I. ESTIMATION AND PROPAGATION OF ERRORS

A. Direct and indirect measurements

Most of the measurements which we encounter (at least in physics) are *indirect measurements*. This means that usually we know mathematical relationship between a physical quantity y that we want to measure and some basic physical quantities x_1, x_2, .. , which we can measure in practice. The number of such quantities is not that large. As an experimental physicist, I mostly deal with *direct measurements* of electrical current and voltage. Other basic quantities that usually you measure directly are length and time. Anyway, if you know an analytical relationship between y and x_1, x_2, .. , that is

$$y = f(x_1, x_2, ..), \tag{1}$$

you can hope to find the numerical value of y from the measured values of x_1, x_2, .. , and estimate the error of your result. Correspondingly, we have to answer two main questions

(1) What are the errors δx_1, δx_2, .. of our direct measurements?

(2) How these errors affect the error δy of our final results, that is how the above errors *propagate* into our final result?

For example, in our first experiment we try to measure the acceleration of a free-fallen body. It is related to the length s of the path and time t to travel along this path (we assume the body started to move from the rest)

$$g = \frac{2s}{t^2}. \tag{2}$$

B. Instrumental errors

Usually direct measurements correspond to just reading a scale. This could be the scale of a meter stick, or the digital display of a voltmeter. Scales are graduated in discrete calibrated values, e.g. marks on the meter stick scale. In this case the measured value is "rounded off" to the closest scale mark, and it is reasonable to choose the error of your measurement as plus/minus half of the distance between the marks. If your meter stick is graduated in millimeters, it is reasonable to take your error as ± 0.5 mm.

Similarly, if you read a digital scale of a voltmeter and it says 10.46 V, it is reasonable to take the error of your measurement as ± 0.005 V. Note that in this case the rounding off is done by the device (voltmeter) itself. The above are examples of what I call the *instrumental errors*.

C. Systematic errors

Besides the instrumental errors, there could be other errors inherent to your measuring device. For example, your meter stick could be calibrated incorrectly, so that the distance between the marks differs from 1 mm. In this case, the reading can give you lower or higher value than it should be providing that the difference between your reading and the actual length is larger than the instrumental error of your device. Indeed, let's say your meter stick is supposed to be graduated in millimeters but the actual distance between any two neighboring marks is 0.9 mm. You are measuring a length of about 9 mm. In this case you round off your reading to a mark marked as 10 mm, and obtain the result 10 ± 0.5 mm. That is you significantly overestimate your result. This is an example of the *systematic errors*.

The systematic errors can be also inherent to you indirect measurement procedure. For example, in your model for the free-fallen body given by Eq. (2) you do not take into account the friction force on the body due to the air. Obviously, if the effect of this force is significant, you will underestimate the measured value of g.

It is hard to find and estimate this type of errors. The only way to handle them is to analyze carefully your procedure and model, check calibration of your instruments etc. One important point is that since systematic error will always push your result in the same direction, that is you will either *underestimate* or *overestimate* your result comparing to the *true* value, it is easy to distinguish it from the third type of errors that we consider next.

D. Random errors

Suppose you made a single measurement of the time needed for a body to fall from known height s to the ground using a stop watch, and your reading is 3.61 s. If you are sure that the systematic errors are negligible, you conclude that you result is 3.61 ± 0.005 s. Are you happy? Just out of curiosity, you decided to repeat your measurement under the same

conditions, and obtained 3.47 s. Certainly, your error is not 0.005 s but significantly larger! Why didn't you get the same result as after the first measurement? Of cause, you suspect that the conditions for two measurements are not exactly the same. For example, you could trigger your stop watch at slightly different times with respect to the start of body's motion. This is an example of the *random errors*.

The random errors arise because the conditions of your repeated measurements are always changing. These changes will affect your result, so the result of your repeated measurements will fluctuate. You hope that it will fluctuate around the true value. How do we find the best estimate for the true value? It's reasonable to expect that if you repeat you measurement many times, the values you obtain will tend to group around the true value. So, logically the best estimate will be the arithmetic mean of the results of your repeated measurements, and the spread of these numbers around the average value will characterize your random error. This logical thinking has its theoretical basis on the statistics. Let's try to describe our results from the statistical point of view.

