your TA can more easily see what you did and give you appropriate partial credit if necessary.

3 Systemic Error

3.1 Isolating Systemic Errors

Systemic errors are a lot harder to deal with than random errors. This is for two main reasons.

Firstly, systemic errors cannot be reduced by running more trials of the same experiment. They will be present every time, with (roughly) the same effect, and will not average out to zero.

For example, let's suppose the lab was doing an experiment to measure $g = 9.8\text{m/s}^2$. Random errors would make one group measure 10m/s^2, another measure 9.7m/s^2, etc., but on average the whole class would probably get something close to 9.8m/s^2. For a systemic error, this isn't true: it might make everyone in the class measure $g = 9.9\text{m/s}^2$, so that even when you average you don't get the right answer.

Systemic errors *can* be reduced by running multiple entirely *different* experiments to find the same quantity, but that requires coming up with multiple ways to test the same thing. Experimental physicists frequently have multiple independent experiments testing the same thing for this exact reason.

Secondly, systemic errors can't be *measured* by taking the same experiment multiple times and observing how much your results move around, so there's no analogy for formulas (3) and (4). The only way to understand the size of your systemic errors (from your experiment alone) is by understanding what causes them.

Therefore, identifying systemic errors requires knowing the physics behind your system, and understanding your experimental apparatus really well. You have to see where some other physical effect could produce an impact on your results, and what impact that would be.

3.2 Determining the Effects of Systemic Errors

There are two aspects of any systemic error that you would want to know: what kind of effect it has, and how large this effect is.

Both can be determined by re-doing calculations after you include some estimates of any new physical parameters. That can usually be done if it needs to be, but it's often impractical.

Sometimes, however, we can make things easier by making quick approximations, so that's what we'll do here. This means we won't actually calculate the quantitative effects of our systemic errors. Instead, we'll get a qualitative

sense of what they're doing and how important they are.

Many systemic errors can be simply stated as making you over- or underestimate some quantity. You can determine, based on that, whether your final quantity would be too small or too large.

For instance, let's suppose you have a mass that tells you it's 1kg, but it's really dirty. Perhaps that dirt is adding extra weight to the mass. This means that the total mass your experiment would be using would be bigger than 1kg, so you underestimated the value of your mass.

Suppose your final goal was to measure the force exerted on this mass, based on the acceleration it undergoes, using $F = ma$. Well, if you underestimate m, then you would also underestimate F.

These kinds of reasoning - "I underestimated this, so I underestimated that" (or, if something is in the denominator, "I underestimated this, so I *over*estimated that") - can tell you what kind of impact your systemic error will have on your final result.[19]

The second thing you can do is estimate the size of these errors. This allows you to determine whether or not such an error is important.

To a rough approximation, you can get a sense of their size with the following rule of thumb: the relative error in your final quantity is usually about the same as the relative error in your original quantity.

So, taking the above example of a dirty mass again, the relative error resulting from the dirt is the mass of the dirt divided by the mass of the original mass (1kg). If you have an estimate of how much the dirt might weigh (say, 1g), you can convert this into a relative error on the mass (in this case, 0.001).

You can then take this relative error as an order-of-magnitude estimate of the relative error in F, so if F was 7.2N, then your absolute error in F would be roughly .0072N. Assuming your random error in F is significantly higher than this, you can neglect this error as insignificant.

[19] Sometimes, your systemic errors don't just increase or decrease everything, or they do so in a complicated way. For instance, let's suppose you fit a line $y = kx$ to a supposedly-linear graph. However, there's a systemic error that makes a corrected formula $y = kx + ax^2$, for some small number a. Will not including this systemic error make your measurement bigger or smaller? There's no easy way to tell, unfortunately - you would have to go through a more detailed analysis. Systemic errors are generally hard!

4 Other Things

4.1 Things that are NOT errors

There are many things which one would call "errors" in colloquial language, but are not errors in a *scientific* sense. These include:

- Calculation errors: If you think your calculations have a problem, fix the problem. If you can't figure out how to fix the problem, but you know you made one (and perhaps can even identify where it is), feel free to mention it, but know that it's a separate concern from the uncertainties and systemic errors we've discussed here.
- Human error: again - if there's a problem, fix it. If it's too late to fix when you realize that you made a significant mistake (say, you realize it after you went home), that is worth mentioning, but it's not a "source of error" in the sense we've been discussing - it's just a mistake.
- Random deviations from procedure that don't affect results: Let's say the lab tells you to take a length of 0.3m for some piece of rope, in order to have the experiment work well. If you have a 31cm rope, and then do your calculations with $L = 31\text{cm}$, then that's not an error, even if it's a deviation from procedure. If you do all calculations in a way that is accurate to your numbers, then your results should still give you the right result, even if you deviate from procedure, so this won't result in errors![20]

4.2 Significant Figures

In previous science classes, a great deal of attention may have been paid attention to significant figures (how many decimal places you should work to) and rules for manipulating them, trailing zeros, etc.

In general, significant figures serve two roles:

1. A stand-in for uncertainties. Significant figure rules are easy to follow - easier than uncertainty propagation, certainly - and serve as a functional replacement for them if actual uncertainties are not being calculated.
2. Aesthetic reasons. There's no practical reason to go to the twentieth decimal point if your error picks up at the second; those last decimal places aren't important, because your error is larger than them.

Since we'll be using actual uncertainties, the first reason isn't relevant. Therefore, significant figures are, in this class, largely a matter of taste. In the same sense as English grammar: you are allowed to break the rules if you

[20] Exception: some deviations will cause a systemic error to be less small, perhaps to the point of being problematic. For instance, if you are doing an experiment involving an ideally massless string, and you use way too much string, perhaps the approximation of a massless string isn't valid anymore.

have a good reason for doing so. The rules serve as a guide to good communication, but should be ignored where they inhibit it.

