KAPITAŁ LUDZKI
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA
EUROPEJSKI
FUNDUSZ SPOŁECZNY

Projekt pn. „*Wzmocnienie potencjału dydaktycznego UMK w Toruniu w dziedzinach matematyczno-przyrodniczych*”
realizowany w ramach Poddziałania 4.1.1 Programu Operacyjnego Kapitał Ludzki

Chemistry IT

Piotr Szczepański

UMK Toruń 2012

LITERATURE

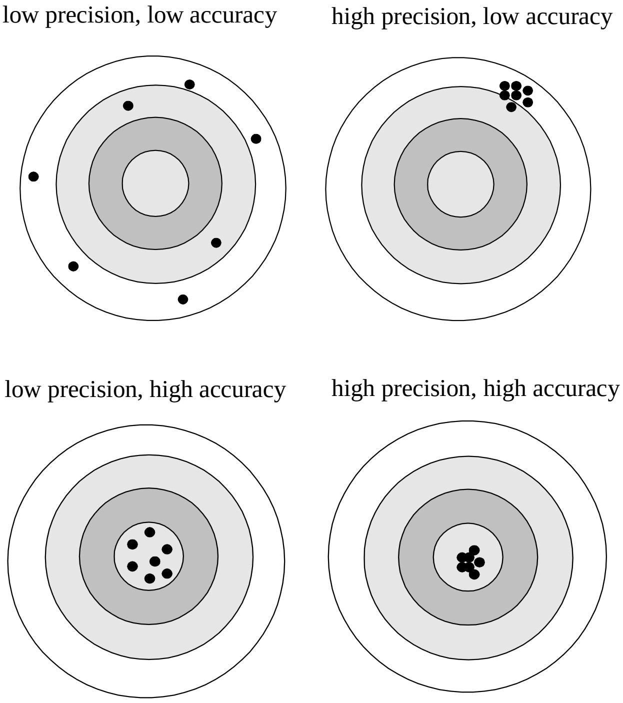
1. R. Wódzki, *Zastosowanie informatyki w chemii*, Toruń 1999
2. Z. Fortuna, B. Macukow, J. Wąsowski, *Metody numeryczne*, WNT 2006
3. H. Hänsel, *Podstawy rachunku błędów*, WNT 1968
4. A. Ralston, *Wstęp do analizy numerycznej*, PWN 1983
5. J.B. Czermiński, A. Iwasiewicz, Z. Paszek, A. Sikorski, *Metody statystyczne dla chemików*, PWN 1992
6. A. Łomnicki, *Wprowadzenie do statystyki dla przyrodników*, PWN, Warszawa 2000
7. J. Arendalski, *Niepewność pomiarów*, Oficyna wydawnicza PW 2006
8. J. Koronacki, J. Mielniczuk, *Statystyka dla studentów kierunków technicznych i przyrodniczych*, WNT 2001
9. E. Bulska, *Metrologia chemiczna*, Wydawnictwo MALAMUT, Warszawa 2008
10. P. Konieczka, J. Namieśnik, *Ocena i kontrola jakości wyników pomiarów analitycznych*, W. N.-T., Warszawa 2007
11. E. Steiner, *Matematyka dla chemików*, Wydawnictwo Naukowe PWN, Warszawa 2001
12. T.E. Shoup, *Applied numerical methods for the microcomputer*, Prentice-Hall, Inc. 1984
13. *Guide to the Expression of Uncertainty in Measurement*, ISO, Switzerland 1995.
14. J. Mazerski, *Chemometria praktyczna. Zinterpretuj wyniki swoich pomiarów*, Wydawnictwo MALAMUT, Warszawa 2009
15. D.W. Rogers, Computational chemistry using the PC, John Wiley & Sons. Inc., 2003
16. P. Gemperline, Practical guide to chemometrics, CRC Press, 2006
17. P. C. Jurs, Computer software applications in chemistry, John Wiley & Sons. Inc., 1996
18. M. Otto, Chemometrics. Statistics and computer application in analytical chemistry, WILEY-VCH Verlag GmbH, 1999
19. J. N. Miller, J. C. Miller, Statistics and chemometrics for analytical chemistry, Pearson Education Limited, 2000

CONTENTS

1. Introduction to statistical analysis of experimental data

In the course of repeated experiments, errors varying according to their nature may occur. One of them is a systematic uncertainty (formerly a systematic error) which is typical for experiments conducted in exactly the same conditions. It results from imperfect instruments, errors during calibration, an instrument's drift over time, optical parallax, and imperfections of an observer. This error can be corrected or eliminated by conducting a so called blind test, proper calibration and a careful experiment. This type of error determines the accuracy of an experiment, i.e. the closeness of a measurement result to an actual value.

A random uncertainty (formerly a random error) stands for small, uncontrolled fluctuations in the experimental measurements that result from a myriad of causes affecting conditions of an experiment (a random variable). This error alone can be analysed using statistical methods. It is critical for the precision of a measurements i.e. the reproducibility of a result in repeated experiments. The two terms accuracy and precision are schematically compared in Fig. 1.1.



low precision, low accuracy

high precision, low accuracy

low precision, high accuracy

high precision, high accuracy

Fig. 1.1. Accuracy vs. precision scheme.

Sometimes we can also distinguish a mistake (gross error) which is associated with inattention of an experimenter (e.g. poor reading, damage to equipment). Therefore, a measurement outcome may significantly deviate from others.

The general rule in data analysis is that questionable results cannot be rejected without mathematical justification. A questionable result (outlier) can be rejected only if it is indicated by a result of an appropriate test such as: Dixon's, 3d, or Grubbs's tests, etc. Each has its advantages and disadvantages.

The Dixon's test (test Q) is a common test used for identification and rejection of outliers. In this test, the Q ratio is calculated from the formula:

$$\mathrm{Q} = \frac{\left| x_{\mathrm{suspect}} - x_{\mathrm{nearest}} \right|}{x_{\mathrm{max}} - x_{\mathrm{min}}} \tag{1.1}$$

i.e. the difference between the outlier and the closest result divided by the range ($x_{max} - x_{min}$). The suspect result is rejected if the calculated value Q is larger than the critical value (Q_{cr}) indicated in the table (Tab. 1.1), dependent on the number of measurements *n*.

Tab. 1.1. Critical values of Q (Q_{cr}) for Dixon's test

n	3	4	5	6	7	8	9	10	∞
Q_{cr}	0.94	0.76	0.64	0.56	0.51	0.47	0.44	0.41	0.00

The Dixon's test allows for rejection of only one outlier in a given series of measurements.

The 3d test calculates the arithmetic mean of deviations of points from the mean, without regard to a outlier:

$$d = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n} \tag{1.2}$$

According to this method, if a suspected result is not within a prescribed range (±3d), it should be rejected.

1.1. Significant digits (figures)

The way a numerical quantity is recorded is closely related to the precision with which this value has been set. A correct recording of measurement results related to the calculus of errors generally requires that a result and its uncertainty are rounded off. The reason why uncertainties and final results should be rounded off can be presented in an example. The mean value and its uncertainty obtained after a few hundred measurements of polyester coating thickness using micrometer are presented below:

120.342525794323 ± 9.722742949332 μm

Putting the result and its uncertainty in such a form suggests that the precision of the measurements is larger than the size of an atom (the fourth decimal place), the size of a nucleus of an atom (the eighth decimal place) and comparable to the size of a quark (the last, twelfth decimal place). The value and its uncertainty so recorded are far from the acceptable precision with which the measurement was made. According to the error calculus, this result should be written as follows:

120.3 ± 9.8 μm

The example shows that measurement results should be given together with an uncertainty and a unit.
An uncertainty value is given with an accuracy of up to two significant digits.
If an uncertainty value (rounded) does not increase by more than 10%, only one digit may be left (for example 0.88 is round up to 0.9). It must also be noted that uncertainties are always rounded up. In selecting the number of significant digits of a result, the last digit of the result and an uncertainty should be in the same decimal place (e.g. 32.3 ± 0.7).