1. Statistical distribution

In the previous section we saw that it is a very good idea to repeat your measurement several times and check for consistency. Obviously, this helps you to reveal if random errors are important. If the results of all your repeated measurements are the same within the instrumental error, obviously the random errors are too small and not important. However, suppose you got an array of N different values. Going back to the previous section's example, suppose we made $N = 10$ measurements of time t and obtained results t_i, $i = 1, 2, .. , N$.

$$3.61 \quad 3.47 \quad 3.92 \quad 3.55 \quad 3.33 \quad 3.27 \quad 3.65 \quad 3.27 \quad 3.56 \quad 3.42 \tag{3}$$

First, we have to find a way to represent these date. Note that the actual time spans continuous values. Let's divide the time axis t into equally separated values t_k, such that $\Delta t = (t_{k+1} - t_k) >> \delta t$, where δt is the instrumental error. Now, let's count number of times when a single measurement gaves us a value t_i such that $t_k \leq t_i < t_{k+1}$. We'll call this number n_k. Obviously we have $\sum_k n_k = N$, and the arithmetic mean of our N measurements is given by

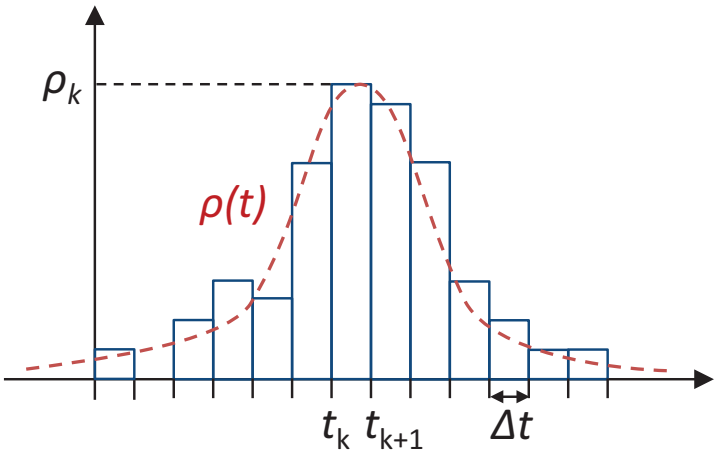

FIG. 1: Graphical representation of the results of repeated measurements and their continuous limit.

$$\langle t\rangle_N = \frac{1}{N}\sum_{i=1}^{N} t_i \approx \frac{1}{N}\sum_k n_k t_k. \tag{4}$$

Note that we choose Δt sufficiently small so that different t_i from the same interval $[t_k, t_{k+1}]$ do not differ significantly. It is convenient to write the above expression in the form which do not explicitly depend on N. Introducing $\rho_k = n_k/N$, that is the *fraction* of measurements which give result $t \in [t_k, t_{k+1}]$, we can write

$$\langle t\rangle_N \approx \sum_k \rho_k t_k, \tag{5}$$

where, obviously, each $\rho_k \leq 1$ and

$$\sum_k \rho_k = 1. \tag{6}$$

It is convenient to represent this result graphically in the form of a *histogram*, see Fig. 1. Usually we will see that maximal ρ_k group around certain value of $k = \tilde{k}$. So, logically the true value of t should belong to the interval $[t_{\tilde{k}}, t_{\tilde{k}+1}]$. To increase accuracy of our true-value estimation, we can try to decease the interval length Δt. However, note that if we just decrease Δt keeping N unchanged, many intervals will not contain any measured numbers t_i at all. So, as a rule with decreasing Δt we should also simultaneously increase the number

of measurements N.

Logically, since the quantity t assumes continuous values, we can consider the limit $\Delta t \rightarrow 0$ (and simultaneously $N \rightarrow \infty$). It is natural to expect that in this *continuous limit* we will obtain a continuous line, something like shown by a red dashed line in Fig. 1. It is easy to see that this line is described by a continuous function $\rho(t)$ such that

$$\lim_{t_{k+1} \rightarrow t_k} \rho_k = \rho(t_k) \Delta t. \tag{7}$$

The function $\rho(t)$ of continuous variable t is called the *probability distribution function.* Summing over all k and using Eq. (6), we obtain

$$1 = \lim_{t_{k+1} \rightarrow t_k} \sum_k \rho_k = \int_{-\infty}^{\infty} \rho(t) dt. \tag{8}$$