As a general rule of thumb, work to three significant figures in your quantities in this class, with a number on your uncertainty to match. You can go further if you have a very-precisely measured quantity, or two if your error is very large (say, bigger than 10%). Generally, though, three sig figs will be both enough for rounding errors to not matter and few enough that it won't be aesthetically abhorrent.

(As a matter of presentation in our lab: don't feel obligated to round on your data table. However, when you write your report, you should use a reasonable number of digits in your discussion - don't copy ten decimal places every time you write your answer.)

4.3 Number Sense for Uncertainties

Finally, here are a few guidelines to help you identify when your uncertainty propagation has problems:

- Generally, unless you're doing addition, relative uncertainty shouldn't change too much. A good check is to calculate relative uncertainties for every kind of basic measurement you have; if your final relative uncertainty is *much* higher or lower than the largest of these, there's probably a mistake somewhere.
- When you make a plot, look at your error bars. Are they larger than the range of your data? (Does your plot look almost like a bunch of vertical lines, rather than dots with error bars?) If so, you probably have a mistake somewhere (unless you have reason to think otherwise).
- If your uncertainty is larger than the value you measured, it is *almost certainly* wrong. There are a few exceptions (e.g., position, which can be measured as zero with an uncertainty), but generally, uncertainty greater than value is bad.
- Finally: *think* about the uncertainty you calculated, physically.[21] Does it make sense? (If it's bigger than you expect, is there a part of your experiment that was obviously highly uncertain? If it's smaller than you expect, why - i.e., what "experimental tricks" enable you to make it small?)

If you identify a mistake, remember the saying "garbage in, garbage out" - an early bad propagation will (usually) make later ones not make sense. Try to find the *earliest* place where your results don't make sense (using the above tricks), since that's where you are most likely to find your mistake.

[21] This really extends beyond uncertainties, of course - think about your measurements, your calculated quantities, your final results, and *all* of their uncertainties.

A Mathematical Details

A.1 What is an uncertainty really?

A proper definition of what an "uncertainty" really is was briefly explained in footnote 13, but here's an elaboration in a bit more detail.

Suppose we are taking a measurement of some quantity X. What, then, really, is σ_X?

Well, this quantity X will have some average value of its measurement.[22] Let's call this value μ (not $\overline{X}$, for reasons to be explained in the next section).

For any given measurement of X, we can then talk about $X - \mu$ for that measurement. However, on average, this difference will be zero - X will be above μ about as often as it will be below. So that doesn't help.

It might seem reasonable to talk about the average distance of X from μ - that is to say, the average value of $|X - \mu|$. One can do this, but it's very messy, and doesn't follow nice rules.

A better quantity to look at is the *root mean square* distance of X from μ (with "root mean square" often abbreviated as *rms*). This is given by taking the square, then the mean, then the root, yielding the formula:

$$\sigma_X = \text{RMS Distance} = \sqrt{\overline{(X-\mu)^2}} \tag{20}$$

This turns out to have much nicer properties than our previous calculation, and so we use this as our definition of σ_X.

Of course, your uncertainty estimates "by eye" won't be this exact quantity, but they'll be vaguely similar, and so we'll use the same formulas that this follows. (We discuss this at more length in Appendix B.1.)

A.2 Why $N - 1$?

Now that (in the previous section) we defined what the uncertainty actually is, we can see some of the motivation behind the formula for the uncertainty in a single measurement, (4). In fact, this is *almost* the same as the calculation we already have, except for that little -1 in the denominator. What gives?

Well, this is where the distinction between μ and $\overline{X}$ comes into play. The quantity we're measuring has some actual value, μ, that would be the average

[22] Hopefully, this will coincide with the "true" value of X (or at least be close enough). This will only happen if our measurement procedure has minimal systemic error (or cancelling systemic errors), and is therefore *unbiased*.

of infinitely many measurements. However, we only took finitely measurements, and the average of our actual data is $\overline{X}$, which will not necessarily equal μ. So let's go through this carefully.

Let's suppose we've taken a bunch of measurements x_1, x_2, ..., x_n. This data has an average of $\overline{X}$, and a theoretical value of μ. Now, let's calculate the uncertainty σ_X based on the definition of σ_X given in equation (20). It's easiest to calculate from the square:

$$\sigma_X^2 = \overline{(X-\mu)^2} \approx \frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}$$

The approximation here means that, on average, that will be the variation in our sample (since the left-hand side is based on the actual probability distribution, and the right-hand side is based on our *finite* set of data). Corrections to this approximation are not important for our purposes.[23]

We'll use a trick of adding and subtracting $\overline{X}$ in each square now:

$$(\sigma_X)^2 = \frac{\sum_{i=1}^{N}((x_i-\overline{X})+(\overline{X}-\mu))^2}{N}$$

FOILing out that square gives:

$$\begin{aligned}
\sigma_X^2 =& \frac{\sum_{i=1}^{N}\left[(x_i-\overline{X})^2+(\overline{X}-\mu)^2+2(x_i-\overline{X})(\overline{X}-\mu)\right]}{N} \\
=& \frac{\sum_{i=1}^{N}(x_i-\overline{X})^2+\sum_{i=1}^{N}(\overline{X}-\mu)^2+\sum_{i=1}^{N}2(x_i-\overline{X})(\overline{X}-\mu)}{N} \\
=& \sum_{i=1}^{N}\frac{(x_i-\overline{X})^2}{N}+\frac{N(\overline{X}-\mu)^2}{N}+\frac{\sum_{i=1}^{N}2(x_i-\overline{X})(\overline{X}-\mu)}{N} \\
=& \sum_{i=1}^{N}\frac{(x_i-\overline{X})^2}{N}+\sigma_{\overline{X}}^2+\frac{2(\overline{X}-\mu)\cancel{\sum_{i=1}^{N}(x_i-\overline{X})}}{N}
\end{aligned}$$

That last term is zero because the sum vanishes, by definition of $\overline{X}$:

$$\sum_{i=1}^{N}(x_i-\overline{X}) = \left(\sum_{i=1}^{N}x_i\right) - N\overline{X} = N\overline{X} - N\overline{X} = 0$$

[23] For those of you who have taken a statistics class, applying this procedure again is what gives you a t-score instead of a z-score. It tends to be more important at small sample sizes, when the variation in the variance is higher (i.e., from set of trials to set of trials, the last term of that sum varies more). However, we're oversimplifying this anyway (since we're not even using z-scores; see Appendix B.2), so we're not concerned about this subtlety here.

