Physics Laboratory
(Vp141/Vp241)

Student Handbook

Introduction to Measurement Data Analysis

This handbook is based on the materials provided by the Department of Physics, Shanghai Jiaotong University. rev 1.2

Edited by: Qin Tian, Zeng Ming, Cao Jianjun, Han Xugen, Feng Yaming, Mateusz Krzyzosiak

Contents

1 Measurement and Measurement Uncertainties

Because of limitations of measurement devices, imperfect measurement procedures and randomness of environmental conditions, as well as human factors related to the experimenter himself, no measurement can ever be perfect. Its result may therefore only be treated as an estimate of what we call the exact value of a physical quantity. The experiment may both overestimate and underestimate the value of the physical quantity, and it is crucial to provide a measure of the *error*, or better *uncertainty*, that is carried by an experimental result. The two common measures are

$$\begin{aligned} \text{absolute measurement error} &= \text{measurement result} - \text{exact value} \\ \text{relative error} &= \frac{\text{absolute measurement error}}{\text{exact value}} \end{aligned}$$

With respect to their origin, measurement uncertainties/errors may be classified as *systematic* or *random*.

1.1 Random Errors

Random errors refer to the situation, when under the same conditions, multiple measurements of a physical quantity yield different results. Both the magnitude and the sign of these errors are not possible to be predicted and have a random nature. Their origin is related to statistical fluctuations in the conditions of the experiment, including the measurement devices used and the experimenter himself. All these factors influence the outcome of the experiment each time it is performed and contribute to the overall random measurement error (random uncertainty).

Although it is not possible to predict random errors for a single experiment, multiple realizations of the experiment allow us to comment on some features that may be effectively analyzed within the theory of probability and statistics.

1.2 Systematic Errors

Systematic errors may be caused by a number of factors. These may be identified as

a) Approximations implied by theoretical models and formulas. For example, the formula for the period of oscillations of a simple pendulum $T = 2\pi\sqrt{l/g}$ is approximate, and was derived under the assumption that the amplitude of oscillations approaches zero, which cannot be realized.

b) Imperfections of measurement devices. For example, inaccuracy of the scale of a thermometer.

c) Environmental conditions, such as temperature and humidity, may also affect the precision of measurement devices. All devices are usually designed to work efficiently in a certain range of environmental conditions.

d) Human and psychological factors that may bias results of measurements based on human action (*e.g.* manual operation of a stopwatch).

A characteristic feature of systematic errors is their permanence and regularity. Their impact on the measurements cannot be diminished by increasing the number of measurements. Therefore it is extremely important to identify and eliminate any systematic errors. They are often the main contribution to uncertainty of measurements.

2 Estimation of Measurement Uncertainties

Every experimental result should not only give an estimate of the value of the physical quantity that it is designed to measure, but also provide information about the uncertainty of the result. Giving the result only, without any information about the uncertainty, cannot be considered as a scientifically accepted outcome of any experiment. Measurement uncertainties are commonly classified into one of two categories A and B with the corresponding uncertainty magnitudes denoted as Δ_A and Δ_B, respectively.

2.1 Type-*A* (Random) Uncertainties

2.1.1 Random Error Distribution and Standard Deviation

Unpredictability is a distinctive feature of random errors. However, with a large number of realizations of a measurement, some statistical properties. In experiments in physics, the distribution of measured values x of a physical quantity X usually follows the Gauss (or normal) distribution described by the probability density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the most probable value (in this case also the average value) of the physical quantity X, and σ is the so-called standard deviation quantifying to what extent are the results of the measurement scattered around μ (see figure 1).

The interpretation of the probability density function is given by the formula

$$\Pr\left(x_1 < X < x_2\right) \stackrel{\text{def}}{=} \int_{x_1}^{x_2} f(x)\, dx$$

which defines the probability that the result x of a measurement lies in the interval (x_1, x_2). In particular, for the normal probability distribution

$$\Pr\left(\mu - \sigma < X < \mu + \sigma\right) = 0.683,$$

that is a single measurement of a physical quantity X, distributed according to the Gauss distribution, yields a value from the interval $(\mu - \sigma, \mu + \sigma)$ with probability 0.683. The interval is often referred to as 0.683 confidence interval.

