FYS2150 Guide - Experimental Physics

Gabriel Sigurd Cabrera

January 21, 2018

Contents

Introduction

Before taking an experimental physics course, physics undergraduated are trained to think in terms of absolutes; with some exceptions, they are exposed to noiseless information that is perfectly recorded and without error. It makes sense to begin a physics degree this way, since one must understand the mathematics and equations needed to understand our world to be able to work in a laboratory, but it is also important to eventually realize that data is very seldom perfectly clear and accurate, and that it is very important to be able to filter out noise from data and understand potential sources of error by using statistical methods.

This pamphlet is intended to be used both as a learning tool and quick reference guide when working in a lab and analyzing data in an experimental physics course; it contains basic explanations of statistical concepts, the equations behind them, as well as images meant to illustrate their usage in practice.

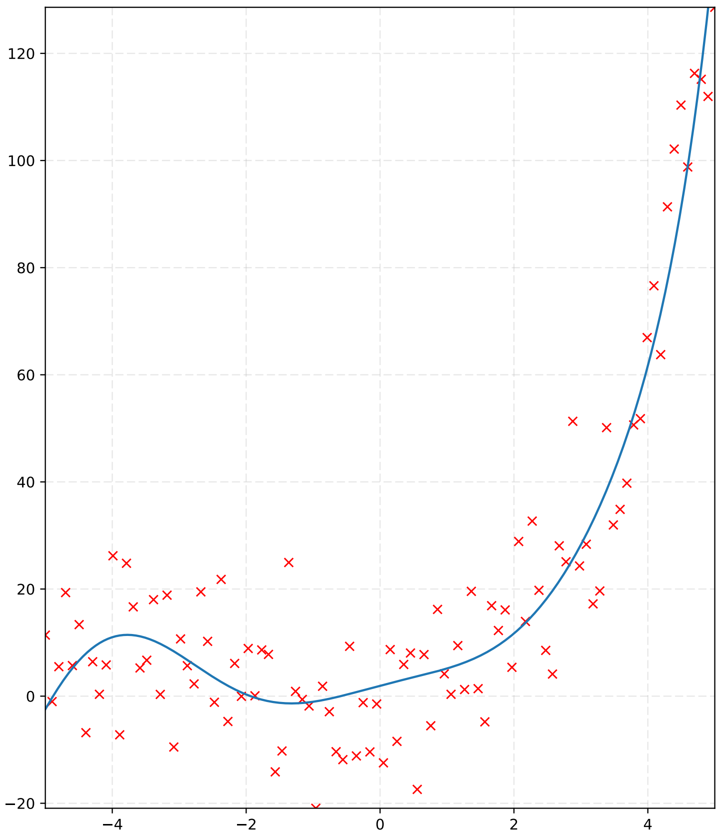


Figure 1: A least-squares approximation of a dataset

1 Principles of Physical Experimentation

There are a specific set of steps that should be taken when performing an experiment – one should always try to follow them as closely as possible so as to minimize inaccuracies when coming to a conclusion:

Identifying the Degrees of Freedom In a physical system, it is nearly always the case that the variables being dealt with can vary in several different ways (or directions, in the case of mechanical systems). These variations are called *degrees of freedom* – they are more formally defined as

> A number of independently variable factors affecting the range of states in which a system may exist, in particular any of the directions in which independent motion can occur.[1]

A concrete example of a system with one degree of freedom is the pendulum of a grandfather clock; as we all know, the pendulum will only swing in one direction without twisting, turning, or bouncing (since it is rigid). Since it can only move in one direction, it has one degree of freedom.

An example of a system with two degrees of freedom would be a rigid pendulum held by a hook in the ceiling – such a pendulum can move both in the x and y directions.

It is very important that one has systematically taken into account *all* possible degrees of freedom when dealing with an experiment, as neglecting a single one can make the results of an experiment nonsensical and incorrect (and very confusing!).

Build a Model Now that all degrees of freedom are taken into account, it is time to attempt to either create a model of the system (either analytically or numerically) to make it possible to compare the future experimental data with a mathematical equivalent with the same initial conditions. This is important both for the sake of verifying the validity of the experiment and physics model, but also to make sure that one has an understanding of the physics behind the experiment.

Create a Hypothesis Now that a model has been created, we can then choose the initial conditions that we will later test for and plug them into our model. This process is how a hypothesis is obtained, since the extent of one's understanding or predictive ability is, by definition, limited to the accuracy of the model; the results of the experiment will be used to either support or contradict this hypothesis.

Test the Hypothesis Now that we have the equipment all set up, and our hypothesis all prepared, it is time to perform the experiment itself. Once this process is complete and all the data is recorded, be certain that absolutely everything that was used (constants, equipment and sources of error) are recorded alongside the data.