Now: we have this nice formula, but don't know either σ_X nor $\sigma_{\overline{X}}$. However, from the general sum formula (17) (which we don't need this formula to derive), we *do* know that equation (5) must hold.

Therefore, we can replace $\sigma_{\overline{X}}^2$ with $\frac{(\sigma_X)^2}{N}$ and finish off our calculation:

$$\sigma_X^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{X})^2 + \frac{\sigma_X^2}{N}$$

$$N\sigma_X^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2 + \sigma_X^2$$

$$N\sigma_X^2 - \sigma_X^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2$$

$$(N-1)\sigma_X^2 = \sum_{i=1}^{N}(x_i - \overline{X})^2$$

$$\sigma_X^2 = \frac{\sum_{i=1}^{N}(x_i - \overline{X})^2}{N-1}$$

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{X})^2}{N-1}}$$

And there you have it - equation (4), with the $N-1$ factor and all. From this, equation (3) is a result of (5), so now we have all of our basic formulas for uncertainty.

TL;DR: The -1 is a result of the fact that the average of our data, $\overline{X}$, is not necessarily the same as the actual true value, μ, when we said the uncertainty was our average distance from the true value. $\overline{X}$ will generically be closer to our data than μ, and the slightly larger denominator compensates this difference.

A.3 A Fuller Explanation of Error Propagation Formulas

Using the definition of uncertainty from section A.1 and some basic notions from calculus, we can get a better idea of where the error propagation formulas come from, and figure out how to make our own.

Following the notation of section 2.3, we'll denote by δX the difference in our measurement from the true value of X, and similarly for δY when we need it.

A.3.1 1 Variable Error Formulas

Let's begin with a function of one variable, $f(X)$. The actual error of our measurement of $f(X)$ (resulting from an actual error in X of δX) is given by:

$$\delta[f(X)] = f(X + \delta X) - f(X)$$

Well, that's not very helpful in general. Fortunately, we have some calculus tools on our side. Let's make the assumption that δX is small, and Taylor-expand it:

$$f(X + \delta X) \approx f(X) + f'(X)\delta X + \dots$$

Now, here's the key approximation we'll make: we'll drop the ... from the above expression. We assume that the second-order errors are negligible, and so use the first order errors. This is good enough for most purposes, although if you were really careful you might want to do something more specific.

Therefore, we can approximate the earlier formula as:

$$\delta[f(X)] \approx [f(X) + f'(X)\delta X] - f(X) = f'(X)\delta X$$

Now, we just plug into the definition for σ_X, (20), and find the result:

$$\sigma[f(X)] = \sqrt{\overline{(f'(X)\delta X)^2}} = \sqrt{(f'(X))^2\overline{(\delta X)^2}} = |f'(X)|\sigma_X$$

Thus, our fully general one-variable error formula is simply:

$$\sigma[f(X)] = |f'(X)|\sigma_X \tag{21}$$

From there, using the fact that if $f(X) = X^n$, the derivative is $f'(X) = nX^{n-1}$, you can directly get the power error propagation formula (11).

A list of error propagation formulas for a variety of functions (found by applying the above formula) can be found in Table 1.

Calculation	Error Formula
X^n	$\lvert n\rvert\lvert X\rvert^{n-1}\sigma_X$
e^X	$e^X\sigma_X$
$\ln(X)$	$\frac{\sigma_X}{\lvert X\rvert}$
$\sin(X)$	$\lvert\cos(X)\rvert\sigma_X$ *24
$\cos(X)$	$\lvert\sin(X)\rvert\sigma_X$ *24
$\sin^{-1}(X)$	$\frac{\sigma_X}{\sqrt{1-X^2}}$ *25
$\tan^{-1}(X)$	$\frac{\sigma_X}{1+X^2}$ *25

Table 1: A list of more exotic error functions.

A.3.2 2+ Variable Error Formulas

Now, let's consider functions of two variables, like addition and multiplication: $f(X, Y) = X + Y$ and so on. We're going to need a bit of multivariable calculus here, but conceptually, it's the same. We begin with a Taylor-expansion:

$$f(X + \delta X, Y + \delta Y) \approx f(X, Y) + \frac{\partial f}{\partial X}\delta X + \frac{\partial f}{\partial Y}\delta Y + \dots$$

Therefore, again dropping the higher-order terms, we similarly find the expression:

$$\delta[f(X, Y)] \approx \frac{\partial f}{\partial X}\delta X + \frac{\partial f}{\partial Y}\delta Y$$

Again, we plug into our formula (20) and (after a bit of algebra) see what we get:

$$\sigma[f(X, Y)] = \sqrt{\left(\frac{\partial f}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial f}{\partial Y}\right)^2 \sigma_Y^2 + 2\left(\frac{\partial f}{\partial X}\right)\left(\frac{\partial f}{\partial Y}\right)\overline{\delta X \cdot \delta Y}}$$

Now, we assume that X and Y are uncorrelated, so $\overline{\delta X \delta Y} = 0$. Otherwise, you need that number, too, to do your calculation properly; that number is called the *covariance* of X and Y.[26]

Under the assumption that that's zero, though, we get our formula in two variables:

$$\sigma[f(X, Y)] = \sqrt{\left(\frac{\partial f}{\partial X}\right)^2 (\sigma_X)^2 + \left(\frac{\partial f}{\partial Y}\right)^2 (\sigma_Y)^2} \tag{22}$$

If you know enough multivariable calculus to understand the partial derivatives, you should now be able to derive the addition and multiplication rules with that formula.