Of course the probability that the measurement yields a value lying in a broader interval centred around the average is larger. In particular

$$\Pr\left(\mu - 2\sigma < X < \mu + 2\sigma\right) = 0.954, \qquad \Pr\left(\mu - 3\sigma < X < \mu + 3\sigma\right) = 0.997.$$

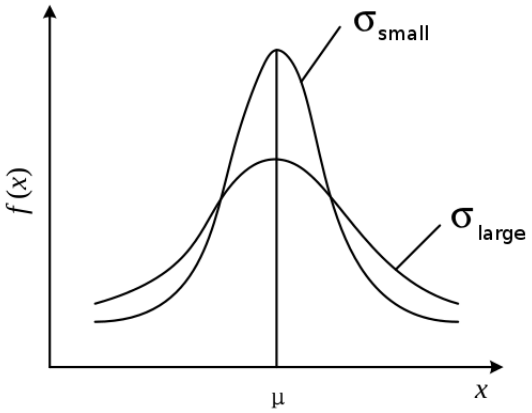

Figure 1: Probability density function for the normal distribution. Note the difference in the shape for σ_{small} and σ_{large} (where $\sigma_{\text{small}} < \sigma_{\text{large}}$) illustrating different dispersion around the average value μ.

2.1.2 Standard Deviation of the Average Value in Multiple Measurements

Although the exact value μ_X of a physical quantity X objectively exists, because of randomness of measurement outcomes, it is impossible to measure it directly and its value may be only estimated. If a large number of independent measurements is performed, it may be shown that the probability density curve assumes the shape of a Gaussian with the *arithmetic mean* being the best estimate of the exact value μ_X.

If we perform n independent measurements of a physical quantity X, obtaining a set of results $x_1, x_2, \ldots, x_n$, the average value $\bar{X}$ of X is defined as

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

and the standard deviation σ_X may be estimated by the quantity

$$s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2},$$

quantifying the dispersion of measurement results around the average value. Again, if the number of measurements is large, the interval $(\bar{X} - s_X, \bar{X} + s_X)$ estimates the 0.683–confidence interval.

Because the arithmetic mean provides us with the best estimate of the exact value of a physical quantity, for obvious reasons we are also interested in the dispersion of $\bar{X}$. Within the probability theory it may be shown that the standard deviation of $\bar{X}$ may be estimated as

$$s_{\bar{X}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2} = \frac{s_X}{\sqrt{n}}. \tag{1}$$

The above discussion clearly indicates that the standard deviation of the arithmetic mean (calculated based on n measurements) is estimated at $1/\sqrt{n}$ of the value corresponding to a single measurement, which means that the exact value of the physical quantity lies in the interval $(\bar{X} - s_{\bar{X}}, \bar{X} + s_{\bar{X}})$ of smaller length than that of $(\bar{X} - s_X, \bar{X} + s_X)$.

It is worth to emphasize again that only if the number of measurements is large enough, the results follow the normal distribution and the above discussion is valid. In the case when the number of measurements is relatively small, statistical distribution of the results is better described by the so called Student's t–distribution (or simply t–distribution). Then the probability that the measurement yields a value from $(\bar{X} - s_{\bar{X}}, \bar{X} + s_{\bar{X}})$ is not 0.683. In this case a ξ–confidence interval is defined as $(\bar{X} - t_\xi s_{\bar{X}}, \bar{X} + t_\xi s_{\bar{X}})$. Usually the confidence level $\xi = 0.95$ is used. The values of $t_{0.95}$ and $t_{0.95}/\sqrt{n}$ are given in Table 1.

n	3	4	5	6	7	8	9	10	15	20	≥ 100
$t_{0.95}$	4.30	3.18	2.78	2.57	2.45	2.36	2.31	2.26	2.14	2.09	≤ 1.97
$\frac{t_{0.95}}{\sqrt{n}}$	2.48	1.59	1.204	1.05	0.926	0.834	0.770	0.715	0.553	0.467	≤ 0.139

Table 1: The values of $t_{0.95}$ and $t_{0.95}/\sqrt{n}$.

2.1.3 Estimation of Type-*A* Measurement Uncertainties

Measurement uncertainties of type A are usually estimated by using formula (1) as the estimate (at the confidence level 0.95) for the standard deviation of the average value obtained from multiple measurements. For example, for $n = 6$, from Table 1 we have $t_{0.95}/\sqrt{6} \approx 1$, and in this case $\Delta_A = s_X$ can be used to quantify the uncertainty of the measurement's result.

2.2 Type-*B* Uncertainties

Measurement uncertainties of type B, denoted as Δ_B, do not have a statistical nature, but result from the design of a measurement instrument itself. These uncertainties may be estimated in a simple way by the maximum uncertainty Δ_{dev} inherent to the instrument working in a given measurement range (see Table 2).

2.3 Presentation of Measurement Results

An experimentally found value of a physical quantity should always include the information about its uncertainty and should be presented in the following form

$$x = \bar{X} \pm u, \qquad u_r = \frac{u}{\bar{X}} \times 100\% \tag{2}$$

where $\bar{X}$ is the average value of a multiple measurement, $u = \sqrt{\Delta_A^2 + \Delta_B^2}$ is the measurement uncertainty (including contributions due to type–A and type–B uncertainties), and u_r denotes the relative uncertainty.