According to the rules, the significant figures of a number are digits from 1 to 9 and zero, if:

a) zeroes are placed between other non−zero digits, or
b) zeroes are placed at the end of a number containing a decimal point.

For example, the number:

6.321 $4.345 \cdot 10^{-3}$ 0.001307 $1.000 \cdot 10^{4}$

have four significant digits. Examples of the result with various number of significant digits and decimal places are presented in Tab. 1.2.

Tab. 1.2. Examples of the result with various number of significant digits and decimal places

Result	Number of significant digits	Number of decimal places
42.8	3	1
0.345830	6	6
0.543	3	3
0.0038	2	4
0.00028040	5	8

The measured numbers are usually rounded off to the degree of accuracy. According to the rules for rounding off, the numeric values are:

a) round up, if the last digit is ≥6,

b) round down, if the last digit is ≤4, or

c) if the last digit is equal 5: round up, if at least one from the removed, non–significant digits is non-zero, or to the nearest even digit.

Some rounding examples are presented below:

A= 0.7756 g round to A= 0.776 g rule a)
A=0.7753 g round to A=0.775 g rule b)
A= 0.77551 g round to A= 0.776 g rule c)
A= 0.7755 g round to A= 0.776 g rule c)
A= 0.7765 g round to A= 0.776 g rule c)

The following scheme (Fig. 1.2) illustrates the rounding of the measured values.

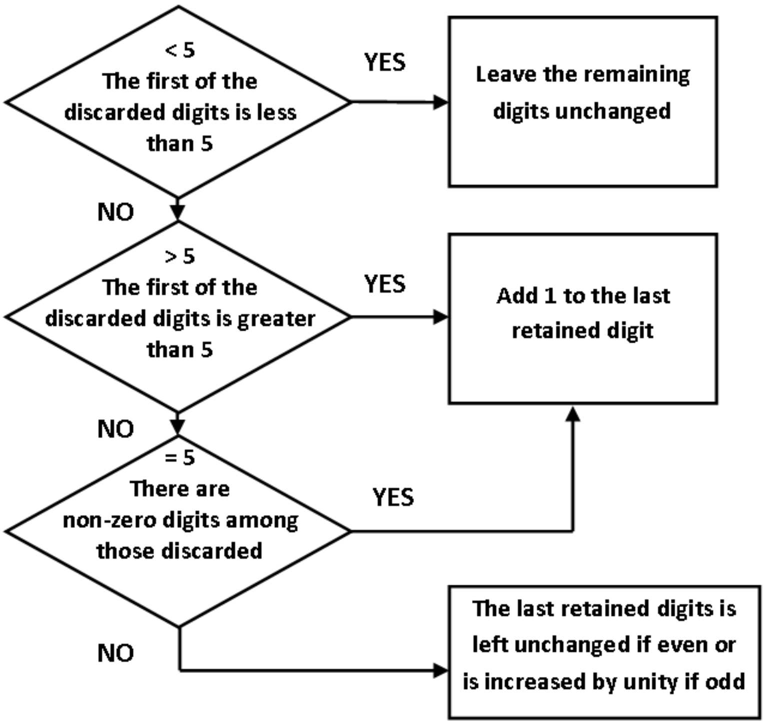

< 5
The first of the discarded digits is less than 5
YES
Leave the remaining digits unchanged
NO
> 5
The first of the discarded digits is greater than 5
YES
Add 1 to the last retained digit
NO
= 5
There are non-zero digits among those discarded
YES
NO
The last retained digits is left unchanged if even or is increased by unity if odd

Fig. 1.2. Rounding of the measured values.

1.2. Statistical analysis of random error

The occurrence of random errors in the course of repeated measurements means that the obtained results (x_i) show distribution (dispersion). Therefore, some values of x_i are more common than others and may be located in the middle range of other values x. Since measurement results are largely determined by a large number of (unidentifiable) random factors, methods of probability calculus and mathematical statistics are used to assess their uncertainties.

A structure of results can be analysed by dividing a range of all results into a number of intervals and assigning values to each class. The resulting frequency distribution for the corresponding classes can be presented in a graph called a histogram (Fig. 1.3). The graph consists of a series of rectangles placed on an axis of coordinates which are based on intervals of length h (Δx), whereas the height is determined by the frequency (or cardinality) of results belonging to a specific class interval.

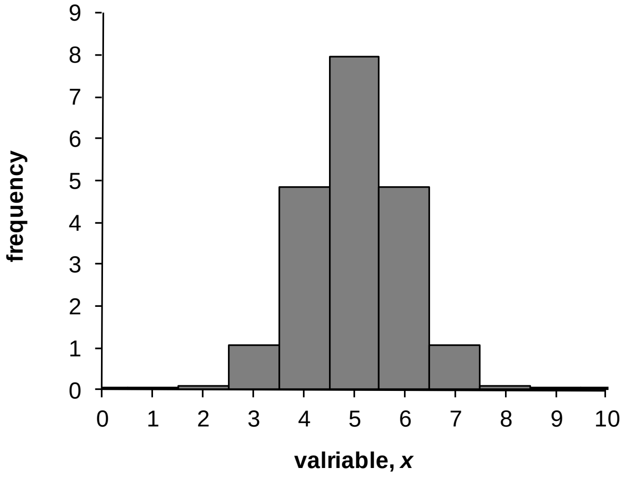


Fig. 1.3. Histogram of the frequency distribution of x_i for each class of data.

If it was possible to repeat the measurements an infinite number of times, then the resulting distribution could be represented as a normal distribution curve for the general population (Fig. 1.4). In statistics, the general population is a set of all possible experiments of a given type.

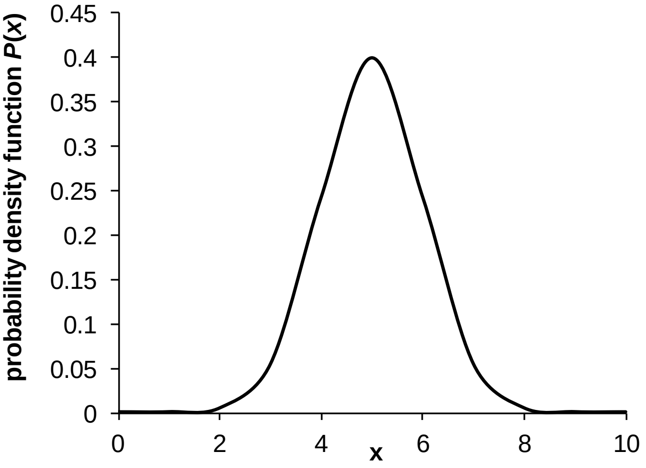


Fig. 1.4. Normal (Gaussian) distribution curve.