The derivation and result also extend straightforwardly to three or more variables: you add more partial derivatives times matching errors in quadrature, under the assumption that all errors are pairwise-uncorrelated.

A.3.3 Statistical Independence

It's worth noting what happens when the statistical independence breaks down.

[24] For trig functions, σ_X must be measured in radians.

[25] For inverse trig functions, $\sigma(\text{trig}^{-1}(X))$ will be given in radians.

[26] The generalization of single-variable uncertainties to multi-variable uncertainties generally entails a *covariance matrix*; see Appendix B.3 for more details.

For instance, consider if we were find the error in $\frac{X}{X+Y}$. Clearly, the numerator and denominator are not independent - when the numerator increases, the denominator increases too, so an error in X produces less error in this ratio than it otherwise would, because the error in the numerator and denominator will cancel.

In the context where the issue is just a complicated function of variables that were *originally* statistically independent, there are two solutions to this. The first is to try to get your equation into a form where the error can be evaluated properly, and each variable only shows up once.

For instance, the above fraction is equal to $\frac{1}{1+Y/X}$. There, if X and Y are uncorrelated, we don't have a computational issue. This is neat when you can do it; unfortunately, it's more of a trick than an easy-to-use procedure.

The other is to use the full calculus formulas. This is how to do this sort of analysis properly, if you ever find yourself in such a situation, and is - if you know calculus - often easier than trying to find a trick procedure.

(If the original quantities were statistically dependent, then you have to quantify that statistical dependence, and things get more complicated - see the discussion in Appendix B.3.)

A.3.4 Higher Order Effects

What if our Taylor series approximation breaks down? For instance, what if $f'(X) = 0$ (or is small enough that the second-order correction matters)?

A number of complicated things can happen, led by our error propagation formulas being no longer approximately correct. That aside, though, more happens than just having different formulas: random errors have biasing effects.

Let's take a simple example: suppose we want to measure X^2 when the true value, μ, is equal to 0. Now: *any* sort of X error is going to result in a positive estimate for X^2. This means that a really small random error in X doesn't make X^2 go up and down by the same amount - it always makes it go up.

This makes our random error in X into a systemic one in X^2. That's quite the messy issue!

Fortunately, it rarely comes up, but if you're at a place where $f'(X) = 0$ (in any variable X), you'll need to look very closely at your experiment to make sure these systemic errors don't matter. (Fortunately, the "errors-are-small" assumption usually means these are still very small, and you can neglect them altogether... but errors aren't always small enough that you can do that!)

Let's take a more concrete example: a warped meter stick. The meter stick can be warped up or down; let's say the amount it is warped up or down is an amount X.

It doesn't matter if X is positive or negative; either way, the meter stick is going to take a curved path between your two points rather than a straight one, and so the measured distance will always be longer than the actual distance ($\sigma[f(x)]$ will always be positive). You are going to consistently overestimate the actual distance, regardless of which direction it was warped.

B White Lies In These Notes

A number of things we have said in previous sections are not strictly true (even if they approximate the truth to a certain degree). In this section, we will rectify that by giving a more accurate presentation of that information.

The more nuanced methods presented here will not be used in our introductory physics labs (PHY121/122 and PHY133/134), but may be relevant to you in a more advanced course.

B.1 Different Uncertainty Quantifications

The first lie we tell is that all of the methods of determining uncertainty in section 2.1 are equivalent. This is, of course, not the case.

There are two reasonable ways to think about uncertainty. One is as an *upper bound* on the variation from the average. The other is the *typical* variation from the average.

Most of the qualitative rules we provide in that section follow the upper-bound rule. When we assume our dominant source of error is roundoff error, we assume that we cannot be off by (significantly) more than the maximum amount by which we round. Therefore, the amount we round can be taken as an uncertainty.

This is also the sort of uncertainty we presume when we make our comparisons. When we ask whether the uncertainty ranges of A and B overlap (and say that if they do they "agree" and if they don't they "disagree"), we are attempting to make the assertion: if they overlap, then there is a possibility that both measurements are actually giving the same number; if they do not overlap, then they cannot be giving the same number.

By contrast, the idea that uncertainty represents the "typical deviation" from the true value underlies the derivations we perform in Appendix A, where

we used the definition of uncertainty (20) that quantified that measurement.

In particular, that means that our statistically-derived uncertainties using formulas (4) and (3) are measuring this typical deviation. So, technically speaking, are all of our propagation formulas.

If we say we are using the second sort of measurement of uncertainty, then we are overestimating many of our basic errors, and are being overly stringent on our comparisons. If we say we are using the first sort of uncertainty, then our statistically-estimated uncertainties are underestimated, and our comparisons are overly strict (since we require[27] the difference to be less than the "typical distance"). In either case, we're making a dishonest leap in our quantifications.

A reasonable compromise heuristic is the following: first, make the uncertainty estimates as referred in section 2.1, except double formulas (4) and (3). We then make propagations as before. Finally, we make comparisons to a constant as before, and when we need to compare two numbers with uncertainty, we take the difference and see if it is compatible with the constant 0.

In essence, if we take the statistical uncertainty formulas as measuring the standard deviation of a normal distribution, then we are working with 95% confidence intervals, and treating our measurement devices' roundoff errors as 95% confidence intervals as well.

The uncertainty formulas are the same for twice the uncertainty as the uncertainty, so that still works, and our comparison technique ensures that at the end of the calculation we are comparing with two-sigma confidence - that is to say, we use what is often taken as standard outside of physics, the famous "$p < 0.05$" confidence result.