Filter out Errors Finally, before making a conclusion, it is absolutely vital that the recorded data is cleaned up and that all sources of error are taken into account and adjusted for. How to deal with error will be explained in §3 (p.9)

[1] https://en.oxforddictionaries.com/definition/us/degree_of_freedom

2 The Normal Distribution

When dealing with a set of data, it is not uncommon for the majority of data points to mostly tend towards an average value (called its *mean*) with a minority of points being outliers. This makes intuitive sense in many cases, like when measuring human height: we expect the majority of people to be within a certain height range, with fewer and fewer exceptions as we digress from the mean. This concept is in fact quantifiable, and is defined in the *central limit theorem*:

> The sum of a number of independent and identically distributed random variables with finite variances will tend to a normal distribution as the number of variables grows.[2]

Normal distributions in one-dimension can be easily visualized using a *Gaussian curve*, as seen in Figure 2

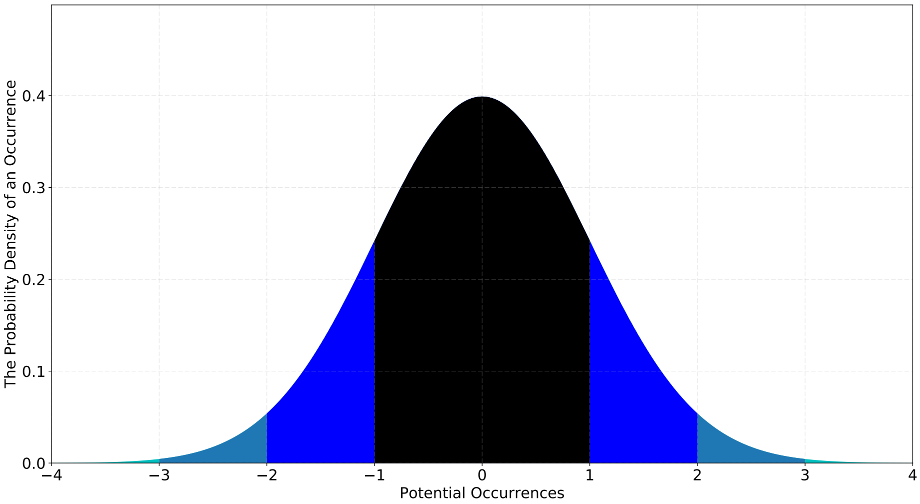


Figure 2: A 1D Gaussian distribution with $\mu = 0$ and $\sigma = 1$

2.1 The Gaussian Function

To plot a Gaussian curve, we use the Gaussian function shown in Equation 1

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1}$$

Where $g(x)$ is the Gaussian function, μ is its mean, and σ is its standard deviation. The latter two variables will be defined in §2.2 and §2.3.

[2] https://en.wikipedia.org/wiki/Stable_distribution#A_generalized_central_limit_theorem

Usage There are a couple of things one should understand before using the Gaussian function to find the probability of an occurrence.

Firstly, the area under a Gaussian curve is always exactly equal to one, corresponding to the 100% likelihood that a generated value from a normal distribution will always be in the range $(-\infty, \infty)$. We see that this is true in Equation 2

$$\int_{-\infty}^{\infty} g(x)dx = 1; \quad \forall\ \sigma, \mu \in \mathbb{R} \tag{2}$$

Secondly, it is *not possible* to find the probability of a single point occurrence; we see in Figure 2 (p.4) that the y-axis does not denote the probability of an occurrence, but rather the *probabilistic density* of an occurrence. In other words, we must always integrate over a *range* of occurrences to find a probability P, otherwise it will simply be zero. For example, let's say we wish to find the likelihood of getting a value $a \in \mathbb{R}$; the *incorrect* process would be attempting to find the probability of a occurring as in Equation 3

$$\int_{a}^{a} g(x)dx = 0; \quad \forall\ \sigma, \mu \in \mathbb{R} \tag{3}$$

We see that attempting to choose a single value leads nowhere; we must instead choose a range $[a - \delta, a + \delta]$ and integrate over that as in Equation 4

$$P = \int_{a-\delta}^{a+\delta} g(x)dx \tag{4}$$

We can also choose to find the probability of getting a value within a range $[a, b]$ as shown in Equation 5

$$P = \int_{a}^{b} g(x)dx \tag{5}$$

2.2 The Sample Mean

For a set of discrete points Let's say we have a dataset containing N points $\{x_1, x_2, \cdots, x_{\text{N}}\}$; whether or not our set is normally distributed, we can calculate its *sample mean* (also known as its *average*) μ by using Equation 6

$$\mu = \tfrac{1}{\text{N}} \sum_{i=1}^{\text{N}} x_i \tag{6}$$

We may also choose to use alternative notation such as $< x >$ or $\overline{x}$ to denote the sample mean of our dataset.

In a Gaussian Curve If we assume that our dataset is normally distributed, our curve's maximum will be located at the x-coordinate corresponding to our sample mean; we can clearly see that this is the case in Figure 2 (p.4) , as our mean is located at $\mu = 0$.

In other words, the values most likely to be generated from a normal distribution are nearest to its sample mean.

For a continuous probability density function Let $f(x)$ be a probability density function such that $\int_{-\infty}^{\infty} f(x)dx \triangleq 1$; we can find the sample mean of $< x >$ based on the given probabilities by using Equation 7

$$< x > \triangleq \int_{-\infty}^{\infty} xf(x)dx \tag{7}$$

2.3 The Standard Deviation and Standard Error

Assuming that we once again have a dataset with N points $\{x_1, x_2, \cdots, x_{\text{N}}\}$, the square of its standard deviation σ is defined as the average of the squared difference between all its points and the sample mean μ, as shown in Equation 8

$$\sigma^2 = < (x-\mu)^2 > = \tfrac{1}{N}\sum_{i=1}^{\text{N}} (x_i - \mu)^2 \tag{8}$$

We can use the standard deviation to understand the probability of an occurrence when generating a random value from a normally distributed set; we classify the ranges of possible values generated from a normal distribution *deviating* from the mean as $n\sigma$ multiples of the standard deviation. Figure 4 (p.7) shows the extents of $1\sigma, 2\sigma$, and 3σ standard deviations on a Gaussian curve, as well as the probabilities of randomly generating a value within their ranges.