A normal (Gaussian) distribution is a continuous distribution with mean x ($\bar{x}$) and variance σ^2, defined for all real x by a probability density function:

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \tag{1.3}$$

The general population can be characterized by such values as:

- population mean of the distribution:

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_i x_i \tag{1.4}$$

- population variance, defined as the mean square of the deviations of the *x* values from the mean (μ):

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \sum_i (x_i - \mu)^2 \tag{1.5}$$

- population standard deviation, which is the square root of the variance:

$$\sigma = \sqrt{\sigma^2} \tag{1.6}$$

A population standard deviation (σ) is the most important measure characterizing dispersion (population coverage measure) and determines an average deviation of a x_i values from a population mean value (μ). An important characteristic feature of a population standard deviation is that for a normal distribution the probabilities that a result is far from a mean value at most by σ, 2σ i 3σ are respectively:

$\mu \pm \sigma$ 68.26%
$\mu \pm 2\sigma$ 95.46%
$\mu \pm 3\sigma$ 99.73%

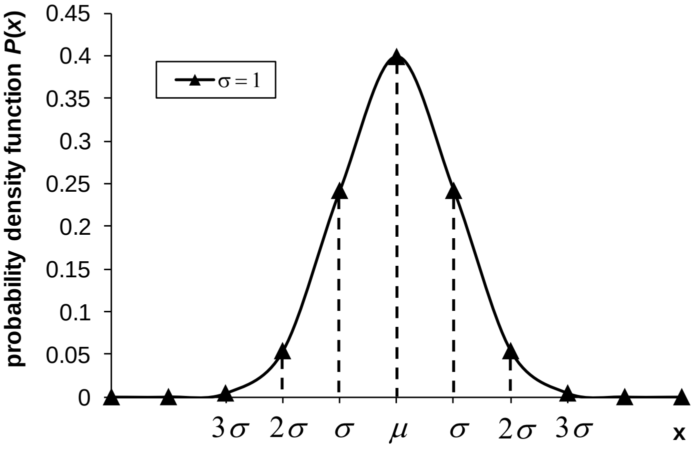


Fig. 1.5. Probability density distribution function (P(x)) of the Gaussian distribution.

The interval $\mu \pm 3\sigma$ means that 99.73% of results will be far from a mean value by no more than 3 standard deviations. This feature of a standard deviation is used in statistical tests (3d, three–sigma rule).

The probability density function (Fig. 1.5) for a normal distribution is symmetric about a population mean (μ), while a population standard deviation value only affects the shape of the distribution (Fig. 1.6).

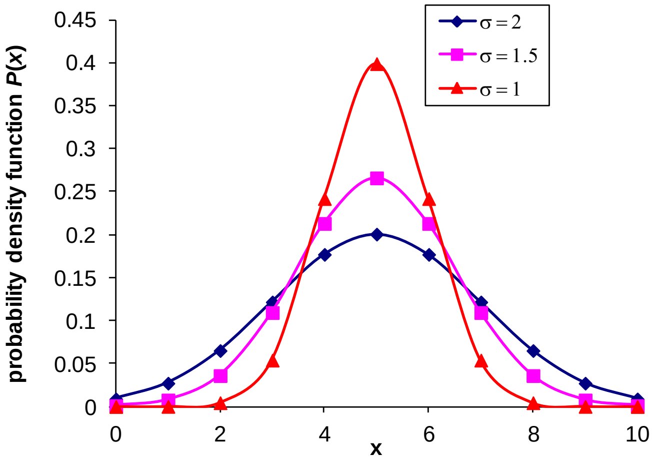


Fig. 1.6. The effect of a population standard deviation value on the shape of a normal distribution.

The density functions presented in Fig. 1.6 indicate that the higher the population standard deviation, the more the results are dispersed about the mean (the distribution curve flattens).

In real measurements, we never have an infinite number of results (general population), what we have is only a random sample. Therefore, based on experimental results, we can obtain only an approximate description of the entire population distribution. Numbers known as statistical parameters (or quantities) are used to describe the structure of results. These parameters can be divided into 4 groups:

a) measures of position – indicate the average or typical level of results, i.e. mean, median, mode (the number that appears most often) etc.,
b) measures of dispersion (variation) – indicate the degree of result dispersion with respect to a mean value (e.g. range, variance, standard deviation, relative standard deviation etc.)
c) measures of asymmetry – indicate the type and degree of deviation from the symmetry of an examined feature (variable) distribution (e.g. skewness)
d) measures of concentration – indicate the concentration of individual observations around a mean (e.g. kurtosis).

In all real experiments, a finite number of measurements (samples, etc.) prevents determination of values μ and σ, and only allows for their estimation using appropriate formulas (estimators). A sample mean, defined as the sum of all values *x* divided by a sample size (*n*), is an estimated mean for a population:

$$\mu \approx \bar{x} = \frac{1}{n}\sum_i x_i \tag{1.7}$$

The variance (s^2) and standard deviation for the sample (*s*) can be calculated from the equations:

$$\sigma^2 \approx s_x^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2 \tag{1.8}$$

$$\sigma \approx s_x = \sqrt{s_x^2} = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2} \tag{1.9}$$

in which $n - 1$ denote the number of degrees of freedom, i.e. the number of independent observations (*x* values), which are used in calculating of *s*.

The standard deviation defined by Equation (1.9) characterizes the mean square error of a single measurement. For the analysis of experimental data, the uncertainty of the final result, i.e. the mean is of greater importance:

$$u(x) = s_{\overline{x}} = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n(n-1)}} \tag{1.10}$$

According to the international standard for the measurement uncertainty analysis (described in more detail in the next chapter), the quantity defined by the Equation (1.10) is called a standard uncertainty. Literature, however, still uses its original name: a standard deviation of a mean. Another concept used and derived from statistics, is a confidence interval for a mean, indicating insufficient number of data sets ($n < 30$). A confidence interval for a mean arithmetic sample is an interval symmetrical with respect to the mean, while the expected value is in it with an assumed probability equal to $1 - \alpha$:

$$P\left(\overline{x} - t_{n-1,\alpha}\frac{s_x}{\sqrt{n}} \le \mu_{\overline{x}} \le \overline{x} + t_{n-1,\alpha}\frac{s_x}{\sqrt{n}}\right) = 1 - \alpha \tag{1.11}$$

In this expression α is the level of significance, and $t_{n-1,\alpha}$ is the Student's distribution parameter (Tab. 1.3.), i.e. a measure of deviations of the distribution of a small number of results from the normal distribution.

Tab. 1.3. Selected values of the Student's distribution parameter (pen name of W. Gosset (1876 – 1937))

Number of degrees of freedom	Confidence level		
	90%	95%	99%
1	6.31	12.7	63.7
2	2.92	4.30	9.92
3	2.35	3.18	5.84
5	2.02	2.57	4.03
7	1.90	2.36	3.50
9	1.83	2.26	3.25

According to the Equation (1.11), the sample mean can be put along with the confidence interval in the following way:

$$\mu_X = \overline{x} \pm t_{n-1,\alpha}\frac{s_x}{\sqrt{n}} \tag{1.12}$$

Expected value = arithmetic mean of the sample ± one half of the confidence interval

The international standard, in force also in Poland, obliges to use an expanded uncertainty (U) which determines (as well as a confidence interval for a mean) an interval around the result of analysis which can be expected (in accordance with the accepted level of significance (probability)) to contain the expected value. An expanded uncertainty is used when the repeatability of measurements is the dominant parameter influencing the estimation of uncertainty. It can be calculated from the formula

$$U = k\frac{s_x}{\sqrt{n}} = k \cdot u(x) \tag{1.13}$$

in which k denote the coverage factor ($2 \le k \le 3$).

The analysis of measurement uncertainty will be discussed more broadly in the next chapter.

EXAMPLE

Ten measurements of the pH of an aqueous solution were conducted, providing the following results:

6.254 6.312 6.277 6.261 6.291 6.330 6.289 6.288 6.326 6.293

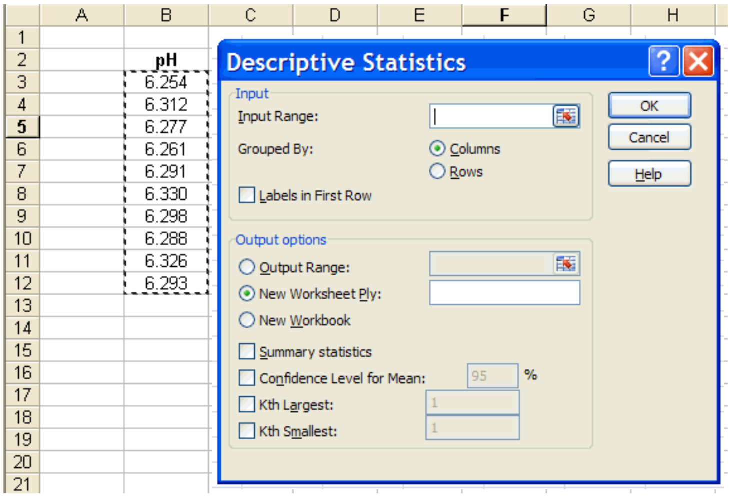
Calculate uncertainty and expanded uncertainty of the mean.