Using the same argument, we obtain for the arithmetic mean value

$$\langle t \rangle = \lim_{t_{k+1} \rightarrow t_k} \sum_k \rho_k t_k = \int_{-\infty}^{\infty} \rho(t) t dt. \tag{9}$$

Another important quantity is the integral of $\rho(t)$ from t_1 and t_2. From our definition, it gives the limit of fraction of measurements that give result t between t_1 and t_2, or the *probability* $P_{[t_1,t_2]}$ to obtain t in the interval from t_1 to t_2 as a result of a single measurement, that is

$$P_{[t_1,t_2]} = \int_{t_1}^{t_2} \rho(t) dt. \tag{10}$$

2. Mean, Standard Deviation, and Standard Deviation of the Mean

The probability distribution function $\rho(x)$ gives you as completely information as possible for the measured quantity x subject to random errors. For most of the probability distribution functions encountered in nature, the mean value $\langle x \rangle$ given by Eq. (9) corresponds to the value which you would measure with the highest probability in a single measurement. This value is your *true* value of the physical quantity x. However, in reality the probability distribution function is not known *a priory.* You can find it approximately by making N independent measurements and plotting a histogram, like in Fig. 1. Obviously, you can

not make an infinite number of measurements N needed to reconstruct the continuous limit $\rho(x)$!

The best you can do is to make a sufficiently large number of measurements N and calculate the arithmetic mean

$$\langle x\rangle_N = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{11}$$

Then, the so called *principle of maximum likelihood*, which can be proven in statistics, tells us that this value is the best estimate of $\langle x\rangle$ from Eq. (9). *Thus, the arithmetic mean of the results of N independent measurement is our best estimate x_{best} for the true value of x.*

Another important quantity that we want to know is how much, on average, the result of a single measurement will deviate from $\langle x\rangle$. In case of N independent measurements, it is logical to estimate this as an arithmetic mean of deviations $(x_i - \langle x\rangle_N)$, $i = 1, 2, .. , N$. However, note that we expect that different x_i will fall on both lower-value and higher-value sides of $\langle x\rangle_N$, so different terms in the sum over $(x_i - \langle x\rangle_N)$ will compensate each other. Instead, we should sum over $|x_i - \langle x\rangle_N|$ or, as is usually done, over $(x_i - \langle x\rangle_N)^2$, that is

$$\sigma_N^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \langle x\rangle_N)^2 . \tag{12}$$

The root square of this quantity is called the *standard deviation*, $SD = \sqrt{\sigma_N^2}$. Again, using the continuous limit, we obtain

$$\sigma_x^2 = \int_{-\infty}^{\infty} \rho(x)\,(x - \langle x\rangle)^2\, dx. \tag{13}$$

This quantity is known as the *variance* of your distribution. Again, the *principle of maximum likelihood* tells you that if you did N independent measurements, σ_N^2 is you best estimate for the variance σ_x^2.

Note that $SD = \sqrt{\sigma_N^2}$ tells you *how much, on average, the result of a single measurement will deviate from your best estimate of the true value of x.* It is not the error of your best estimate itself. The later is given by what is called the standard deviation of the mean, and is given by

$$SDM = \frac{SD}{\sqrt{N}} = \frac{\sqrt{\sigma_N^2}}{\sqrt{N}}. \tag{14}$$

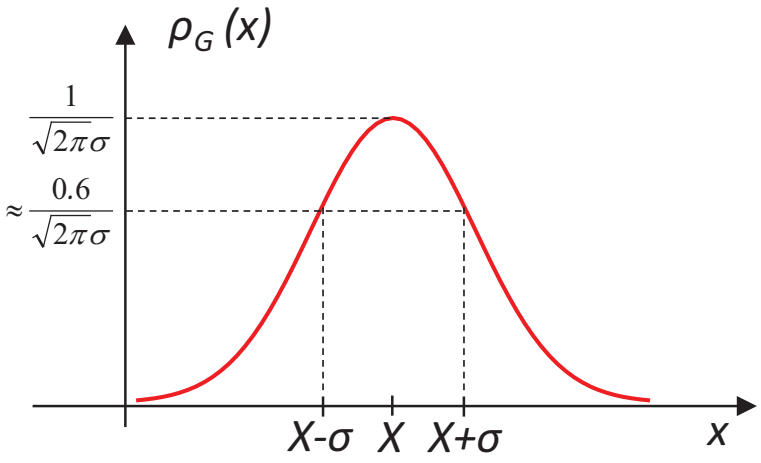

FIG. 2: Graphical representation of the results of repeated measurements and their continuous limit.