The probabilities are also shown in Figure 3

σ-Factor	Probability of a value falling within this σ-factor
1	68.27 %
2	95.45 %
3	99.73 %

Figure 3: The probability of generating a value within $1\sigma, 2\sigma$, and 3σ respectively in a normal distribution

2.4 The Convolution of Multiple Normal Distributions

Let's say that we now wish to combine N separate, normally distributed datasets $\{x_1, x_2, \cdots, x_{\text{N}}\}$ (each with their own respective σ_i and $\overline{x}_i$, $i = 1, 2, ..., N$) so as to create a new total Gaussian function $g(y)$ for the total dataset $y = \sum_{i=1}^{\text{N}} x_i$. To accomplish this, we must use equation 9

$$g(y) = \frac{1}{\sqrt{2\pi\sigma^2_{\text{TOT}}}} e^{-\frac{1}{2}\frac{\left(y - \sum_{i=1}^{\text{N}} \overline{x}_i\right)^2}{\sigma^2_{\text{TOT}}}} \quad \text{where} \quad \sigma^2_{\text{TOT}} = \sum_{i=1}^{\text{N}} \sigma_i^2 \tag{9}$$

Now, let $f(x_1, x_2, ..., x_{\text{N}})$ be the total probability density function of all our datasets such that $\int_{-\infty}^{\infty} f(x_1, x_2, ..., x_{\text{N}}) dx_1 dx_2 \cdots dx_{\text{N}} \triangleq 1$. We can define the sample mean of these multiple datasets by using Equation 10

$$< x_1 x_2 \cdots x_{\text{N}} > = \int_{-\infty}^{\infty} (x_1 x_2 \cdots x_{\text{N}})\, f(x_1, x_2, ..., x_{\text{N}})\ dx_1 dx_2 \cdots dx_{\text{N}} \tag{10}$$

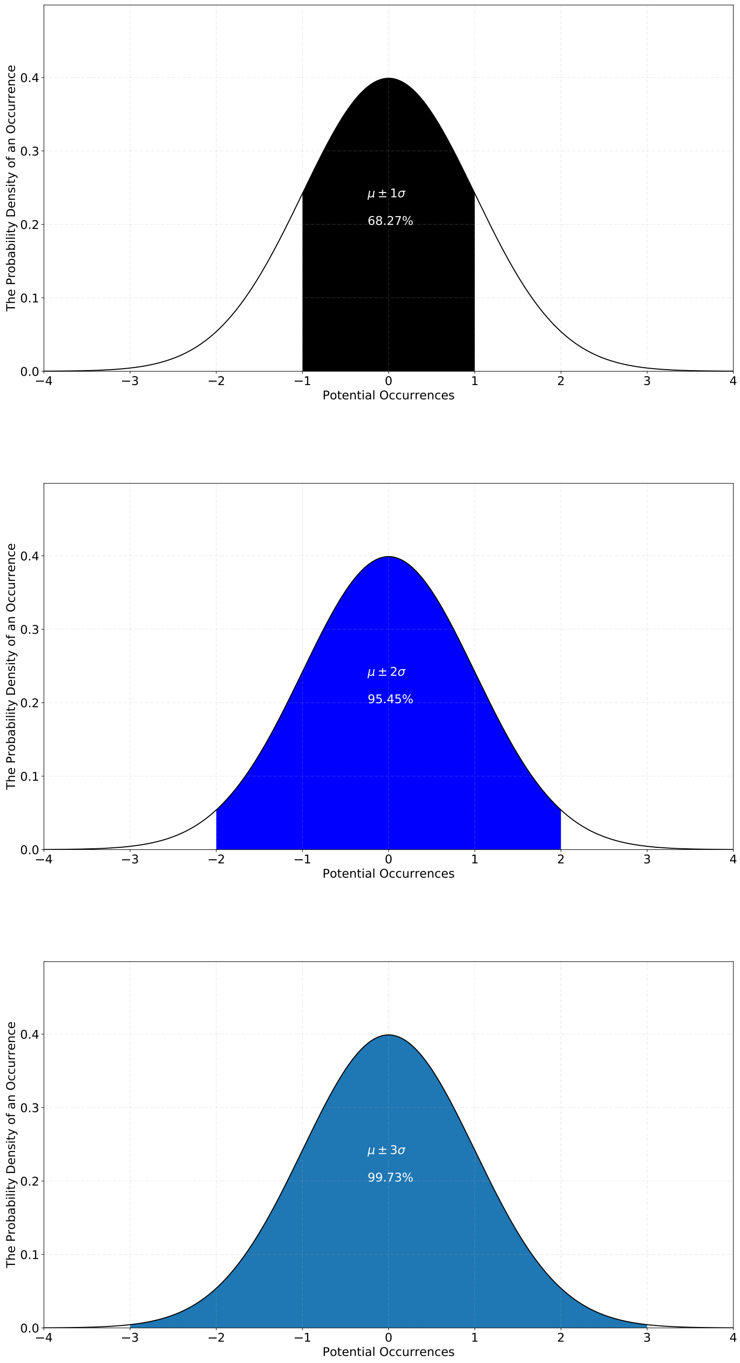


Figure 4: The probability of generating a value within $1\sigma, 2\sigma$, and 3σ respectively in a normal distribution with $\mu = 0$ and $\sigma = 1$