SOLUTION

This task can be solved using appropriate formulas for standard deviation (Eq. (1.9)) and standard uncertainty (Eq. (1.10)) or using standard functions of a calculation spreadsheet, e.g.:

=AVERAGE(range)
=VAR(range) (older versions of Excel)
=VAR.S(range) (Excel 2010)
=STDEV(range) (older versions of Excel)
=STDEV.S(range) (Excel 2010)

Results can be more easily obtained using the *Data analysis* add-in. After selecting from the menu: *Data→Data analysis→Descriptive statistics* (or in older versions of Excel: *Tools→Data analysis→ Descriptive statistics*) and selecting the *Input range*, *Confidence level for the mean (95%)* and *Summary statistics*:



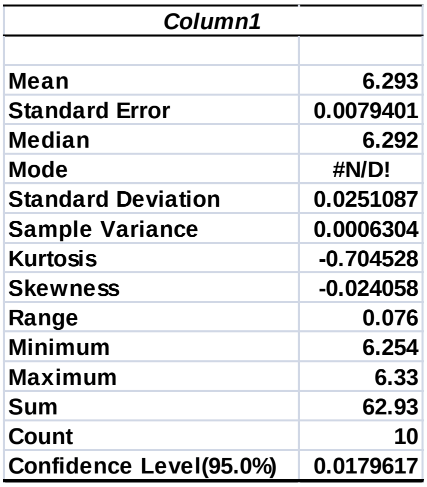
	A	B
1		
2		pH
3		6.254
4		6.312
5		6.277
6		6.261
7		6.291
8		6.330
9		6.298
10		6.288
11		6.326
12		6.293

Descriptive Statistics

Input
Input Range:
Grouped By: ● Columns ○ Rows
□ Labels in First Row

Output options
○ Output Range:
● New Worksheet Ply:
○ New Workbook
□ Summary statistics
□ Confidence Level for Mean: 95 %
□ Kth Largest: 1
□ Kth Smallest: 1

we get the following summary of the analysis:



Column1	
Mean	**6.293**
Standard Error	**0.0079401**
Median	**6.292**
Mode	**#N/D!**
Standard Deviation	**0.0251087**
Sample Variance	**0.0006304**
Kurtosis	**-0.704528**
Skewness	**-0.024058**
Range	**0.076**
Minimum	**6.254**
Maximum	**6.33**
Sum	**62.93**
Count	**10**
Confidence Level(95.0%)	**0.0179617**

This table summarizes the most important statistical parameters. The quantity in Excel called *Standard error* is the standard deviation of a mean, i.e. standard uncertainty, defined by the formula 1.10. *Confidence level (95.0%)* is half the width of the confidence interval (Eq. (1.12)). In order to determine the coefficient of the Student's *t*-distribution occurring in the equation, we can use the functions:

=TINV(α, n–1) (older versions of Excel)
=T.INV.2T(α, n–1) (Excel 2010)

According to the general rule that correctly rounded values of a quantity and its uncertainty have the same number of decimal places, the end result can be put as follows:

pH = 6.293, u(pH) = 0.008 with the standard uncertainty,

pH = (6.293±0.016) for k = 2 with the expanded uncertainty, or

pH = 6.293±0.018 with the confidence interval of the mean (not recommended).

The mean and standard deviation can be calculated by using an alternative recursive method. In this method the first value x_1 is the first trial value of a mean:

$$m_1 = x_1 \tag{1.14}$$

In this case, the initial sum of squared deviations is zero:

$$q_1 = 0 \tag{1.15}$$

In further calculations, recursive formulas for a mean value (m) and the sum of squared deviations (q) are used as follows:

$$m_i = \frac{(i-1)m_{i-1} + x_i}{i} \tag{1.16}$$

$$q_i = q_{i-1} + \frac{(i-1)(x_i - m_{i-1})^2}{i} \tag{1.17}$$

After completing the calculations for all values of i (i = 1, 2, ..n), the final m_i value is the mean of the entire data set (m_n) whereas the standard deviation (s) is computed using the equation:

$$s = \sqrt{\frac{q_n}{n-1}} \tag{1.18}$$

where q_n denote the final value of sum-squared deviations (q_i).

1.3. Uncertainty of measurement

In 1995 a group of international institutions (ISO, BIMP, IEC, IFCC, UIPAC, UIPAP, OMIL, NIST) established international standards of measurement uncertainties. In 1999 such a standard was adopted also in Poland. Legal requirements concerning the analysis of measurement results oblige to follow the recommendations of this standard.

The new standard requires a statistical approach to the uncertainty calculus. In accordance with the accepted principles, a measurement error is a measure of difference between two specific values:

MEASUREMENT ERROR = measured value − true value

For individual measurements the following formulas are used for the absolute error:

$$\varepsilon = \left| x - x_r \right| \tag{1.19}$$

and relative error:

$$\delta = \frac{\varepsilon}{x_r} = \frac{\left| x - x_r \right|}{x_r} \tag{1.20}$$

in which x denote the measured value, while x_r is the true value.

UNCERTAINTY is a parameter associated with a measurement result that characterizes the dispersion of results and can be reasonably attributed to the measured value.

Following the recommendations of the standard, a STANDARD UNCERTAINTY is taken as a measurement uncertainty, and it is calculated as the square root of a variance estimator. The symbol adopted for a standard uncertainty is *u* or *u*(*x*).

An important element of the standard is also a distinction between two ways of assessing uncertainty which are classified into two categories according to a calculation method (type A and type B). Type-A uncertainties characterize random errors and their analysis is based on statistical calculations. Type B uncertainties relate to systematic errors analysed using methods other than statistical calculations.

TYPE A	TYPE B
An analysis based on statistical calculations	Non-statistical methods: - experimenter’s experience, - comparison with previous similar measurements, - manufacturer’s certificate for measuring instruments used (instrument grade), - reference material analysis (references).

Moreover the new standard: makes a distinction between correlated and uncorrelated measurements in indirect measurements, introduces the concept of "expanded uncertainty" and determines how to record measurement results and their uncertainties.

According to the standard, analysed measured values can be divided into two groups:

a) quantities measured in direct measurements (measuring one quantity, e.g. mass, temperature, etc.),

b) quantities measured in indirect measurements (measuring several quantities x_1, x_2, …, and calculating an indirect quantity according to the functional formula $y = f(x_1, x_2, \ldots)$, i.e. density measurement according to the formula $d = m/V$)

The adopted standard also defines how to record uncertainty:

standard uncertainty $m = 0.82$ g, $u(m) = 0.14$ g
expanded uncertainty $m = 0.82$ g, $U(m) = 0.28$ g
$m = (0.82 \pm 0.28)$ g for $k = 2$

In the presented example the uncertainty is given with two significant digits.

1.3.1. Type B evaluation of uncertainty

Type B evaluation of uncertainty is used when we deal with one measurement result or if there is no dispersion in a series of results. A standard uncertainty can be calculated from the corresponding formulas, for example, we can use the following formula to calculate an uncertainty resulting from the accuracy of an instrument (calibration uncertainty):

$$u(x) = \frac{\Delta_d x}{\sqrt{3}} \tag{1.21}$$

where $\Delta_d x$ is the calibration uncertainty, equal to the scale interval of the measuring device used.