B.2 Degrees of Confidence

A proper accounting of comparing with uncertainty is rarely all-or-nothing. Typically, when experimental results disagree, there are varying degrees of disagreement. While this section will not give a full account of these possibilities, it may serve as an introduction to the world of "five-sigma results" versus "two-sigma hints" (etc.) that you often hear in discussions of experimental physics.

When one takes measurements of a quantity, compared to the true value, they come out with various distributions. To account for all these distributions in a detailed way is a statistical nightmare - although, depending on the circumstances of the experiment, it may need to be done.

[27] When one quantity has uncertainty, anyway. When both have uncertainty, our "proper" propagation formulas give us slightly more complicated results, since they add in quadrature.

On a practical level, however, things tend to drastically simplify if you have enough statistics. If you take an average of many measurement, then the mean of those measurements always tends to be distributed according to a fixed distribution, known as a Gaussian. Formally (with all its technical caveats), this is known as the *Central Limit Theorem*.

More specifically, define the *normal distribution* with mean μ and standard deviation σ as:

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{23}$$

Then, if we assume that a quantity is *normally distributed* with mean μ and standard deviation σ, the probability for of getting a result for that quantity in the range (a,b) (from a single measurement) is:

$$P(a,b) = \int_a^b f(x|\mu,\sigma^2)dx \tag{24}$$

Then, the Central Limit Theorem states[28] that, if you have a very large number N of samples, and an initial distribution with mean μ and uncertainty (in the sense defined in equation (20)) σ, the mean of all those measurements is distributed in a way that is approximately normal, with mean μ and standard deviation $\frac{\sigma}{\sqrt{N}}$ (and the approximation gets better as N gets larger).

So, at the end of the day, if we have a bunch of statistics, we can usually get away with assuming our results are distributed in a Gaussian way, even if they are not. If we make such an assumption, then we can much more easily make quantitative comparisons, because now we have a probability distribution to use.

Now, let's assume for simplicity that we're comparing a result with uncertainty (taken, again, in the sense of equation (20)) to a known expectation (if both quantities have uncertainty, we can compare the difference to zero, as described in the previous section). Then, we can ask how many multiples of that uncertainty our result is from the expectation - often phrased as how many sigmas.

With our Gaussian distribution, we expect a result within one σ approximately 68% of the time, two σs 95%, and three σs 99.7%. For this reason, two σs is often taken as a "strong hint" (but not more, since one in twenty experiments will find that by accident[29]).

[28] There's a bit more formalization still, but if you want to learn all the technical details, take a probability theory class.

[29] This tends to be the standard for publication in some other fields (social sciences) by default, which leads to the unfortunate practice of "p-hacking" - repeating many measurements to find one that indicates a statistically significant result, even if it does not stand up to replication and was one of these one-in-twenty chances. Some academic journals in those fields have started instituting protections against this practice, since it is bad use of statistics (whether intentional or not).

The "golden standard" used in fields like particle physics for "discovery" of new physics is typically five σ - that's taken as a known result, unless relevant systemic errors that were not noticed are later discovered. (The ways in which labs account for systemic errors to avoid this problem vary from experiment to experiment.)

Thus, one can have an experiment with a continuous range of results, based on how many σ. Informally, we might say that we have three categories instead of two: practical agreement ($< 2\sigma$), hints but not proof ($> 2\sigma$, $< 5\sigma$), and what is practically proof ($> 5\sigma$), with a significant range of fuzziness in that middle category.

B.3 Correlated Errors: Covariance Matrices

Several times, both in the main text and the appendix, we made note that we were making a simplifying assumption that our different measurements were all independent. We figured out how to work around that assumption if our intermediary calculations were correlated using calculus if our initial measurements were uncorrelated in section A.3, but here, we will figure out how to deal with correlations between variables in a nice, systemic way.

Let's suppose that we have our set of measurements $\vec{X}$, which is a column vector with N components. (Note: this is a vector in the mathematical sense, not the physics sense; it does not lie in physical space, but in some abstract mathematical space.) For instance, we might have measurements of mass m, length L, maximum angle θ, and period T of a pendulum, in which case our measurement vector could be:

$$X = \begin{bmatrix} m \\ L \\ \theta \\ T \end{bmatrix}$$

Or, it could be rearrangement; the exact order of the components isn't important (so long as you're consistent about it). Now, as in sections 2.3 and A.3, we let δX denote the difference in any particular measurement of X - that is to say, the difference in any (hypothetical) *set* of measurements performed to get X. We now define the *covariance matrix* as:

$$\Sigma = \overline{(\delta X)(\delta X)^T} \tag{25}$$

We label entries of this matrix (taking the above X as an example) by, e.g., σ_{mL}. This is called the *covariance* between m and L (or whatever entries it has).

Then, it follows from our definition of uncertainty (20) that the diagonal elements are the squares of uncertainty: $\sigma_{mm} = \sigma_m^2$, etc. Furthermore, if your

measurements are independent, this matrix is diagonal, and that is all that appears.

One can define another matrix, the *correlation matrix*, by dividing by appropriate factors of the uncertainty, with entries:

$$\text{corr}(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \tag{26}$$

In other words, we take the covariance matrix, take the diagonal elements to define σ_m and so forth, then divide each row by the respective uncertainty (i.e., divide the row corresponding to m by σ_m) *and* divide each column by the respective uncertainty.

This makes a matrix with 1s on the diagonal and unitless numbers on the off-diagonal. If all of your measurements are independent, the correlation matrix ends up being the identity matrix. (The converse - that if you have the identity matrix, your measurements are independent - is not necessarily true, but is accurate to the lowest-order approximation that we make in our propagation formulas.)

The correlation matrix is simpler and extract intuitive correlation information from, but the covariance matrix is easier for most calculations (rather like relative and absolute error).