2.5 Variance, Covariance, and Correlation

The **variance** of a dataset measures how far a set of numbers are spread out from their average value[3]. The variance of a dataset $\{x_1, x_2, \cdots, x_{\mathrm{N}}\}$ is defined in Equation 11

$$\mathrm{var}\,(x) \triangleq < (x - \overline{x})^2 > \tag{11}$$

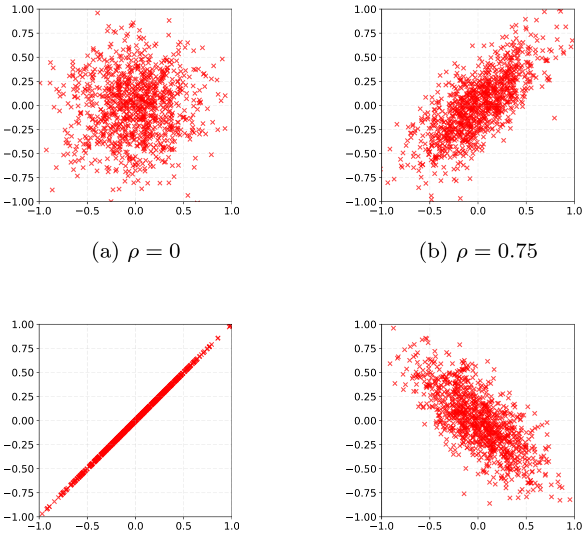
The **covariance** provides a measure of the strength of the correlation between two or more sets of random variates[4]. The covariance of two datasets $\{x_1, x_2, \cdots, x_{\mathrm{N}}\}$ and $\{y_1, y_2, \cdots, y_{\mathrm{N}}\}$ is defined in Equation 12

$$\mathrm{cov}\,(x, y) \triangleq < (x - \overline{x})(y - \overline{y}) > \tag{12}$$

Correlation is a statistical technique that can show whether and how strongly pairs of variables are related[5]. The correlation ρ between two datasets $\{x_1, x_2, \cdots, x_{\mathrm{N}}\}$ and $\{y_1, y_2, \cdots, y_{\mathrm{N}}\}$ is defined in Equation 13

$$\rho = \frac{\mathrm{cov}\,(x, y)}{\sqrt{\mathrm{var}\,(x)\,\mathrm{var}\,(y)}} = \frac{< (x - \overline{x})(y - \overline{y}) >}{\sqrt{< (x - \overline{x})^2 >< (x - \overline{y})^2 >}} \tag{13}$$

Figure 5 shows how two such datasets would appear when plotted relative to each other for cases where the data is not correlated, 75% correlated, 100% correlated, and 75% negatively correlated, respectively; note that a 100% correlation implies that $y = x$.



(a) $\rho = 0$

(b) $\rho = 0.75$

(c) $\rho = 1$

(d) $\rho = -0.75$

Figure 5: A variety of differently correlating datasets

[3] https://en.wikipedia.org/wiki/Variance

[4] http://mathworld.wolfram.com/Covariance.html

[5] https://www.surveysystem.com/correlation.htm

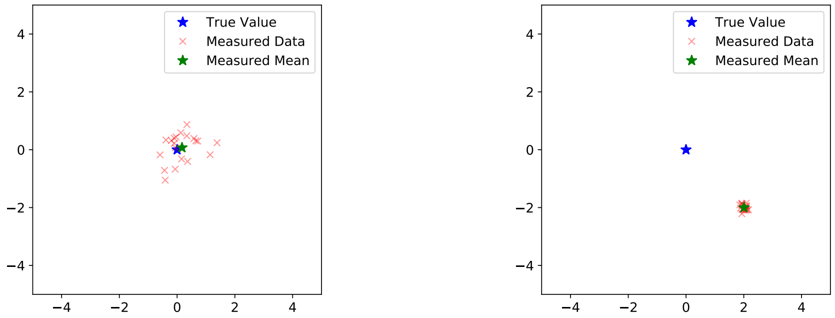
3 Error

As mentioned in the introduction, physics undergraduates are normally not accustomed to dealing with unreliable measurements in their first year or so. This is because the numbers they are given in exercises tend to be absolute and exact, since the focus at that time is less about how those numbers were found, and rather about understanding the mathematical fundamentals needed to have a general understanding of the biggest subjects within physics.

Now that we are in an experimental physics course, we must approach the subject of error in a more formal manner, since experimentation with real data and equipment will always lead to some slight measurement error. In addition to this, it is important to note that there exists both *avoidable* and *unavoidable* error (the unavoidable error is referred to as **uncertainty**, and will be discussed in §4), and within avoidable error there exists two categories – **systematic** and **random** error. Keep in mind that although uncertainty is unavoidable, it is possible to make it worse than necessary through incorrect lab procedures!

3.1 Precision vs. Accuracy

At first glance, these terms appear to be synonymous, but this is a severe misconception! Within the realm of physics, they have completely different meanings. Firstly, a dataset with a high **precision** is a dataset whose mean value is very close to the true value we are trying to reach. On the other hand, a dataset a with high **accuracy** is one whose standard deviation is very small (all the data points are very close to each other). These are not mutually exclusive, and a dataset can be neither accurate nor precise, one or the other, or both. Figure 6 is a visual representation of two datasets, the first being precise but inaccurate since its points are centered around the true value yet relatively distant from each other, and the other being unprecise and accurate since its points are not centered around the true value and yet are very compact and close together.