Once it can be assumed on the basis of general knowledge that a variable has a triangular distribution, the standard uncertainty is calculated from the formula:

$$u(x) = \frac{\Delta_d x}{\sqrt{6}} \tag{1.22}$$

Another factor affecting a measurement uncertainty is the uncertainty of an experimenter triggered by causes beyond his/her control. In most cases the uncertainty can be calculated from the expression:

$$u(x) = \frac{\Delta_e x}{\sqrt{3}} \tag{1.23}$$

For uncertainties of literature data or values calculated using a calculator (no standard deviation values), the following equation is applied:

$$u(x) = \frac{\Delta_t x}{\sqrt{3}} \tag{1.24}$$

The total standard uncertainty (type B) for a single measurement can be calculated from the formula:

$$u(x) = \sqrt{\frac{(\Delta_d x)^2}{3} + \frac{(\Delta_e x)^2}{3} + \frac{(\Delta_t x)^2}{3}} \tag{1.25}$$

EXAMPLE

Calculate the standard uncertainty of a volume measured by a volumetric flask of 250±0.4ml and calculate the standard uncertainty of mass measurement on an analytical balance ±0.0001g.

SOLUTION

Using the Equation (1.21) for calibration uncertainty, we obtain:

$$u(V) = \frac{\Delta_d V}{\sqrt{3}} = \frac{0.4}{\sqrt{3}} = 0.231\,ml = 0.24\,ml$$

which according to the rules can be put as:

V = 250.00 ml, $u(V)$ = 0.24 ml

and

$$u(m) = \frac{\Delta_d m}{\sqrt{3}} = \frac{0.0001}{\sqrt{3}} = 0.000058\,g$$

1.3.1.1. Propagation of uncertainty

If the measured quantity y is a function of several input (independent) variables $y = f\,(x_1, x_2, \ldots, x_n)$, the *combined standard uncertainty*, in accordance with the rule of error propagation, can be calculated from the formula:

$$u_c(y) = \sqrt{u_1^2\left(\frac{\partial y}{\partial x_1}\right)^2 + u_2^2\left(\frac{\partial y}{\partial x_2}\right)^2 + \ldots + u_n^2\left(\frac{\partial y}{\partial x_n}\right)^2} \tag{1.26}$$

In Equation (1.26) symbol $\left(\frac{\partial y}{\partial x}\right)$ denote the partial derivative of y function with respect to a given variable (x). Uncertainty evaluation therefore requires basic knowledge about the derivatives of functions. In order to explain this issue, the basic information and formulas for calculating derivatives are presented below. The table 1.4 shows some examples of elementary functions and their derivatives.

Tab. 1.4. Some examples of elementary functions and their derivatives

Function $f(x)$	Derivative $f'(x)$	Comments
c	0	constant function
x^n	nx^{n1}	$n \in$ N $x \in$ R
a^x	$a^x \ln a$	a$\in$R$^+ - \{1\}$ $x \in$ R$^+$
e^x	e^x	$x \in$ R
ln x	1/x	$x \in$ R$- \{0\}$
sin x	cos x	
cos x	–sin x	

In order to calculate derivatives of functions that are a combination of elementary functions, the following formulas can be used:

Product of a function and a constant:

$$[c\,f(x)]' = c\,f'(x) \tag{1.27}$$

Sum of the function:

$$[f(x)+g(x)]' = f'(x) + g'(x) \tag{1.28}$$

Product of the function:

$$[f(x)\ g(x)]' = f'(x)\ g'(x) + f(x)\ g'(x) \tag{1.29}$$

Quotient of the function:

$$\left[\frac{f(x)}{g(x)}\right]' = \frac{f'(x)g(x) - f(x)g'(x)}{(f(x))^2} \tag{1.30}$$

In the case of functions of several variables $y = f(x_1, x_2, \ldots, x_n)$, a derivative with respect to one variable, assuming that other variables are constant, is called a *partial derivative*.

EXAMPLE

Calculate the partial derivatives of the function:

$$z = f(x, y) = 4x^2 + y$$

SOLUTION

Using the formulas from table 4 and definition of a partial derivative we get:

$$\frac{\partial z}{\partial x} = 8x \qquad \text{and} \qquad \frac{\partial z}{\partial y} = 1$$

The expression $\frac{\partial z}{\partial x}$ is read as „the partial derivative of z with respect to x”.

EXAMPLE (MODIFIED)
[L. Sobczyk, A. Kisza, K. Gatner, A. Koll, Eksperymentalna chemia fizyczna, PWN, Warszawa 1982, str. 27]

The molecular weight of substance was determined with the Mayer’s method, yielding the following results:

Mass of the substance m = 0.1250 g = 0.125×10^{-3} kg
Volume of displaced air V = 32.18 cm^3 = 32.18×10^{-6} m^3
Air pressure p = 748.2 mm Hg = 99750.0 Pa (minus the saturated vapor pressure)
Temperature T = 298.2 K

Based on the sensitivity of the apparatus, the following maximum errors were determined:
Δm = 0.0005 g = 5.0×10^{-7} kg
ΔV = 0.05 cm^3 = 5.0×10^{-8} m^3
Δp = 1.1 mm Hg = 146.6 Pa
ΔT = 0.1 K

Based on the results, calculate the molecular weight of substance and the combined standard uncertainty.

SOLUTION

Using the Equation (1.21) for the type B uncertainty and omitting the experimenter's uncertainty and the uncertainty of gas constant (R), we get:

$u(m)$ = 0.000289 g
$u(V)$ = 0.0288 cm^3
$u(p)$ = 0.635 mm Hg
$u(T)$ = 0.0577 K

Substituting the data in the formula for the molecular weight, we obtain:

$$M = \frac{m\mathrm{R}T}{pV} = 96.54\text{g/mol}$$

To calculate the combined standard uncertainty, it is necessary to designate partial derivatives with respect to each variable occurring in the equation. After inserting the appropriate values, we get:

$$\frac{\partial M}{\partial m} = \frac{\mathrm{R}T}{pV} = 772.35 \qquad \frac{\partial M}{\partial V} = -\frac{m\mathrm{R}T}{pV^2} = -3.00 \quad \frac{\partial M}{\partial p} = -\frac{m\mathrm{R}T}{p^2V} = -0.01 \quad \frac{\partial M}{\partial T} = \frac{m\mathrm{R}}{pV} = 0.32$$

Now we can use the formula for the combined standard uncertainty which is as follows:

$$u(M) = \sqrt{\left(\frac{\partial M}{\partial m}\right)^2 u^2(m) + \left(\frac{\partial M}{\partial V}\right)^2 u^2(V) + \left(\frac{\partial M}{\partial T}\right)^2 u^2(T) + \left(\frac{\partial M}{\partial p}\right)^2 u^2(p)} = 0.24g$$

The final result can therefore be put in the following way:

M = 96.54 g/mol, $u(M)$ = 0.24 g/mol

1.3.2. Type A evaluation of uncertainty

Type A evaluation of uncertainty involves determining the uncertainty of a measurement result series using statistical analysis. In the case of a simple quantity, derived from direct measurements, the *standard uncertainty* of a mean is calculated as the standard deviation of the mean:

$$u(x) = s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n(n-1)}} \tag{1.31}$$

If the repeatability of measurements is the dominant parameter influencing the evaluation of uncertainty, an *expanded uncertainty*, defining the interval around a measurement result, can be calculated from the equation:

$$U = k\frac{s_x}{\sqrt{n}} = k \cdot u(x) \tag{1.32}$$

in which $s_{\bar{x}}$ denote the standard deviation, n – the number of measurements, while k – denote the coverage factor. The dimensionless coverage factor usually takes values from k=2 (recommended) to k=3, which corresponds to 95 or 99% probability of finding a result in this range.