This result will be derived below after we discuss the propagation of errors.

Question 1

Note that if $N \to \infty$, then $\sqrt{\sigma_N^2} \to \sqrt{\sigma_x^2}$, which is finite, therefore $SDM \to 0$. Does this make sense to you?

3. Normal distribution

As I mentioned earlier, usually you don't know the exact probability distribution function $\rho(x)$ describing your random errors. However, in many cases the form of this distribution function is known. The distribution most frequently encountered in nature is the so called *normal* (or *Gauss*) distribution given by

$$\rho_G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-X)^2}{2\sigma^2}}. \tag{15}$$

The plot of normal distribution is shown in Fig. 2. The function $\rho_G(x)$ depends on two parameters, X and σ, the physical meaning of which you can figure out by working out the following two exercises.

Exercise 1

Find the mean value of quantity x obeying the random normal distribution

$$\langle x \rangle = \int_{-\infty}^{\infty} \rho_G(x) x dx. \tag{16}$$

Exercise 2

Find the variance and, therefore, the standard deviation of quantity x obeying the random normal distribution

$$\sigma_x^2 = \int_{-\infty}^{\infty} \rho_G(x) \left(x - \langle x \rangle\right)^2 dx. \tag{17}$$

Hint

You can use the following (so called Gauss) integral

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}. \tag{18}$$

If you do it right, you should find out that the parameters X and σ coincide with the mean and the square root of variance, respectively, that is $X = \langle x \rangle$ and $\sqrt{\sigma_x^2} = \sigma$. There are a few other helpful results you should remember about the normal distribution. It is a good idea to check it yourself using Matlab or Mathematica.

(1) The probability to find the result of a single measurement of quantity x obeying the normal distribution within interval $[X - \sigma, X + \sigma]$ is about 68%

$$\int_{X-\sigma}^{X+\sigma} \rho_G(x) x dx = 0.68. \tag{19}$$

In other words, σ is you *68% confidence limit*.

(2) The probability to find the result of a single measurement of quantity x obeying the normal distribution within interval $[X - 2\sigma, X + 2\sigma]$ is about 95%

$$\int_{X-2\sigma}^{X+2\sigma} \rho_G(x) x dx = 0.95. \tag{20}$$

E. Propagation of errors

1. *Arbitrary case*

Suppose we know now how to deal with estimation of true values and their uncertainties for each of the directly measured quantities $x_1, x_2, ..$. How do we estimate the true value and the error of our estimation for the quantity y given by Eq. (1)? It seems reasonable to assume that the best estimate for y is given by

$$y^{best} = f(x_1^{best}, x_2^{best}, ..). \tag{21}$$

Let's find uncertainty δy if we know uncertainties δx_1, δx_2,.. . For clarity, let's start with a much simpler case

$$y = x_1 + x_2 = (x_1^{best} \pm \delta x_1) + (x_2^{best} \pm \delta x_2). \tag{22}$$

Obviously, we can write for the *maximum* and *minimum* values of y within the accuracy of our measurements

$$y^{max} = (x_1^{best} + x_2^{best}) + (\delta x_1 + \delta x_2), \quad y^{min} = (x_1^{best} + x_2^{best}) - (\delta x_1 + \delta x_2), \tag{23}$$

therefore $y = y^{best} + \delta y$, where $\delta y = \delta x_1 + \delta x_2$. Thus, to obtain the error in y we just need to add error in x_i *linearly.*

Assuming that errors δx_i are small ($\delta x_i << |x_i|$) and using differentials, it is straightforward to use a similar argument and obtain in the general case of $y = f(x_1, x_2, ..)$ that

$$\delta y = \left|\frac{\partial f}{\partial x_1}\right| \delta x_1 + \left|\frac{\partial f}{\partial x_2}\right| \delta x_2 + .. \quad , \tag{24}$$

where $\partial f / \partial x_i$ means the *partial* derivative with respect to x_i.

Exercise 3

Please carry out this argument and obtain the Eq. (24).