(a) A precise but inaccurate dataset

(b) An accurate but unprecise dataset

Figure 6: The difference between precision and accuracy

3.2 Systematic Error vs. Random Error

There are plenty of things that can go wrong in an experimental setting; bad equipment, human error, an incorrect setup and a variety of other occurences can lead to trouble when not identified early on enough in the experiment. To better deal with such possibilites, we can divide our errors into two categories.

The first type of error is called **systematic error**, and is characterized by the fact that it is both *reproducible* (meaning that an experiment can be repeated with the same systematic error yielding the same incorrect results) and very difficult to identify or notice, since its source may originate from anything from badly calibrated equipment, to bad laboratory conditions such as high humidity or temperature.

The second type of error is called **random error**, and is different to systematic error in the sense that it cannot be reproduced in a future repetition of the same experiment, as it is by its very nature due to human error or any number or random one-time occurrences. The positive thing about random error is that it can usually be identified by repeating an experiment, and is easy to account for in the future.

4 Uncertainty

Due to the nature of measurement, there is always an inherent amount of error that will manifest itself in recorded data; through advances in technology, these error may be reduced over time, but one can never truly take a measurement that is noiseless and exact. The real danger here is not the error itself however, since we may still extract very useful data from slightly incorrect data – rather, it is vital to take care to manage the error such that all calculations being made based on our experimental data take this **uncertainty** into account.

Think about it this way, if we are certain that we have a glass containing 1L of water, and we have another glass which may contain anything between 0.25 and 0.75 liters of water (in other words, 0.5 ± 0.25L of water, 0.25 being the uncertainty) and proceed to combine them, we will then be unable to know exactly how much liquid is now in the container due to the uncertainty from the second glass. In the end, we must conclude that we have 1.5 ± 0.25L of water.

If we then restart this exercise, and instead have one glass with 1 ± 0.1L and another with 0.5 ± 0.2L and proceed to combine them, we must then take the sum of the uncertainties and add it to the sum of the median values such that we have a total of 1.5 ± 0.3L of water. This makes intuitive sense when viewed from the perspective that there exists an (unmeasureable) true exact amount of water in each glass – since we could potentially have true values of 0.9L and 0.3L due to the aforementioned uncertainties, it is possible for our true total amount of water to be 1.2L, which is 1.5L − 0.3L, or the lowest possible value of 1.5 ± 0.3L.

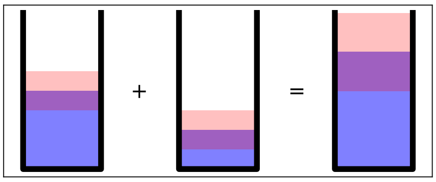


Figure 7: A visualization of the addition of uncertainties

4.1 Absolute vs. Relative Uncertainty

Up to this point, the type of uncertainty we've been discussing is known as **absolute** uncertainty – this is the most intuitive of the two, and is simply due to the inaccuracies of measuring devices. Firstly, the absolute uncertainty is always represented in the same units as that of the measured value (for example liters in our previous example) and secondly, it is always added together when dealing with addition or subtraction (this is explained later in §4.2).

Generally speaking, the absolute uncertainty is normally represented by Δ; as an example, if we have a mass m, then it may be represented as $m \pm \Delta m$ when including uncertainty.

Due to the nature of other arithmetic operations, such as multiplication, we must also take into account another value known as the **relative** uncertainty. Although it may appear as such, this is not a particularly difficult concept to grasp, as it simply represents the percentage of uncertainty relative to the median value.

For relative uncertainty, we may use the symbols δ or ε; δ may be found by using Equation 14

$$\delta x = \frac{\Delta x}{x} \tag{14}$$

It is useful to look at the relative uncertainty as a middleman used to perform certain arithmetic operations, since it tends to be best to represent the uncertainty as absolute in our final results.

4.2 The Arithmetic of Uncertainty

Now that we've explained the addition of uncertainty in Figure 7, as well as understood the differences between absolute and relative uncertainties in §4.1 it is important to understand that the subtraction of two uncertain values *still involves the addition of absolute uncertainty!* This is a vital detail to keep in mind, and does make sense when you think about it for a moment, since it is not possible for our experimental results to become more and more certain as we introduce uncertainty (in fact, that would even lead to negative uncertainty, which makes absolutely no sense whatsoever). Figure 8 (p.12) is a compilation of all arithmetic operations with uncertainty included.

For a general solution to determine the total error Δy in a function $y = y(x_1, x_2, ..., x_{\text{N}})$, where $x_1, x_2, ..., x_{\text{N}}$ are a set of variables of the form $x_i \pm \Delta x_i$, we can use Equations 15 and 16

$$\Delta y_{x_i} = \left(\frac{\partial y}{\partial x_i}\right) \Delta x_i,\ \ i = 1, 2, ..., N \tag{15}$$

$$\Delta y = \sqrt{\left(\Delta y_{x_1}\right)^2 + \left(\Delta y_{x_2}\right)^2 + \cdots + \left(\Delta y_{x_{\text{N}}}\right)^2} \tag{16}$$

It is completely acceptable to solve Equations 15 and 16 using numerical methods, as long as a small enough step is used when making the calculation, just be aware that it won't give an exact answer in most cases, though that may be unnecessary in some cases.

It is also not a bad idea to keep in mind that it is possible that the systematic error and random error in a measurement is separate, meaning that it is necessary to combine the two by using the methods shown in this section to find the total error.