Device	*Range*	*Graduation*	*Maximum Error*
steel ruler	150 mm	1 mm	±0.10 mm
	500 mm	1 mm	±0.15 mm
	1000 mm	1 mm	±0.20 mm
steel tape meter	1 m	1 mm	±0.8 mm
	2 m	1 mm	±1.2 mm
calliper	125 mm	0.02 mm	±0.02 mm
		0.05 mm	±0.05 mm
micrometer screw gauge	$0 \sim 25$ mm	0.01 mm	±0.004 mm
class-7 scale	500 g	0.05 g	0.08 g (at full-load)
		0.05 g	0.06 g (at 1/2 load)
		0.05 g	0.04 g (at 1/3 load)
class-3 scale	200 g	0.1 mg	1.3 mg (at full-load)
		0.1 mg	1.0 mg (at 1/2 load)
		0.1 mg	0.7 mg (at 1/3 load)
standard thermometer	$0 \sim 100^\circ$ C	1° C	$\pm 1^\circ$ C
high-precision thermometer	$0 \sim 100^\circ$ C	0.1° C	$\pm 0.2^\circ$ C
electronic meter (class 0.5)			0.5%× range
electronic meter (class 0.1)			0.1%× range
electronic multimeter			*

* $\alpha\% \times U_x + \beta\% \times U_m$ (where U_x is the actual reading, and U_m is the range.

The parameters α and β have different values for different functions of the meter.

Table 2: Measurement devices and measurement uncertainties.

On graphs, the uncertainty of the data is indicated by error bars.

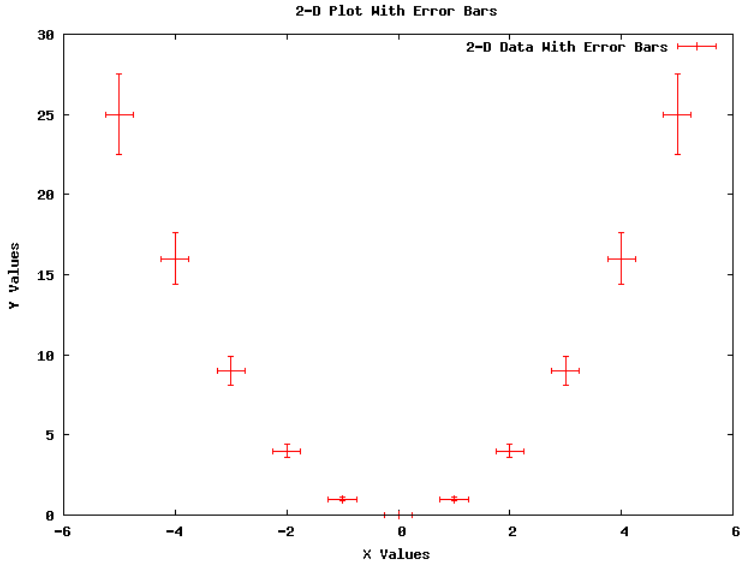

Figure 2: Experimental data with uncertainties indicated by error bars (Source: nsnam.org).

2.4 Measurement Uncertainty of Directly Measurable Quantities

2.4.1 Single Measurement

A single measurement is performed in the following situations:

a) The resolution of a measurement device is relatively low, so that multiple measurements would yield the same result.

b) Very high precision of the measurement is not required.

c) Because of limitations due to the environment, it is possible to complete only one measurement.

In the case of a single measurement, its result is also presented in the form (2) with u usually taken equal to Δ_{dev}.

2.4.2 Multiple Measurements

In the case of a multiple measurement, yielding the results $x_1, x_2, \ldots, x_n$, the uncertainty of the value of the physical quantity being measured is calculated by in the following procedure:

(1) Calculate the average value (arithmetic mean) of measurement results $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$

(2) Calculate $s_X = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2}$

(3) Find $\Delta_A = s_X\ t_{0.95}/\sqrt{n}$. For example, if the number of measurements $n = 6$, then $t_{0.95}/\sqrt{n} = 1$ and $\Delta_A = s_X$

(4) According to the resolution of the measurement device, determine the maximum error Δ_{dev} and assume $\Delta_B = \Delta_{\text{dev}}$

(5) Calculate the total uncertainty $u = \sqrt{\Delta_A^2 + \Delta_B^2}$ and relative uncertainty $u_r = \frac{u}{\bar{X}} \times 100\%$.

(6) Write down the result as $x = \bar{X} \pm u$ and $u_r = \frac{u}{\bar{X}} \times 100\%$.