In the case of experimental studies of simple quantities (x_1, x_2, …, x_n) being part of a complex quantity ($y = f(x_1, x_2, \ldots, x_n)$), as in type B uncertainty analysis, the value of a *combined standard uncertainty*, when variables x are independent, can be determined from the formula:

$$u_c(y) = \sqrt{(u(x_1))^2\left(\frac{\partial y}{\partial x_1}\right)^2 + (u(x_2))^2\left(\frac{\partial y}{\partial x_2}\right)^2 + \ldots + (u(x_n))^2\left(\frac{\partial y}{\partial x_n}\right)^2} \qquad (1.33)$$

EXAMPLE

Determine the concentration of substance A (c_A) prepared by dissolving 1g of A in 1 dm^3 of water. The experiment was repeated five times, giving the corresponding standard uncertainties: $u(m_A) = 0.001\,\text{g}$, and $u(V) = 0.002\,\text{dm}^3$

SOLUTION

The partial derivatives c_A relative to m_A and V are as follows:

$$\frac{\partial[A]}{\partial m_A} = \frac{1}{V} = 1 \text{ and } \frac{\partial[A]}{\partial V_A} = -\frac{m}{V^2} = -1$$

Substituting the calculated values into the equation (1.32) yields:

$$u_c(c_A) = \sqrt{(u(m_A))^2\left(\frac{\partial A}{\partial m_A}\right)^2 + (u(V))^2\left(\frac{\partial A}{\partial V}\right)^2} = \sqrt{(0.001)^2(1)^2 + (0.002)^2(-1)^2} = 0.002236\,\text{g/dm}^3$$

The final result can be presented in the form:

c_A = 1.0000 g/dm^3, $u(c_A)$ = 0.0023 g/dm^3 with the standard uncertainty, or
c_A = 1.0000 g/dm^3, $U(c_A)$ = 0.0046 g/dm^3 with the expanded uncertainty.

EXAMPLE

The refractive index (n) and density (d) of benzene (M = 78.114 g/mol) at a temperature of 25°C were examined to determine the molar refraction according to the formula:

$$R = \frac{n^2 - 1}{n^2 + 2} \times \frac{M}{d}$$

The average results were as follows:

d=0.8737 g/cm^3
n=1.4979

for which corresponding standard uncertainties were calculated:

$u(d)$ = 0.0002 g/cm^3
$u(n)$ = 0.0003

Calculate the molar refraction (R) and the standard and expanded uncertainties for this value.

SOLUTION

Substituting the obtained values in the formula for the molar refraction, we get R = 26.20225 cm^3/mol. In order to calculate the standard uncertainty, we need to have the values of appropriate partial derivatives R with respect to d and n:

$$\frac{\partial R}{\partial d} = -\frac{R}{d} = 29.98 \text{ and } \frac{\partial R}{\partial n} = R\frac{6n}{(n^2-1)(n^2+2)} = 44.61$$

From Equation (1.33), we obtain:

$$u(R) = \sqrt{(u(d))^2\left(\frac{\partial R}{\partial d}\right)^2 + (u(n))^2\left(\frac{\partial R}{\partial n}\right)^2} = \sqrt{(2\cdot 10^{-4})^2\left(29.98\right)^2 + (3\cdot 10^{-4})^2\left(44.61\right)^2} = 0.014 \quad \text{cm}^3/\text{mol}$$

Therefore, we get:

R = 26.202 cm^3/mol, *u*(*R*) = 0.014 cm^3/mol
or *R* = 26.202 cm^3/mol, *U*(*R*) = 0.028 cm^3/mol

According to the rule of error propagation, equations for calculating uncertainties arising from basic arithmetic operations can be derived from the Equation (1.33):

Addition and subtraction – for the function in the form:

$$y = ag_1 + bg_2 \tag{1.34}$$

the partial derivatives are:

$$\frac{\partial y}{\partial g_1} = a \text{ and } \frac{\partial y}{\partial g_2} = b \tag{1.35}$$

and the standard uncertainty can be calculated from the formula:

$$u(y) = \sqrt{a^2(u(g_1))^2 + b^2(u(g_2))^2} \tag{1.36}$$

Multiplication and division – for the function in the form:

$$y = ag_1g_2 \tag{1.37}$$

the partial derivatives are given by:

$$\frac{\partial y}{\partial g_1} = ag_2 \text{ and } \frac{\partial y}{\partial g_2} = ag_1 \tag{1.38}$$

and the expression for the standard uncertainty is defined as:

$$u(y) = \sqrt{a^2g_2^2(u(g_1))^2 + a^2g_1^2(u(g_2))^2} \tag{1.39}$$

or:

$$\frac{u(y)}{y}=\sqrt{\frac{(u(g_1))^2}{g_1^2}+\frac{(u(g_2))^2}{g_2^2}} \tag{1.40}$$

1.3.3. Combined standard uncertainty

If A and B type uncertainties occur simultaneously, based on the known standard uncertainties of direct measurements, a *combined standard uncertainty* is determined according to the equation:

$$u(x)=\sqrt{(u_A(x))^2+(u_B(x))^2}=\sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i-\overline{x})^2+\frac{(\Delta_d x)^2}{3}+\frac{(\Delta_e x)^2}{3}+\frac{(\Delta_t x)^2}{3}} \tag{1.41}$$

where:
u(*x*) – combined standard uncertainty,
$u_A(x)$ – uncertainty calculated from the statistical dispersion of a measurement result series,
$u_B(x)$ – uncertainty calculated otherwise than from the dispersion of results.

2. Linear regression – least square method

In experimental sciences, fitting mathematical equations to measurement results (in the form of numbers) is a routine practice. The aim of this procedure is to:

a) generalize a data set using an appropriate mathematical function with several parameters (coefficients), or
b) make a theoretical model fit (which results from our knowledge) to check a particular hypothesis.

The equation thus determined can be used inter alia for the purpose of:

a) integration (calculating the area under a curve connecting experimental points),
b) interpolation, i.e. predicting values that have not been measured and are within the range of independent variables used to determine the equation parameters,
c) differentiation, and the consequent calculation of the slopes of tangents to a curve to calculate the instantaneous reaction rates, physicochemical partial quantities, etc.,
d) apparatus calibration (chromatograph, refractometer, spectrophotometer, etc.).

The method of least squares is one of the oldest methods for fitting curves to experimental data. It involves minimizing the sum of squared deviations between the observed and calculated from the model values of a dependent variable (*y*). In this case the minimized value is squared deviations (*Q*), as defined by the equation:

$$Q=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{2.1}$$

where *n* denote the number of data points (pairs of $x-y$) being fitted, y_i – the values of the observed dependent variables *y*, $\hat{y}_i$ – the values of the dependent variable calculated on the basis of the fitted equation ($\hat{y}_i = f(x)$). The Equation (2.1) can be put as follows:

$$Q=\sum_{i=1}^{n}(y_i-f(x_i))^2 \tag{2.2}$$

The difference $(y_i-\hat{y}_i)$ can be presented in a graph (Fig. 2.1. a) as a vertical line segment between the observed value and the value calculated from the model (deviation of the *i*-th point from the regression line). *Q* is the sum of all squared deviations (Fig. 2.1. b). According to the method of least squares, the curve is located relative to experimental points so that the value of *Q* is the smallest

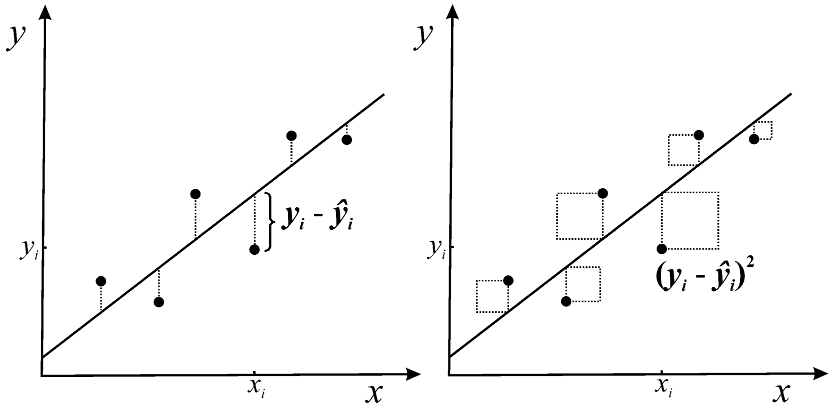


Fig. 2.1. Graphical interpretation of the method of least squares.