Operation	Generalized Formula
Addition	$(x \pm \Delta x) + (y \pm \Delta y) = (x + y) \pm \sqrt{(\Delta x)^2 + (\Delta y)^2}$
Subtraction	$(x \pm \Delta x) - (y \pm \Delta y) = (x - y) \pm \sqrt{(\Delta x)^2 + (\Delta y)^2}$
Multiplication	$(x \pm \Delta x)\,(y \pm \Delta y) = xy \pm xy\sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}$
Division	$\frac{x \pm \Delta x}{y \pm \Delta y} = \frac{x}{y} \pm \frac{x}{y}\sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2}$
Exponentiation	$x^n \pm \Delta x = x^n \pm x^n n \frac{\Delta x}{x}$
Natural Logarithms	$\ln(x \pm \Delta x) = \ln(x) \pm \frac{\Delta x}{x}$
Natural Exponentiation	$\exp(x \pm \Delta x) = \exp(x) \pm \exp(x)\,\Delta x$

Figure 8: Generalized arithmetic operations with uncertainty

5 The Mean and Standard Deviation

Although we have already discussed the mean, variance and standard deviations of datasets in §2 (p.4), it is a good idea to have all our equations listed together due to the relationships they hold to each other as well as for ease of access.

The Sample Mean The sample mean is, in short, the arithmetic mean or average of a dataset. It can be found using Equation 17

$$\overline{x} = \frac{1}{N}\sum_{i=1}^{\mathrm{N}} x_i \tag{17}$$

The Sample Variance The sample variance is, in essence, the quantification of how much a dataset varies relative to its mean. It may be determined with Equation 18

$$s^2 = \frac{1}{N}\sum_{i=1}^{\mathrm{N}} (x_i - \overline{x})^2 \tag{18}$$

The Unbiased Square of the Standard Deviation If we wish to have an idea of how to model a Gaussian curve based on a normally-distributed dataset, we can use Equation 19 to find its standard deviation

$$\sigma^2 \approx \frac{1}{N-1}\sum_{i=1}^{\mathrm{N}}(x_i - \overline{x})^2 \tag{19}$$

The Unbiased Standard Error in the Mean We may at some point be interested in determining how accurate our mean is, relative to the standard deviation. Logically speaking, it makes sense that the more points we are given, the more representative its mean becomes. Take for a example, a dataset containing 3 points; although it is possible to find a sample mean from these points, it doesn't have much weight statistically speaking, due to the lack of data. On the other hand, a dataset containing 100,000 points would have a mean that is far more meaningful, due to the abundance of data. This is a value that can be quantified, and is known as the **unbiased standard error**; it is defined in Equation 20

$$\sigma_m = \frac{1}{\sqrt{N}}\sigma \tag{20}$$

6 The Method of Least Squares

If we are given a dataset of points that appear to trend, either linearly, quadratically or in any way matching some kind of function, we may wish to find such a function so as to match the points as accurately as possible. To accomplish this, we must delve into some linear algebra and understand how to use QR-Factorization to find our least-squares solution.

6.1 QR Factorization

Theorem Let A be an $m \times n$ matrix with linearly independent columns. It is then possible to write the following:

$$A = QR \tag{21}$$

Where Q is an $m \times n$ orthogonal matrix and R is an $n \times n$ upper triangular matrix with all diagonal elements $r_{i,i} > 0$

Definition Let A be an $m \times n$ matrix with $m > n$, and let $\vec{x}, \vec{y} \in \mathbb{R}^n$. Then, a **least squares** solution of the equation $A\vec{x} = \vec{y}$ is:

$$\hat{x} \in \mathbb{R}^n \text{ such that } \|\vec{y} - A\hat{x}\| \leq \|\vec{y} - A\vec{x}\|\,, \quad \forall\, \vec{x} \tag{22}$$

It may seem like a daunting task to find $\hat{x}$, however we are fortunate in that the following theorem will help us find our desired solution:

Theorem The set of *least squares solutions* of $A\vec{x} = \vec{y}$ is equal to the set of solutions of:

$$\left(A^T A\right)\vec{x} = A^T\vec{y} \tag{23}$$

The above is known as the **normal equation**

Note that $\left(A^T A\right)$ is an $n \times n$ matrix, and that $A^T\vec{y}$ is a vector in $\mathbb{R}^n$

Note Assume that A has a *QR factorization* $A = QR$; to help us solve the equation given in the above theorem, we have that:

$$A^T A\vec{x} = A^T\vec{y} \iff R^T R\vec{x} = R^T Q^T\vec{y} \iff R\vec{x} = Q^T\vec{y} \tag{24}$$

Since R is a triangular matrix, this simplifies our problem significantly.

Terms In the above, we have that:

- A is the **design matrix**
- $\vec{y}$ is the **observation vector**
- $\vec{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}$ is the **parameter vector**

6.2 Usage

Here are a few generalized situations - note that these are meant to be read in order, and that each $\alpha_i \in \mathbb{R}$ represents a constant coefficient, and each β_i represents an arbitrary function of some variable:

Linear Approximation Suppose we wish to approximate a data set (of n points) with a linear function of the form $y(x) = \alpha_0 + \alpha_1 x$; in essence, we wish for the following to be as small as possible:

$$S = \sum_{i=1}^{n} \left(y_i - (\alpha_0 + \alpha_1 x_i)\right)^2 \tag{25}$$

We cannot control our y_i values, so we must find a α_0 and α_1 that minimizes our S.