Then, let's suppose we have a function of multiple variables $f(X)$, and want to calculate its uncertainty. In terms of the derivative of f (in all variables) at the measured value of X, ∇f (which we take as a column vector), we can calculate the uncertainty as:

$$\sigma[f(x)] = \sqrt{(\nabla f)^T \Sigma(X) (\nabla f)} \tag{27}$$

If we have multiple functions of interest $g(X)$, $h(X)$, etc.; then we can actually calculate the covariance matrix using the same formula: assemble them into a single row vector $f(X) = [g(X)\ h(X)\ \ldots]$, at which point ∇f is a matrix (with which variable's derivative you take changing in the columns and the function you take changing in the rows). Then, the above equation still holds, but gives the covariance matrix between $g(X)$, $h(X)$, etc.

If you take the simplest case, that your initial variables are uncorrelated, then this gives the same results as in A.3.

C Data Fitting: Procedures and Uncertainties

In this section, we'll explain some of the details behind how the plotting tool works, and what a "line of best fit" means, quantitatively speaking. An under-

standing of this will let you fit curves to a wider variety of circumstances than the plotting tool(s) provided for the PHY121/122 and PHY133/134 courses.

C.1 Fitting a Line Without Error Bars

Let's start with the simplest case: the data doesn't have error bars to speak of (or when we don't know our errors very well, and just want something easy and reasonable).

We say we want a "line of best fit," which presumably means we want something optimized - we want it to be the "best." But, in this context, what, exactly, is "good" - that is to say, what makes one line a "better" fit than another?

What we ultimately want is some sort of measure of *closeness to the data*, since an ideal line of best fit would pass as close to the data as possible. So let's try to measure "distance from the data" and see what we can do.

One naïve thing to try to measure total distance from a line to our points is the following:

1. Take the line and data points as given.
2. For each data point, take the distance from the line to the data point as the *closest* the line gets to the data point. (For a straight line, this will always be done with a perpendicular projection.)
3. Take the total distance as the sum of individual distances.

This sounds reasonable, except that the computations (since we're working with distances in 2 (or higher!) dimensional space), which involve square roots, are terrible. So, let's simplify things a bit, and use the same trick that we do for uncertainties themselves (as discussed in section A.1): we'll work with the sum of distances squared, which gives us what is (therefore) called a *least squares fit*.

Unfortunately, this still isn't easy, because we have to find the distance of closest approach, which is nontrivial even in the linear case (and it gets even worse for nonlinear fitting). So we'll simplify even more, and take just the *vertical* distance between the point and the line. In addition to being simpler in the linear case, this extends much more easily to a wide class of problems (as will be discussed in C.2).

C.1.1 Solving the Linear Fit

Now: let's suppose we have a bunch of measurements x_i and corresponding y_i. Let's arrange the y_i into a column vector, $\vec{y}$, with a number of entries equal to the number of data points. Let's do the same with the x_i (into a vector $\vec{x}$).

Note that we want to minimize $\sum_i (y_i - (ax_i + b))^2$.

If we let $\vec{1}$ be the column vector of all 1s (with a number of entries equal to the number of data points), then this quantity can be rewritten:

$$\sum_i (y_i - (ax_i + b))^2 = (\vec{y} - a\vec{x} - b\vec{1})^T (\vec{y} - a\vec{x} - b\vec{1})$$

We can now minimize it in a and b by taking derivatives and setting them equal to zero, which gives us that $\vec{x}^T(\vec{y} - a\vec{x} - b\vec{1})$ and $\vec{1}^T(\vec{y} - a\vec{x} - b\vec{1})$ vanish. This is now a system of equations linear in a and b, and so it can be solved by linear algebra.

This system of equations can be written in terms of the matrix $X = \begin{bmatrix} \vec{x} & \vec{1} \end{bmatrix}$ and the column vector $c = \begin{bmatrix} a \\ b \end{bmatrix}$ as:

$$X^T \vec{y} = X^T X c$$

If we suppose that $X^T X$ is invertible, then this gives us:

$$c = (X^T X)^{-1} X^T y$$

This is an equation for a and b in terms of x and y.[30]

Since that is now simply a calculation of the coefficients in terms of an explicit formula, one can, in principle, do an associated error propagation and get an associated uncertainty on a and b. The full propagation is messy, but if you neglect uncertainties in x, you can derive an uncertainty based on the uncertainty in y:

$$\delta c = (X^T X)^{-1} X^T \delta y$$
$$\Sigma_c = (X^T X)^{-1} X^T \Sigma_y X (X^T X)^{-1}$$

Assuming we know the uncertainty in y (to avoid certain complications[31]) and assuming that uncertainty is fixed over y (such that $\Sigma_y = \sigma_y^2 I$), we can get

[30] This can, in principle, be written more explicitly in terms of other quantities, but we don't do that here - the idea behind the calculation suffices for our purposes, since the Plotting Tool can do the rest for us. If you want the final formulas, Google exists.

[31] In the absence of error bars, we make the assumption that $\Sigma_y = \sigma_y^2 I$ - that is to say, that all data points have the same uncertainty and there are no correlations in our data. Furthermore, we make the assumption that the scatter in our data from the line is all a result of vertical scatter in y, at which point we can use our data to get an estimate in σ_y, just as we use the scatter in our data to get an estimate of uncertainty in any particular measurement in equation (4).[32] The details in this case get a bit more subtle, though, for the same reason as discussed in section A.2: we are estimating uncertainty based on data's difference from an expectation that is *also* based on that data, and the scatter in our data from the average based on our data will always be a bit less than the scatter in our data from the true value.

[32] Note that σ_y here is the y-scatter for any particular y for *fixed* x, not the overall y-scatter in our data. Thus, we don't naïvely apply equation (4) to get this value; we have to measure the difference of each y from the expected value for that y, which is dependent on our fit.