For a linear function in the form:

$$\hat{y}_i = f(x_i) = a_0 + a_1 x_i \tag{2.3}$$

equation (2.1) can be rewrite as:

$$Q = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \tag{2.4}$$

As for the linear regression, Q is a function of two regression coefficients, it can be presented in a figure (Fig. 2.2) as a parabolic surface with a minimum for only one pair of a_0 and a_1.

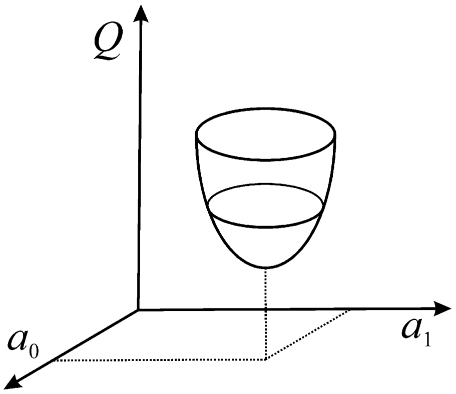


Fig. 2.2. The sum–squared error (Q), as a function of the model parameters a_0 and a_1.

To calculate the values of a_0 and a_1 corresponding to Q_{min}, we can use a standard procedure in which the calculated partial derivatives Q with respect to a_0 and a_1 are compared to zero and then a system of equations relative to these variables is solved. This procedure gives the following results:

$$\frac{\partial Q}{\partial a_0} = 2\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)(-1) = 0 \tag{2.5}$$

Divide through by –2 we get:

$$\frac{\partial Q}{\partial a_0} = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) = 0 \tag{2.6}$$

After multiplication the above expression takes the form:

$$\frac{\partial Q}{\partial a_0} = \sum_{i=1}^{n} y_i - na_0 - a_1 \sum_{i=1}^{n} x_i = 0 \tag{2.7}$$

In order to obtain an equation for a_0 for calculating the intercept (Fig. 2.3), the Equation (2.7) is multiplied by *n* to give the formula:

$$\frac{\sum y_i}{n} - a_0 - a_1 \frac{\sum x_i}{n} = 0 \tag{2.8}$$

which can be put as follows:

$$a_0 = \bar{y} - a_1 \bar{x} \tag{2.9}$$

In this equation $\bar{y}$ and $\bar{x}$ denotes the mean value of the dependent and independent variable, respectively.

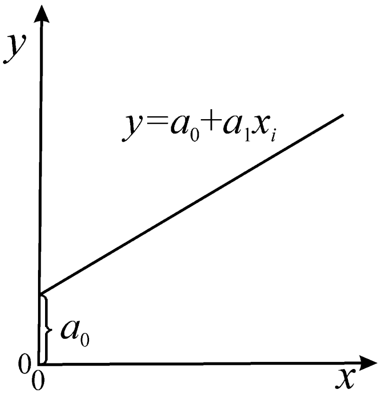


Fig. 2.3. Graphical interpretation of the a_0 coefficient (intercept).

Similar calculation for a_1, yields:

$$\frac{\partial Q}{\partial a_1} = 2\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)(-x_i) = 0 \tag{2.10}$$

and:

$$\frac{\partial Q}{\partial a_1} = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) x_i = 0 \tag{2.11}$$

After multiplication:

$$\sum_{i=1}^{n} y_i x_i - a_0 \sum_{i=1}^{n} x_i - a_1 \sum_{i=1}^{n} x_i^2 = 0 \tag{2.12}$$

After inserting equation (2.9) into the above expression, we obtain the following equation:

$$a_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{2.13}$$

which allows determination of the value a_1, called a slope. Graphical interpretation of the slope is presented in Fig. 2.4.

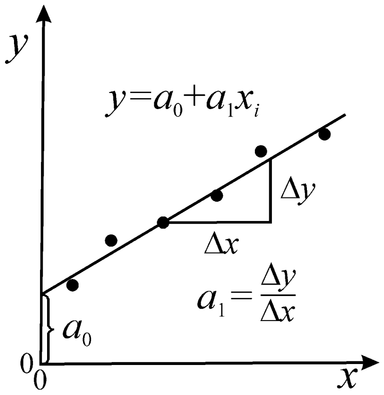


Fig. 2.4. Graphical interpretation of the a_1 (slope) and a_0 (intercept) coefficients.

The Equation (2.13) can also be put in the following, frequently encountered form:

$$a_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{2.14}$$

The Equation (2.13) is considerably simplified in the case of regression analysis with no a_0 (intercept) coefficient (if $a_0 = 0$ a line passing through the center of the coordinate system). The general equation takes the following form:

$$\hat{y}_i = f(x_i) = a_1 x_1 \tag{2.15}$$

and a_1 coefficient can be calculated from:

$$a_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \tag{2.16}$$

Graphical interpretation of the linear regression (without intercept) is presented in Fig. 2.5.

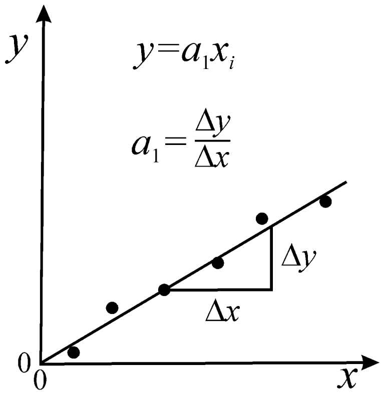


Fig. 2.5. Graphical interpretation of the a_1 coefficient (slope) without intercept (a_0 = 0).

EXAMPLE:

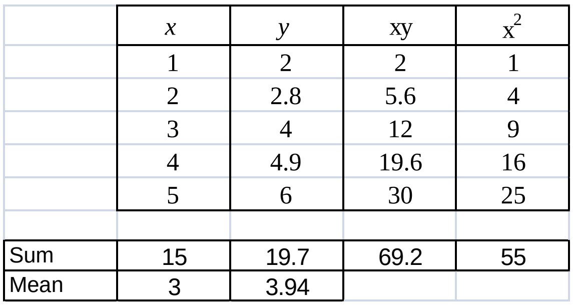
For the following results:

x	*y*
1	2
2	2.8
3	4
4	4.9
5	6

determine the linear equation (a_1 and a_0 coefficients) using method of least squares.

SOLUTION

To determine a_0 and a_1 coefficients from the equations (2.9) and (2.13), it is necessary to make simple calculations of corresponding means and sums which can be computed in a calculation spreadsheet or using a calculator:



	x	*y*	xy	x^2
	1	2	2	1
	2	2.8	5.6	4
	3	4	12	9
	4	4.9	19.6	16
	5	6	30	25
Sum	15	19.7	69.2	55
Mean	3	3.94		

Substituting the calculated values into the equation (2.9) and (2.13) we get:

$$a_1 = \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{5\cdot 69.2 - 15\cdot 19.7}{5\cdot 55 - 15^2} = 1.01$$

and

$$a_0 = \bar{y} - a_1\bar{x} = 3.94 - 1.01\cdot 3 = 0.91$$

The regression parameters (coefficients) in Excel ca be calculated from worksheet functions:

=SLOPE(known_ys; known_xs) (a_1 coefficient) and
=INTERCEPT(known_ys;known_xs) (a_0 coefficient).