2. *Uncorrelated quantities*

The above argument seems to be solid and fair. However, is there possibility that in certain cases we would overestimate the error in y if we just add the uncertainties δx_i? The

answer is "yes"! Let's go back to the simplest case $y = x_1 + x_2$. When estimating the maximum/minimum values of y, we assumed that both x_1 and x_2 take maximum/minimum values *simultaneously* during the same single measurement. However, if this two quantities are completely independent, or *uncorrelated*, this would rarely happen. Instead, when e.g. x_1 takes the maximum values $x_1^{best} + \delta x_1$ because of random changes in the environment, the quantity x_2 has a large probability to take value close to x_2^{true}. Similarly, if x_2 takes the maximum value during the given measurement, there is a big chance that x_1 will be close to x_1^{true}. So, in the case of independent quantities x_1 and x_2 the maximum vales of y is likely to be significantly less then in Eq. (23).

How do we estimate uncertainty δy in this case? Consider again the simplest case $y = x_1 + x_2$. Let's assume that both variables are completely uncorrelated and their random errors obey the normal distribution. As we saw above, the probabilities P_{x_1} and P_{x_2} to measure the given values of x_1 and x_2 are proportional to the exponential factor in Eq. (15), that is

$$P_{x_1} \propto e^{-\frac{(x_1 - X_1)^2}{2\sigma_{x_1}^2}}, \quad P_{x_2} \propto e^{-\frac{(x_2 - X_2)^2}{2\sigma_{x_2}^2}}. \tag{25}$$

Since the variables x_1 and x_2 are completely uncorrelated, the probability to measure the given values simulteniously, that is probability P_y to measure given value of $y = x_1 + x_2$, is given by the product of the above probabilities, that is

$$P_y \propto e^{-\frac{(x_1 - X_1)^2}{2\sigma_{x_1}^2}} e^{-\frac{(x_2 - X_2)^2}{2\sigma_{x_2}^2}}. \tag{26}$$

Since X_1 and X_2 are just fixed numbers, we can consider the new variables $\tilde{x}_1 = x_1 - X_1$ and $\tilde{x}_2 = x_2 - X_2$, and rewrite the above equation as

$$P_y \propto e^{-\frac{(\tilde{x}_1 + \tilde{x}_2)^2}{2(\sigma_{x_1}^2 + \sigma_{x_2}^2)}} e^{-\frac{(\sigma_{x_2}\tilde{x}_1 - \sigma_{x_1}\tilde{x}_2)^2}{2\sigma_{x_1}^2\sigma_{x_2}^2(\sigma_{x_1}^2 + \sigma_{x_2}^2)}}. \tag{27}$$

Note that for a given value of y, the factor in the second exponent can assume any value because x_1 and x_2 are completely independent. In other words, the given value of y can be realized by infinite number of combinations x_1 and x_2. Thus, the factor in the second exponent can run through a continuous set of values, so lets call this factor z^2, where $-\infty < z < \infty$. Integrating the above equation over z and using Eq. (18), we obtain

$$P_y \propto e^{-\frac{(\tilde{x}_1+\tilde{x}_2)^2}{2(\sigma^2_{x_1}+\sigma^2_{x_2})}} = e^{-\frac{(y-Y)^2}{2\sigma^2_y}}, \tag{28}$$

where $Y = X_1 + X_2$ and $\sigma^2_y = \sigma^2_{x_1} + \sigma^2_{x_2}$. This result means that the random error in y is distributed normally around $X_1 + X_2 \equiv x_1^{true} + x_2^{true}$ with 68% confindence limit given by

$$\sigma_y = \sqrt{\sigma^2_{x_1} + \sigma^2_{x_2}}. \tag{29}$$

It is straightforward to use a similar argument for the general case $y = f(x_1, x_2, ..)$ and show that

$$\sigma_y = \sqrt{\left(\frac{\partial f}{\partial x_1}\sigma_{x_1}\right)^2 + \left(\frac{\partial f}{\partial x_2}\sigma_{x_2}\right)^2 + ...} \tag{30}$$

Thus, for completely independent quantities subject to normally distributed random errors we have

$$\delta y = \sqrt{\left(\frac{\partial f}{\partial x_1}\delta x_1\right)^2 + \left(\frac{\partial f}{\partial x_2}\delta x_2\right)^2 + ..} \quad , \tag{31}$$

that is we add errors in *quadratures*, compare with Eq. (24).