(after some algebra) the equation:

$$\Sigma_c = \frac{1}{N}\frac{\sigma_y^2}{\frac{1}{N}\sum_i(x_i^2) - \left(\frac{1}{N}\sum_i x_i\right)^2}\begin{bmatrix} 1 & -\frac{1}{N}\sum_i x_i \\ -\frac{1}{N}\sum_i x_i & \frac{1}{N}\sum_i(x_i^2)\end{bmatrix}$$

Note that the denominator is always positive, using the simplifying identity $(\sum_i x_i^2) - (\sum_i x_i)^2 = \sum_i \left(x_i - \left(\frac{1}{N}\sum_j x_j\right)\right)^2$. Also, assuming the x are drawn from a fixed distribution, the matrix and the denominator of the fraction are approximately constant for large N (the sums cancel the $\frac{1}{N}$ factors), so $\Sigma \simeq \frac{1}{N}$, and in particular the uncertainties go like $\frac{1}{\sqrt{N}}$ (as usual for information based on an aggregate of N data points).

The nontrivial correlation matrix here deserves a bit of thought, but it's intuitively obvious: if all of our data takes place in positive x and we increase our guessed intercept, we get an optimal fit by decreasing the slope to a corresponding degree (to go back through the data on average), so we expect a negative cross-correlation between intercept and slope if our data is generally positive, as observed. (E.g., in the simplifying (although technically invalid) case of one data point at a positive x, if we raise the intercept, we have to decrease the slope to still pass through that data point and be the best fit.)

All of this analysis can be done without including the constant b term. The only change that needs to be made is that you drop the column of 1s from the x matrix and the vector c just turns into the constant a. This is the natural result if you apply the methods of section C.2.

C.1.2 Linear Fits Without Error Bars In Context

In this course, that fit is exactly what the PHY121/122 plotting tool does: it neglects the error bars entirely, and simply draws a line of best fit. The PHY133/134 plotting tool does a more subtle fit that includes error bars, but runs this kind of fit if no error bars are provided.

Most of your default fitting tools (which won't require error bars to run a fit) will do this kind of fit. In particular, Excel's (Google Sheets', etc.) linear fitting utility (and associated built-in function LINEST, which also gives uncertainties) do this sort of fit.

All of these tools (since they entirely neglect error bars) infer Σ_y from the scatter in your data, and take into account the subtleties discussed in footnote 31.

C.2 Fitting General Linear Models

We'll here extend the methods of section C.1.1 to a broader class of fitting problems known as *linear models*. These have some number of inputs x and

some outputs y.

We want to be fitting an equation which can be written in the form $f(x) = \sum a_i f_i(x)$, where the a_i are unknown constants and the $f_i(x)$ are known functions of the input(s).

For instance, we might be fitting an oscillation of known frequency with something of the form $A\cos(\omega t) + B\sin(\omega t)$, with A and B unknown constants (and ω known, since it's not a linear coefficient!), which is linear in the unknowns A and B (even though it is not in t). Similarly, for our standard linear fit discussed earlier, we would take $f_1(x) = x$ and $f_2(x) = 1$, with $a_1 = a$ and $a_2 = b$.

We will be using vertical least squares as our quantity to optimize for a best fit, as before. We also assume, for simplicity, a single y output.

We suppose we have a set of x measurements x_i and a corresponding set of y measurements y_i. Note that the indices here refer to different corresponding *sets* of measurements; i.e., if we have multiple input variables x_1 and x_2, the x_i are sets of pairs of (x_1, x_2), not the individual elements of that pair.

We now write a matrix F (replacing what we called X in section C.1.1) such that the ith column has the jth row be $f_i(x_j)$. I.e., as you go down the matrix, you change the x-measurement; as you go across the matrix, you change the function f_i taken. We also write down the coefficients a_i as a column vector a. For a given set of a_i, the expected set of y is then Fa.

We then want to minimize the norm of y minus its expectation (squared), which means we want to minimize $(y - Fa)^T(y - Fa)$. As before, we take a derivative with respect to a (now done all-at-once as a vector), giving us the criterion $F^T(y - Fa) = 0$.

Applying linear algebra techniques gives us the expansion:

$$a = (F^T F)^{-1} F^T y \tag{28}$$

We can, as before, also calculate the uncertainty:

$$\Sigma_a = (F^T F)^{-1} F^T \Sigma_y F (F^T F)^{-1}$$

As before, this is assuming no x-uncertainties (see also the discussion of estimating Σ_y using the scatter in y in section C.1.1).

If y has multiple data points (and we want to minimize the overall norm), then F becomes more than a matrix (a tensor). If we have y^a as our different outputs, one can make a corresponding F^a (a matrix for each a), with our linear

model being $y^a = \sum_i a_i f_i^a(x)$ (a different function for each a). At this point, equation (30) turns into:

$$a = \sum_{b,c} (F^{b,T} F^b)^{-1} F^{c,T} y^c \tag{29}$$

(The uncertainty formula can, of course, be similarly extended, but we don't reproduce it here.)

C.2.1 Linearizing Nonlinear Models

As we just showed, all models which can be written in the "linear" form of $f(x) = \sum a_i f_i(x)$ (for known f_i) are fairly easy to write down a least-squares optimization for explicitly. This covers a wide class of models, including all polynomial models.

However, some models of interest are nonlinear. For instance, suppose we know $y = Ae^{-\Gamma t}$ - we might reasonably want to fit an exponential function. Unfortunately, there's no way to fit these kinds of models in an analytic, easy way in general.

For many of these models, there is a solution: you convert it to a form where it *is* a linear model. For instance, in the above case, you can take a logarithm, and show that $\ln(y) = \ln(A) - \Gamma t$. Since $\ln(A)$ is just another constant, we now have a fit of the form $\tilde{y} = \tilde{A} - \Gamma t$, and this is a linear model in the new coefficients. So we can run a fit of this quite easily!