Example

At the room temperature, using the resonance interference method, the length λ of an ultrasonic wave propagating in air has been measured (see Table 3). What is the result of this measurement and its uncertainty?

i	1	2	3	4	5	6
λ [cm]	0.6872	0.6854	0.6840	0.6880	0.6820	0.6880

Table 3: Measurement data.

The result, *i.e.* the average value of the wavelength measured in this experiment

$$\bar{\lambda} = \frac{1}{6}\sum_{i=1}^{6} \lambda_i = 0.6858 \text{ cm},$$

and its standard deviation of the result of a measurement of the wavelength may be estimated

$$s_\lambda = \sqrt{\frac{1}{6-1}\sum_{i=1}^{6}\left(\lambda_i - \bar{\lambda}\right)^2} = \sqrt{\frac{1}{5} \times 2.9 \times 10^3 \times 10^{-8}} \text{ m} \approx 0.0024 \text{ cm}.$$

The instrument used to measure the wavelength has the maximum error $\Delta_{\text{dev}} = 0.002$ cm. Hence, the A–type uncertainty of the wavelength found by this experimental method is $\Delta_A = s_\lambda = 0.0024$ cm and the corresponding B–type uncertainty is $\Delta_B = \Delta_{\text{dev}} = 0.002$ cm.

Hence, the combined uncertainty is

$$u_\lambda = \sqrt{\Delta_A^2 + \Delta_B^2} = \sqrt{(0.0024)^2 + (0.0020)^2} \text{ cm} \approx 0.0031 \text{ cm}$$

and corresponding relative uncertainty

$$u_{r\lambda} = \frac{u_\lambda}{\bar{\lambda}} \times 100\% = 0.5\%.$$

The experimentally found wavelength is

$$\lambda = 0.6858 \pm 0.0031 \text{ cm}, \qquad u_{r\lambda} = 0.5\%.$$

2.5 Indirect Measurements and Propagation of Uncertainties

Usually we are not able to measure the physical quantity in question directly. Instead, we measure other physical quantities and then find the quantity by applying a formula (a law of physics or a definition) that relates the physical quantities involved.

Let us assume that a physical quantity F is expressed in terms of other physical quantities $X, Y, Z, \ldots$ by the formula

$$F = F(X, Y, Z, \ldots). \tag{3}$$

If $X, Y, Z, \ldots$ are measured, some values $x, y, z, \ldots$, are obtained with the corresponding measurement uncertainties $u_x, u_y, u_z, \ldots$. Of course these uncertainties contribute to uncertainty u_F of the value f of the physical quantity F that is to be determined from the experiment.

The way the uncertainties propagate can be determined from the formula (3). Treating the uncertainties as small increments and recalling the formula for the total differential of a function of many variables, the propagated uncertainty is usually estimated by the formula

$$u_F = \sqrt{\left(\frac{\partial F}{\partial X}\right)^2 (u_X)^2 + \left(\frac{\partial F}{\partial Y}\right)^2 (u_Y)^2 + \left(\frac{\partial F}{\partial Z}\right)^2 (u_Z)^2 + \ldots} \tag{4}$$

with the relative uncertainty

$$u_{rF} = \frac{u_F}{\bar{F}} = \sqrt{\left(\frac{\partial \ln F}{\partial X}\right)^2 (u_X)^2 + \left(\frac{\partial \ln F}{\partial Y}\right)^2 (u_Y)^2 + \left(\frac{\partial \ln F}{\partial Z}\right)^2 (u_Z)^2 + \ldots.} \tag{5}$$

The procedure of determining uncertainties of indirectly measurable quantities consists of the following steps

(1) Find individual measurement uncertainties $u_x, u_y, u_z, \ldots$ corresponding to the results $x, y, z, \ldots$ of measurements of the physical quantities $X, Y, Z, \ldots$.

(2) Use the relation $F = F(X, Y, Z, \ldots)$ to calculate the corresponding partial derivatives $\partial F/\partial X, \partial F/\partial Y, \partial F/\partial Z, \ldots$ or $\partial \ln F/\partial X, \partial \ln F/\partial Y, \partial \ln F/\partial Z, \ldots$.

(3) Use the formulas (4) and (5) to find u_F and u_{rF}, respectively.

Formulas for the uncertainty of an indirectly measured quantity for some common relations are listed in Table 4.