We can reinterpret the above as a formula of two vectors, such that:

$$D = \text{dist}\left(\vec{y}, \hat{y}\right) = \|\vec{y} - \hat{y}\| \tag{26}$$

Where we wish to minimize D; we also have that $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and $\hat{y} = \begin{bmatrix} \alpha_0 + \alpha_1 x_1 \\ \alpha_0 + \alpha_1 x_2 \\ \vdots \\ \alpha_0 + \alpha_1 x_n \end{bmatrix}$.

This can in turn be rewritten as:

$$A \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \vec{y} \tag{27}$$

Where $A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix}$

It is now simply a matter of solving for $\vec{y}$ using the *method of least squares*

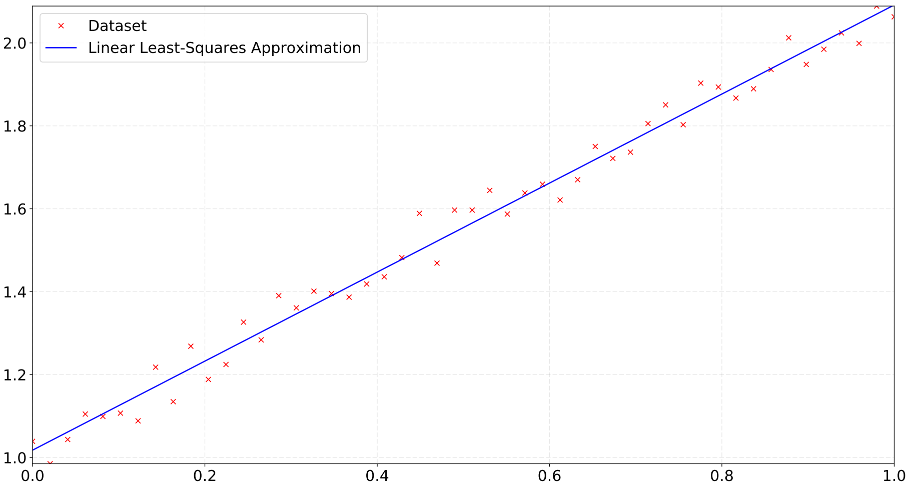


Figure 9: A linear least-squares approximation of a dataset

Polynomial Approximation Now, suppose we wish to approximate another data set (of n points) with a m^{th} degree polynomial of the form $y(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_m x^m$; we can generalize the previous example as follows (we still wish to minimize S):

$$S = \sum_{i=1}^{n} \left(y_i - \left(\alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \cdots + \alpha_m x_i^m \right) \right)^2 \tag{28}$$

We must now find a set $\{\alpha_0, \alpha_1, ..., \alpha_m\}$ that minimizes our S.

We can once again interpret the above as a formula of two vectors where we wish to minimize a value D:

$$D = \operatorname{dist}(\vec{y}, \hat{y}) = \|\vec{y} - \hat{y}\| \tag{29}$$

Where:

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \hat{y} = \begin{bmatrix} \alpha_0 + \alpha_1 x_1 + \alpha_2 x_1^2 + \cdots + \alpha_m x_1^m \\ \alpha_0 + \alpha_1 x_2 + \alpha_2 x_2^2 + \cdots + \alpha_m x_2^m \\ \vdots \\ \alpha_0 + \alpha_1 x_n + \alpha_2 x_n^2 + \cdots + \alpha_m x_n^m \end{bmatrix} \tag{30}$$

This can once again be rewritten as:

$$A \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \vec{y} \tag{31}$$

Where:

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix} \tag{32}$$

Finally, we can solve for $\vec{y}$ using the *method of least squares*

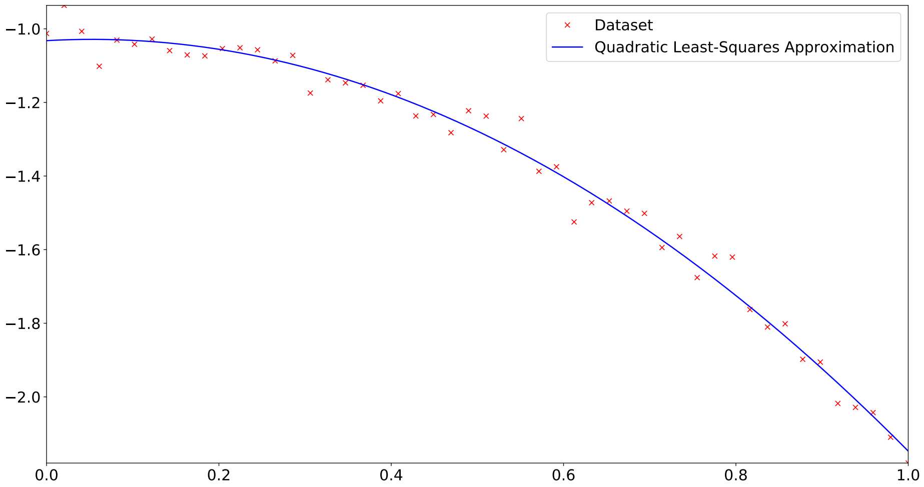


Figure 10: A quadratic least-squares approximation of a dataset

General Approximation Let us now get a generalize this completely - suppose we once again have a data set consisting of n points, but this time we wish to approximate it with $y(x) = \alpha_0 + \alpha_1\beta_1(x) + \alpha_2\beta_2(x) + \cdots + \alpha_m\beta_m(x)$. As previously, we wish to minimize the following S:

$$S = \sum_{i=1}^{n} \left(y_i - (\alpha_0 + \alpha_1\beta_1(x_i) + \alpha_2\beta_2(x_i) + \cdots + \alpha_m\beta_m(x_i))\right)^2 \tag{33}$$