It's worth noting that this isn't quite the same fit as we would have run to our original distribution, because we're minimizing the distance in $\ln(y)$ instead of in y. However, the hope is that we will still get a reasonable fit out of this method, which should hold so long as the transformed variable's probability distribution looks roughly like the original variable's. This is always true if the errors are "small" (in the sense we exploit in error propagation), such that we can transform the errors from y to $\ln(y)$ in a straightforward way, but need not be the case in general.

Here are a few different versions of linearizations we can use:

- An exponential *without constant*, $y = Ae^{\Gamma t}$, can be linearized by taking a logarithm of both sides.[33]
- A sine wave with *known* frequency ω but *unknown* phase ϕ can be calculated using trigonometric identities: if we have the model $y = A\sin(\omega t + \phi)$, we can replace it with $y = B\cos(\omega t) + C\sin(\omega t)$

[33] This is not the case with a nonlinear fit, $y = Ae^{\Gamma t} + B$, unfortunately - although if you have a lot of long-time data, you can typically estimate B well enough to calculate it based on that data alone, and work with $y - B$ on your remaining data, from which you calculate Γ.

- A power law *without constant*, $y = ax^\gamma$, can be linearized with a logarithm: $\ln(y) = \ln(a) + \gamma \ln(x)$ (noting that $A = \ln(a)$ is itself is a free parameter in which one can work linearly). This is sometimes phrased as that γ is the linear slope of a log-log plot.[34]

Unfortunately, not all models can be easily linearized, so depending on what you need, you may need to use some clever tricks - or just bite the bullet and numerically fit a nonlinear model.

C.3 Fitting With Vertical Uncertainties

Now, using those same methods, we can add in uncertainties in y into our fit fairly easily. The key thing to understand is that we are no longer optimizing least-squares. Instead, we will optimize something similar, but we will normalize the differences by the respective differences.

First, assume for simplicity that the covariance Σ_y is diagonal, so that there are no correlations, but data points can have different uncertainties σ_i. Now, instead of wanting to optimize $\sum_i (y_i - f(x_i))^2$, we instead optimize $\sum_i \left(\frac{y_i - f(x_i)}{\sigma_i}\right)^2$. If we add in correlations and again write $y_i - f(x_i)$ as a column vector δy, we want to minimize $\delta y^T \Sigma_y^{-1} \delta y$. Expanding this as before slightly extends our formula to:

$$a = (F^T \Sigma_y^{-1} F)^{-1} F^T \Sigma_y^{-1} y \tag{30}$$

Similarly, the uncertainty equation:

$$\Sigma_a = (F^T \Sigma_y^{-1} F)^{-1} F^T \Sigma_y^{-1} F (F^T \Sigma_y^{-1} F)^{-1}$$

Messy, but ultimately just a bunch of linear algebra (and thus pretty easy to do on a computer, if you really want to!)

C.4 Horizontal Error Bars

A reasonable way to account for horizontal error bars that preserves the computational simplicity of a vertical least-squares fit is the following:

- Figure out a probability distribution you want to use in the horizontal direction (say, a Gaussian). That is to say, if your measurement has a value of x_0 and an uncertainty of σ_x, figure out the probability that the actual value is some other x, and write $P(x|x_0, \sigma_x)$ (the probability of the actual value being x given that the measured value is x_0 with an uncertainty of σ_x).

[34] This is one reason why power laws are often taken as phenomenological fits to quantities which operate on a wide range of scales, such as astronomical relations: they are "linear" in the space of exponents.

- Treat your measurement of x_0 as actually an infinite number of data points: one for every possible x, with weight $P(x|\ldots)$. This means you now have a smooth number of data points, but they all only count as a fraction of a point. (This is the more analytical way to do it; if you want a computationally viable way, discretize this probability distribution - it won't make too much of a difference.)

- Linearly fit, as before, with this data-point distribution (finite or infinite).

The math is conceptually fancy (especially if you don't discretize), but in principle is still entirely doable with linear algebra, and therefore, as before, is easy to write into a program.

C.5 Nonlinear Fitting of Models

One might wonder what happens when one lifts the assumption of vertical least squares and goes back to perpendicular least squares, or when one wants to fit nonlinear (and non-linearizable[35]) models.

Unfortunately, this isn't generally possible to do exactly; therefore, we leave it to numerical optimization routines. If you need to do something fancy like this, find a package that will do it for you, or code it - don't do it by hand.

One optimization routine you might try, if you want to run this in Excel or Google Sheets, is the "Solver" tool. It's more than is available in either program by default, but with a little googling, you should be able to figure out how it works.

Python also has a few algorithms to do this, in scipy. scipy.optimize.curve_fit runs a vertical least squares on a general model, which can be linear or nonlinear. scipy.odr runs an orthogonal weighted least-squares ("orthogonal distance regression").

One feature all of these routines have in common is that they optimize based on an initial "guess." Hence, you have to be able to come up with some reasonable conjecture of the parameters to begin with, and then these algorithms find the best fit that is "close" to what you guessed. (This is a feature of numerical optimization more than the particulars of curve fitting.)

C.6 PHY133/134 Plotting Tool

Now that we've explained all these fitting algorithms, let's discuss what our PHY133/134 tool does.

[35] See section C.2.1 for what we mean by "linearizable"

If no error bars are input, it fits with a least-squares regression, as discussed in section C.1.2.

If only x or y error bars are input, it runs a weighted least-squares.[36] The weights are based on the input standard deviations.

Finally, if both x and y error bars are input, we use scipy.odr to compute the weighted orthogonal distance regression.

As mentioned in section C.5, in all cases of nonlinear fits, a "guess" is first required; you may wonder where the tool gets such a guess from. The tool determines the "guess" parameters by running a least-squares regression *first*, then using those as a starting point for the nonlinear optimization.

If you want to explore further (or replicate the procedure), we recommend playing around with the aforementioned libraries in scipy. Have fun!

[36] We technically use scipy.odr to do so rather than scipy.optimize.curve_fit, but we adjust the weights so that effectively only one direction is considered.