Relation $F = F(X, Y, Z)$	*Uncertainty (or relative uncertainty) propagation formula*
$X + Y$	$u_F = \sqrt{u_X^2 + u_Y^2}$
$X - Y$	$u_F = \sqrt{u_X^2 + u_Y^2}$
αX	$u_F = \alpha u_X,\ u_r = \frac{u}{x}$
$\sqrt[k]{X}$	$u_r = \frac{1}{k}\frac{u}{x}$
XY	$u_r = \sqrt{u_{rX}^2 + u_{rY}^2}$
$\frac{X}{Y}$	$u_r = \sqrt{u_{rX}^2 + u_{rY}^2}$
$\sin X$	$u_F = \lvert\cos x\rvert u_X$
$\ln X$	$u = \frac{u_X}{x} = u_{rX}$

Table 4: Uncertainty propagation formulas.

(4) Write down the result as $F = \bar{F} \pm u_F$ and $u_{rF} = \frac{u_F}{\bar{F}} \times 100\%$, where $\bar{F} = F(\bar{X}, \bar{Y}, \bar{Z}, \ldots)$.

Example

The inner diameter of a hollow metallic cylinder has been measured as $D_1 = 2.880 \pm 0.004$ cm and the outer diameter $D_2 = 3.600 \pm 0.004$ cm. The cylinder's height is $H = 2.575 \pm 0.004$ cm. Find the volume of the cylinder and the uncertainty of the result.

Given the data, the volume of the cylinder is

$$\bar{V} = \frac{\pi}{4}\left(D_2^2 - D_1^2\right) H = \frac{\pi}{4} \times \left(3.600^2 - 2.880^2\right) \times 2.575 \text{ cm}^3 = 9.436 \text{ cm}^3.$$

In order to find the relative uncertainty, first find the partial derivatives

$$\frac{\partial \ln V}{\partial D_2} = \frac{2D_2}{D_2^2 - D_1^2}, \qquad \frac{\partial \ln V}{\partial D_1} = -\frac{2D_1}{D_2^2 - D_1^2}, \qquad \frac{\partial \ln V}{\partial H} = \frac{1}{H}.$$

Then, according to formula (5), the relative uncertainty

$$u_{rV} = \frac{u_V}{\bar{V}} = \sqrt{\left(\frac{2D_2}{D_2^2 - D_1^2}\right)^2 (u_{D_2})^2 + \left(-\frac{2D_1}{D_2^2 - D_1^2}\right)^2 (u_{D_1})^2 + \left(\frac{1}{H}\right)^2 (u_H)^2}$$

Substituting the numbers we obtain $u_{rV} = 0.008 = 0.8\%$. The absolute uncertainty $u_V = \bar{V} u_{rV} = 9.436 \times 0.008 \approx 0.08$ cm^3. Hence the result

$$V = 9.44 \pm 0.08 \text{ cm}^3, \qquad u_{rV} = 0.8\%.$$

3 Significant Digits

3.1 Idea

Since measurement results are subject to uncertainty, the number of digits after the decimal place must not be set arbitrarily and is limited. Measurement results should only be written up to one or two digits determined by the result's uncertainty. The subsequent digit should follow a rounding rule. Obviously, the last one or two significant digits are uncertain and significant digits my be regarded as a rough method to present uncertainty.

The number of significant digits is independent of the position of the decimal point. For instance, 1.23 and 123 both have three significant digits. As to the question whether zero is a significant digit, it can be explained by formulating the following rule: moving from the left to the right, start with the first non-zero digit. A zero to its left is not a significant digit, while a zero to its right is a significant digit. *E.g.*, 0.0123 has three significant digits, but 0.01230 has four significant digits. So "0" should not be omitted when it is a significant digit: for 1.3500 cm (with the last zero being a significant digit), it is wrong to write as 1.35 cm for they differ in the degree of accuracy.

The number of significant digits can roughly reflect the magnitude of relative error. The more significant digits, the smaller the relative error is and the higher the accuracy of the measurement results.

3.2 Notation

Significant digits in a measurement result are determined by the measurement's uncertainty. As uncertainty itself is just an estimated value, only one or two significant digits of uncertainty are taken in general. The last digit of the measured value should correspond with the last digit of uncertainty. In the initial stage, it may be assumed that only the last digit is uncertain. Accordingly, the uncertainty is also determined up to only one significant digit, such as $L = (1.00 \pm 0.02)$ cm. Significant digits of a single direct measurement are determined by the instrument error or uncertainty of estimation. As for multiple measurements, significant digits in their average value are determined by its uncertainty. Finally, significant digits of an indirect measurement are also determined by the (propagated) uncertainty of the result, which was calculated beforehand.

Scientific notation is usually used to denote numerical values. For instance, if in a given year the population of China was seven hundred and fifty million and its uncertainty was twenty million, it should be written as $(7.5 \pm 0.2) \times 10^8$ people, where (7.5 ± 0.2) indicate significant digits and uncertainty, and 10^8 people is the unit.