We can rewrite this as a function D:

$$D = \text{dist}\,(\vec{y}, \hat{y}) = \|\vec{y} - \hat{y}\| \tag{34}$$

Where:

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \hat{y} = \begin{bmatrix} \alpha_0 + \alpha_1\beta_1(x_1) + \alpha_2\beta_2(x_1) + \cdots + \alpha_m\beta_m(x_1) \\ \alpha_0 + \alpha_1\beta_1(x_2) + \alpha_2\beta_2(x_2) + \cdots + \alpha_m\beta_m(x_2) \\ \vdots \\ \alpha_0 + \alpha_1\beta_1(x_n) + \alpha_2\beta_2(x_n) + \cdots + \alpha_m\beta_m(x_n) \end{bmatrix} \tag{35}$$

Once again, this can be written as:

$$A \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix} = \vec{y} \tag{36}$$

Where:

$$A = \begin{bmatrix} 1 & \beta_1(x_1) & \beta_2(x_1) & \cdots & \beta_m(x_1) \\ 1 & \beta_1(x_2) & \beta_2(x_2) & \cdots & \beta_m(x_2) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \beta_1(x_n) & \beta_2(x_n) & \cdots & \beta_m(x_n) \end{bmatrix} \tag{37}$$

Finally, we can solve for $\vec{y}$ using the *method of least squares*

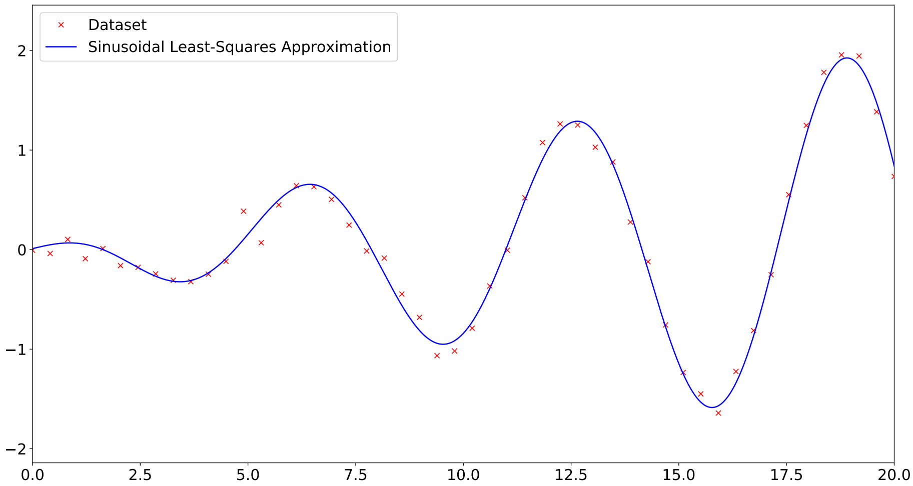


Figure 11: A linear-sinusoidal least-squares approximation of a dataset

6.3 A Python Algorithm

Below is a useful algorithm I developed to find a least-squares solution for a dataset, where the user can input any number of functions to attempt and match the dataset.

```python
import sympy as sp
import numpy as np

def least_squares_functions(points, functions, return_coefficients= False,
include_one= True):

    '''Uses the method of least squares to match a set of points

    <points> should be an array of dimensions (n, 2), where n is the number of
    points we are trying to match.  <functions> should be a string or array of
    strings that can be numpy evaluated mathematically in terms of a variable
    <x> - do not include coefficients. <return_coefficients> determines whether
    or not the list of coefficients is returned - if false, simply returns the
    function'''

    points= np.array(points)

    if len(points.shape) == 2 and points.shape[0] == 2 and points.shape[1] != 2:
        points= points.transpose()
    point_error_msg= 'Invalid format for argument <points>'
    assert len(points.shape) == 2 and points.shape[1] == 2, point_error_msg

    if isinstance(functions, str):
        if include_one == True:
            functions= ['1', functions]
        else:
            functions= [functions]
    functions_error_msg= 'Invalid format for argument <functions>'
    assert isinstance(functions, (list, tuple)), functions_error_msg
    if isinstance(functions, tuple):
        functions= list(functions)
    if include_one == True and '1' not in functions:
        functions= ['1'] + functions

    math_error_msg= 'Non-evaluable function in list <functions>'

    X, y= np.tile(points[:,0], reps= (len(functions),1)).transpose(), points[:,1]
    A= np.zeros((len(X), len(functions)))
    for n,f in enumerate(functions):
        x= X[:,n]
        try:
            A[:,n]= eval(f)
        except ValueError:
            assert False, math_error_msg

    A= np.asmatrix(A)
    AT= A.transpose()
    Y= AT*np.asmatrix(y).transpose()
    coefficients= sp.Matrix(np.concatenate((AT*A, Y), axis= -1)).rref()[0][:,-1]
    coefficients= np.array(coefficients)

    if return_coefficients == False:
        eval_string= ''
        for n,(c,f) in enumerate(zip(coefficients, functions)):
            eval_string+= '(%f*(%s))'%(c,f)
            if n < len(functions) - 1:
                eval_string+= '+'
        def f(x):
            return eval(eval_string)
        return f
    return coefficients
```