It is easy to see that $(\delta y)_{quadrature} \leq (\delta y)_{linear}$. Adding errors linearly is fair but it can overestimate your error in case when the variables are *uncorrelated*. If you are not sure whether your variables are correlated or not, you can always use Eq. (24). If your are reasonably sure that variables are uncorrelated, add the errors in quadratures according to Eq. (31). Note that while above is proven for normally distributed error, it is usually always true for any errors as long as they are completely uncorrelated.

3. Standard Deviation of the Mean

Finally, let's prove Eq. (14) given earlier. It is easy to prove it using the addition of errors in quadratures. Suppose we performed N independent measurements of a quantity, for example a time interval t between two events, and obtained numbers t_1, t_2, .., t_N. As discussed earlier, our best estimate for the true value is the arithmetic mean

$$\langle t \rangle = \frac{t_1 + t_2 + .. + t_N}{N}. \tag{32}$$

How do we estimate error in $\langle t \rangle$? It seems reasonable to do another set of N measurements, determine $\langle t \rangle$ and compare with the previous result. It is reasonable to expect that if these two values are different, this is due to random errors. So, as discussed in Section D the best way to find the best estimate for $\langle t \rangle$ and its error is to repeat the above procedure many times and determine the mean value of $\langle t \rangle$ and its Standard Deviation, that is the *Standard Deviation of the Mean*. Let's formally treat t_i in Eq. (32) as independent variables. Then, the mean value $\langle t \rangle$ is a function of these variables according to Eq. (32). Using addition in quadratures, Eq. (30), we obtain

$$\sigma_{\langle t \rangle} = \sqrt{\left(\frac{\partial \langle t \rangle}{\partial t_1}\sigma_{t_1}\right)^2 + \left(\frac{\partial \langle t \rangle}{\partial t_2}\sigma_{t_2}\right)^2 + .. + \left(\frac{\partial \langle t \rangle}{\partial t_N}\sigma_{t_N}\right)^2}. \tag{33}$$

Obviously $\sigma_{t_1}^2 = \sigma_{t_2}^2 = .. = \sigma_{t_N}^2 = \sigma_t^2$, therefore we immediately obtain

$$\sigma_{\langle t \rangle} = \sqrt{N\frac{\sigma_t^2}{N^2}} = \frac{\sqrt{\sigma_t^2}}{\sqrt{N}}. \tag{34}$$

From the principle of maximum likelihood, our best estimate for this quantity after N independent measurement is given by Eq. (14).

F. Summary

To summarize I provide that main equations discussed in this handout.

(1) *Estimation of random errors*

The best estimate for the true value of quantity x that you can obtain after N independent measurements is given by the mean

$$x^{best} = \langle x \rangle = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{35}$$

The best estimate for error in this quantity is given by the Standard Deviation of the Mean

$$\sigma_{\langle x \rangle} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}(x_i - \langle x \rangle)^2}. \tag{36}$$

So, in the case of random errors, your final result should be represented in the form $x^{best} \pm \sigma_{\langle x \rangle}$.

The best estimate for how much, on average, the result of a single measurement will deviated from x^{best} is given by the Standard Deviation

$$\sigma_N = \sqrt{\sum_{i=1}^{N} (x_i - \langle x \rangle)^2}. \tag{37}$$

In particular, if your random error is normally distributed, σ_N gives you the best estimate for the 68% confidence limit.

(2) *Propagation of errors*

For the quantity $y = f(x_1, x_2, ..)$ depending of several other quantities, the *upper limit* of estimated error is given by

$$\delta y = \left|\frac{\partial f}{\partial x_1}\right| \delta x_1 + \left|\frac{\partial f}{\partial x_2}\right| \delta x_2 + .. \quad . \tag{38}$$

If quantities x_1, x_2, .. are *completely uncorrelated*, the error is given by

$$\delta y = \sqrt{\left(\frac{\partial f}{\partial x_1}\delta x_1\right)^2 + \left(\frac{\partial f}{\partial x_2}\delta x_2\right)^2 + ...} \tag{39}$$

Note that the same equation can be used to add different errors for the same measured quantity. For example, if the quantity x that you measure suffers from both instrumental and random errors, $(\delta x)_{instrument}$ and $(\delta x)_{random}$, and these two errors are completely uncorrelated (which most probably would be the case), then you add them in quadrature to obtain the total error, that is

$$(\delta x)_{total} = \sqrt{(\delta x)^2_{instrument} + (\delta x)^2_{random}}. \tag{40}$$