In another example, (0.000623 ± 0.000003) m should be denoted as $(6.23 \pm 0.03) \times 10^{-4}$ m.

3.3 Operation Rules

The following operation rules should be observed while operating with measurement results:

a) When quantities are being added or subtracted, the number of decimal places (not significant digits) in the answer should be the same as the least number of decimal places in any of the numbers being added or subtracted.

b) In multiplication or division, the number of significant digits in the result should equal the least number of significant digits in any one of the numbers being multiplied or divided.

c) In evaluating a power or a root, the number of significant digits in the answer should be the same as the base.

d) In principle, the number of significant digits when evaluating the value of a function should be determined by uncertainty analysis. In the physics laboratory however, the rules for logarithmic, exponential or trigonometric functions are usually simplified.

e) There is a class of numbers called "correct numbers". For example, when the radius is used to find the diameter with the formula $d = 2r$, the factor 2, which is not determined by measurements, is called a "correct number". The "correct numbers" do not obey the operation rules formulated for significant digits. The number of significant digits is determined by other numbers.

f) When a constant such as π or g appears in a calculation, the number of digits in the numerical value of the constant should be taken to be the same as the measured quantities. For example, the circumference of a circle is $l = 2\pi R$. If $R = 2.356$mm, π should be taken as 3.142.

4 Presentation of Experimental Data

The data acquired in the experiment needs to be processed in order to facilitate analysis and formulate conclusions.

4.1 Tables

Data should be listed in tables when recorded, as it is convenient to check for potential errors. Data tables should be brief and clear. The symbols representing physical quantities and their units should be placed in the first row of the table and not put in each row repeatedly. The data in the table should include the correct number of significant figures. If the final results are presented in the table they should be accompanied by the value of the results' uncertainty.

4.2 Graphs

Graphs are often used to illustrate and discuss results of experiments. They are intended to reveal relations between physical quantities. For example, for a linear relation $Y =$

$mX + b$ between two physical quantities X and Y, it is possible to determine the value of m from the slope of the line and the value of b from the line intercept. In general, a value y, not measured in the experiment, corresponding to the value x can be also found from the line $Y = Y(X)$ (interpolation). Under certain conditions it is possible to read a value beyond the range of the measurement data (extrapolation).

When the functional relation to be tested in the experiment is nonlinear, a change of variables can be performed in order to transform it to a linear one. For example, to illustrate the relation $PV = C$, the graph P vs. V can be replaced by the graph P vs. $1/V$.

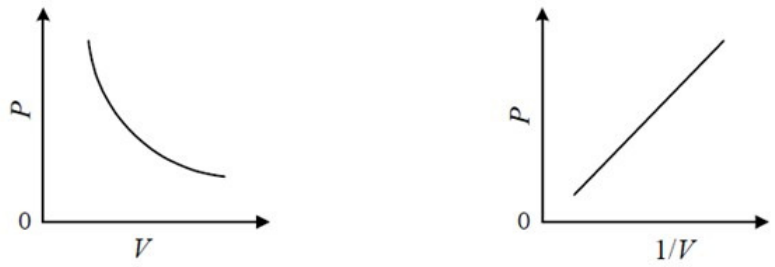

Figure 3: Transformation of variables.

Plotting rules

1) Choose an appropriate scale: linear, logarithmic, or double logarithmic.

2) Generally, the horizontal axis represents the independent variable and the vertical axis represents the dependent variable. Indicate the physical quantities and their units next to the axes. Adjust the scale of the horizontal and vertical axes and the position of the origin properly, in order to have the curve plotted in the center of the graph covering $70\% - 80\%$ of the graph sheet.

   Notes for the scale:

   1) The number of significant digits for experimentally measured data read from the graph should not be less than that obtained in the experiment. For example, for physical variables measured directly, the scale of the smallest axis grid should not be greater than the graduation of the measuring instrument.
   2) The scales of the horizontal and the vertical axis can be different in order to accommodate the graph.
   3) If numbers are particularly large (or small), a multiplication factor, such as $\times 10^3$ or $\times 10^2$ can be written in front of the units.

3) The coordinates of the measured data points should be marked with a small "+" character and uncertainties should be indicated by error bars. When several curves are to be drawn in a graph, each curve can be marked by different symbols like $\times$, $\odot$, $\triangle$, $\boxdot$, *etc.* Depending on circumstances draw a curve. Remember that the curve does not have to (and usually will not) pass all points (see the next section).

4) To plot graphs with the experimental data, use a data processing and plotting software (*e.g.* Origin).

5 The Least-Squares Method

This section contains a brief presentation of the least-squares which is used to fit a curve, given a set of experimental data.

In general, there are situations when given the experimental data, one looks for a curve that fits the data in the "best" way. One case is when the form of the functional dependence between two physical quantities is known, *i.e.* it is known that $y = f(x)$, and the "best" values of uncertain parameters should be found. In the other case, the functional dependence (empirical) should be first determined and then the "best" values of parameters should be found. In experimental physics most experiments belong to the first class, thus only this method will be introduced.

Let us assume that the form of the known function is

$$y = b_0 + b_1 x \tag{6}$$

i.e. there is only one independent variable x. In this case the method is known as simple linear regression.

The experimental data is given as two sets of numbers

$$x_1, x_2, ..., x_k$$
$$y_1, y_2, ..., y_k.$$

If the experimental data did not carry any uncertainty, both sides of Eq. (6) should be equal when the pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_k, y_k)$ are substituted for (x, y). However, in practice, some uncertainty is always present and it is represented by the residuals $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i$. Then

$$\begin{aligned} y_1 - b_0 - b_1 x_1 &= \varepsilon_1, \\ y_2 - b_0 - b_1 x_2 &= \varepsilon_2, \\ &\vdots \\ y_k - b_0 - b_1 x_k &= \varepsilon_k. \end{aligned}$$

The idea of the method is to determine the uncertain parameters b_0 and b_1, and also require the sum of squares of the residuals $\sum_{i=1}^{k} {\varepsilon_i}^2$ to be minimum. Rewriting the sum

as

$$\sum_{i=1}^{k} {\varepsilon_i}^2 = \sum_{i=1}^{k} (y_i - b_0 - b_1 x_i)^2, \tag{7}$$

the problem simplifies to finding the minimum of a function of two variables b_0 and b_1 defined by the right hand side of Eq. (7) has to be found. The outlined procedure allows one to find

a) The optimal (in the least-squares sense) estimate of the slope and the intercept of the regression line

$$\begin{aligned} b_1 &= \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}, \\ b_0 &= \overline{y} - b_1 \overline{x}, \end{aligned}$$

where the bar denotes the arithmetic average.

b) The standard deviation of all quantities: the standard deviation of the dependent variable

$$\sigma_y = \sqrt{\frac{1}{k-n} \sum_{i=1}^{k} {\varepsilon_i}^2}$$

where k is the number of measurements, and n is the number of unknown quantity; the standard deviation of the slope estimate

$$\sigma_{b_1} = \frac{\sigma_y}{\sqrt{\overline{x^2} - \overline{x}^2}}$$

and the standard deviation of the intercept estimate

$$\sigma_{b_0} = \sqrt{\overline{x^2}} \sigma_{b_1}.$$

Various fitting procedures (including the least-squares method) are implemented in data processing software (*e.g.* Origin).

6 Exercises

1. Figure out the number of significant digits for the following quantities. Then keep **three** significant digits and re-express the quantities using scientific notation.

   (a) 1.0850 cm

   (b) 2575.0 g

   (c) 3.141592654 s

   (d) 0.86249 m

   (e) 0.0301 kg

   (f) $979.436\ \text{cm} \cdot \text{s}^{-2}$

2. Correct the following statements using appropriate rules for significant digits.

   (a) 0.30 m is equal to 30 cm, and 30 cm is equal to 300 mm;

   (b) Student A states that 0.1230 has five significant digits, while student B argues that it has three significant digits. Who is correct? Explain.

   (c) In a lab report experiment results were found to be expressed as follows

   - $d_1 = (10.800 \pm 0.02)$ cm,
   - $d_2 = (10.800 \pm 0.123)$ cm,
   - $d_3 = (10.8 \pm 0.002)$ cm,
   - $d_4 = (10.8 \pm 0.12)$ cm.

   Express the results correctly and calculate their relative uncertainty.

3. Find significant digits for the following values

   (a) $\sin 20°16'$,

   (b) $\ln 280.3$,

   (c) $e^{3.250}$.

4. A physical quantity X is found indirectly by measuring quantities X_1 and X_2, and using the formula $Y = X_1 + X_2$. Calculate the value y of the physical quantity Y yielded by this experiment and determine its uncertainty if

   (a) $x_1 = 1.1 \pm 0.1$ cm, $x_2 = 2.387 \pm 0.001$ cm.

   (b) $x_1 = 37.13 \pm 0.02$ mm, $x_2 = 0.623 \pm 0.001$ mm;

5. $Z = \alpha + \beta + \gamma$, where $\alpha = 1.218 \pm 0.002\ \Omega$, $\beta = 2.1 \pm 0.2\ \Omega$, $\gamma = 2.14 \pm 0.03\ \Omega$. Calculate Z.

6. Let $I = 1.00 \pm 0.05$ A, $R = 1.00 \pm 0.03\ \Omega$. Use Ohm's law ($U = IR$) to determine U.

7. Use operation rules for significant digits to calculate the results of the following operations (clearly state all steps)

   (a) $\frac{76.000}{40.00-2.0}$

   (b) $\frac{50.00\times(18.30-16.3)}{(103-3.0)(1.00+0.001)}$

   (c) $\frac{100.0\times(5.6+4.412)}{(78.00-77.0)\times 10.000} + 110.0$

8. In the auto-collimation method, a 1-meter steel tape is used to measure the focal length f of a convex lens. The eight results are as follows: 116.5 mm, 116.8 mm, 116.5 mm, 116.4 mm, 116.6 mm, 116.5 mm, 116.7 mm, 116.2 mm. Calculate and write down the final result of this experiment. Remember to include its uncertainty.

9. Mass of an object is measured with an analytical balance. The measurement is repeated six times with the results: 3.6127 g, 3.6122 g, 3.6121 g, 3.6120 g, 3.6123 g and 3.6125 g. What is the result of this measurement?

10. A period of a simple pendulum has been measured with a stopwatch. The experimenter has measured the time of one period as 1.9 s, 10 periods – 19.3 s, and 100 period – 192.8 s. He thought that he used the same stopwatch and each measurement was a single measurement. A stopwatch normally has an error of 0.1 s, so he thought each measurement had an error of 0.2 s. What is your comment on that?

    If 10 periods are measured, and the results are: 19.3, 19.2, 19.4, 19.5, 19.3, 19.1, 19.2, 19.5, 19.4, 19.5 [s], what is the result from this group of data?

11. The acceleration due to gravity g is measured by measuring the period of a simple pendulum of different lengths with the experimental data recorded in the table below

Pendulum length L [cm]	61.5	71.2	81.0	89.5	95.5
Period T [s]	1.571	1.696	1.806	1.902	1.965

    Plot the graphs L vs. T and L vs. T^2 and find g.

12. The temperature of a sample (liquid) has been measured for ten times yielding the following data

	T [°C]		T [°C]
1	20.42	6	20.43
2	20.40	7	20.43
3	20.42	8	20.43
4	20.39	9	19.20
5	20.30	10	20.43

    Calculate the average value of T and determine whether there is a gross error.

13. Derive the propagation formula for uncertainty of the following indirectly measured quantities

    (a) $I = I_0 e^{-\beta x}$, where I_0, β are constant;

    (b) $Y = AX^B$, with A, B – constants;

    (c) $N = \frac{\sin\frac{A+D}{2}}{\sin\frac{A}{2}}$;

    (d) $E = \frac{Mgl}{\pi r^2 L}$;

    (e) $R_x = \frac{R_1}{R_2} R$.

14. Find mistakes in the following expressions and correct them

    (a) $a = 8.524$ m $\pm$ 50 cm;

    (b) $t = 3.75$ h $\pm$ 15 min;

    (c) $g = 9.812 \pm 14 \times 10^{-2}$ [m/s^2];

    (d) $S = 25.400 \pm \frac{1}{30}$ [mm].

15. The depth of water in a tank is h. There is a hole of diameter d at the bottom of the tank. The time needed to drain the tank completely is measured with the results listed in the following table (in seconds).

d [cm] \ h [cm]	30.0	10.0	4.0	1.0
1.5	73.0	43.5	26.7	13.5
2.0	41.2	23.7	15.0	7.2
3.0	18.4	10.5	6.8	3.7
5.0	6.8	3.9	2.2	1.5

    Try to find the rule (an empirical formula) for the tank drainage time $t = f(d, h)$ by the graphical method.

    (a) Plot the graphs d vs. t, $1/d^2$ vs. t, $\ln d$ vs. $\ln t$. Here h is constant.

    (b) Plot the graphs h vs. t, $\ln h$ vs. $\ln t$. Here d is constant.

    (c) Predict t based on the family of curves plotted, if

    - $d = 4$ cm or $d = 6$ cm,
    - $d = 4$ cm and $h = 20$ cm.

    (d) Deduce the family of curves $t = f(d, h)$. Calculate the value of t in (c) and compare with your prediction. Explain why they are different.

16. Resistance is measured with a voltmeter. Values of the current at different voltages are listed in the table below. Use Origin to plot the V–A characteristics, calculate the resistance and compare with the result directly obtained from the average value.

U [V]	0.400	0.600	0.800	1.000	1.200	1.400	1.600	1.800
$I \times 10^{-3}$ [A]	10.4	15.5	23.5	25.6	30.5	35.5	40.2	45.2