2.1.Weighted linear regression

In the regression equations made so far, all values y_i were assumed to be encumbered with an identical error. This assumption is usually not true for real experimental data, because values y_i are subject to various errors. A proper analysis requires the use of a weighted least squares method and the application of an appropriate statistical weights during the calculations.

According to this method, the general equation for Q (sum–squared error, Eq. (2.2)) is given by:

$$Q = \sum_{i=1}^{n} w_i (y_i - f(x_i))^2 \tag{2.17}$$

Thus, by analyzing the simplest case of a weighted linear regression, the above equation can be presented as follows:

$$Q = \sum_{i=1}^{n} w_i (y_i - a_0 - a_1 x_i)^2 \tag{2.18}$$

where the weighting factors w_i (statistical weight) corresponds to the i–th point. If the point (x_i, y_i) was determined with greater accuracy, the fitting curve should move closer to this point and hence the value w_i should be higher. If w_i = 1 for all values i, Equation (2.18) is reduced to Equation (2.4) and the weighting coefficients are called absolute weights. For different experimental data, numerical values of weights (w_i) can be determined in various ways, i.e. as the inverse of a dependent variable:

$$w_i = \frac{1}{y_i} \tag{2.20}$$

or in the most common way, as the inverse of a variance for each value y_i:

$$w_i = \frac{1}{s_i^2} \tag{2.21}$$

The a_0 and a_1 coefficients n the weighted linear regression can be calculated by the following equations:

$$a_0 = \bar{y}_w - a_1 \bar{x}_w \tag{2.22}$$

$$a_1 = \frac{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i y_i - \sum_{i=1}^{n} w_i x_i \sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{i=1}^{n} w_i x_i \right)^2} \tag{2.23}$$

where the values of $\bar{y}_w$ and $\bar{x}_w$ (weighted centroids) are defined by equations:

$$\bar{y}_w = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \tag{2.24}$$

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \tag{2.25}$$

The goodness of fit of the linear function to experimental data can be assessed by calculating the average deviation from the regression line (residual standard deviation, standard error of estimate):

$$s_y = \sqrt{\frac{Q}{n-2}} \tag{2.26}$$

which is a measure of the accuracy of predictions based on the regression equation and determines the standard deviation of all points around the regression. In this equation *Q* is defined by Equation (2.18), while *n*–2 corresponds to the number of degrees of freedom.

Standard deviations for each of the regression coefficients (a_0 and a_1) can be calculated from the equations:

$$s_{a_0} = \sqrt{\frac{s_y^2}{M} \sum_{i=1}^{n} w_i x_i^2} \tag{2.27}$$

$$s_{a_1} = \sqrt{\frac{s_y^2}{M} \sum_{i=1}^{n} w_i} \tag{2.28}$$

where:

$$M = \sum_{i=1}^{n} w_i \sum_{i=1}^{n} w_i x_i^2 - \left(\sum_{i=1}^{n} w_i x_i \right)^2 \tag{2.29}$$

In practical calculations, due to the fact that s_i^2 is a function of a number of measurements, weights are redefined as follows:

$$w_i' = \frac{s_i^{-2}}{\sum_{i=1}^{n} s_i^{-2} / n} \tag{2.30}$$

therefore Equations (2.22) – (2.25) take the following form:

$$a_0 = \bar{y}_w - b\bar{x}_w \tag{2.31}$$

$$a_1 = \frac{\sum_{i=1}^{n} w_i' x_i y_i - n\bar{x}_w \bar{y}_w}{\sum_{i=1}^{n} w_i' x_i^2 - n\bar{x}_w^2} \tag{2.32}$$

and

$$\bar{x}_w = \sum_{i=1}^{n} w_i' x_i / n \tag{2.33}$$

$$\bar{y}_w = \sum_{i=1}^{n} w_i' y_i \, / \, n \tag{2.34}$$

EXAMPLE

For six standard solutions with a concentration of C (mol/dm^3), absorbance measurements (A) were made and the corresponding standard deviations (s_i) were determined:

C	A	s_i
1.0E-06	0.02	0.009
1.0E-05	0.22	0.02
2.0E-05	0.49	0.013
3.0E-05	0.62	0.046
4.0E-05	0.78	0.051
5.0E-05	1.152	0.011

Determine the weighted linear regression equation and compare it with the standard linear regression. Based on the equations, determine the concentration of an unknown sample (A = 1.1) and calculate the relative error.

SOLUTION

The application of the equations (2.31) and (2.32) in the calculations requires an initial calculation of respective sums and means. The results are compared below:

C	**A**	s_i	$1/(s_i^2)$	w'_i	$w'_i x_i$	$w'_i y_i$	$w'_i x_i y_i$	$w_i x_i^2$
1.0E-06	0.02	0.009	12346	2.48	2.48E-06	0.050	4.96E-08	2.48E-12
1.0E-05	0.22	0.02	2500	0.50	5.02E-06	0.110	1.10E-06	5.02E-11
2.0E-05	0.49	0.013	5917	1.19	2.38E-05	0.582	1.16E-05	4.75E-10
3.0E-05	0.62	0.046	473	0.09	2.85E-06	0.059	1.76E-06	8.54E-11
4.0E-05	0.78	0.051	384	0.08	3.09E-06	0.060	2.41E-06	1.24E-10
5.0E-05	1.152	0.011	8264	1.66	8.30E-05	1.912	9.56E-05	4.15E-09
n		sum	29884.4	6	1.202E-04	2.773	1.125E-04	4.88E-09
6		Sum/n	4980.73					
					mean of y $_w$	***mean of x*** $_w$		
					0.462	2.00E-05		

For the weighted linear regression, substituting the appropriate values (from Table) to the equations we get:

$$a_1 = \frac{\sum_{i=1}^{n} w_i' x_i y_i - n\bar{x}_w \bar{y}_w}{\sum_{i=1}^{n} w_i' x_i^2 - n\bar{x}_w^2} = \frac{1.13 \cdot 10^{-4} - 6 \cdot 0.462 \cdot 2 \cdot 10^{-5}}{4.9 \cdot 10^{-9} - 6 \cdot (2 \cdot 10^{-5})^2} = 23024$$

$$a_0 = \bar{y}_w - b\bar{x}_w = 0.462 - 230242 \cdot 2 \cdot 10^{-5} = 1.52 \cdot 10^{-3}$$

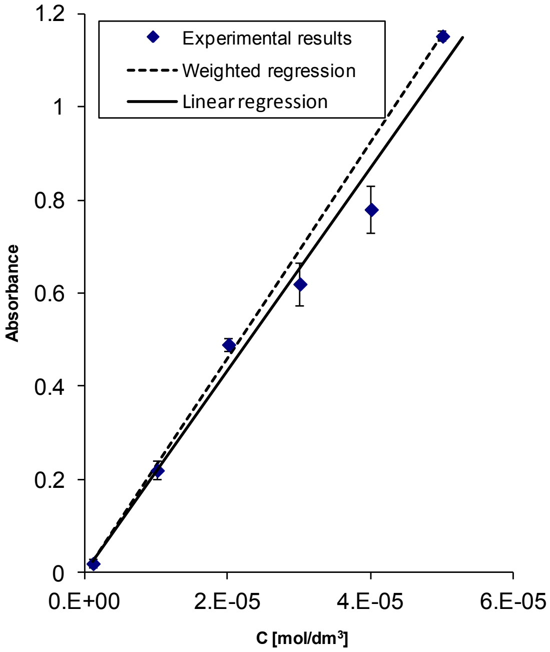
Thus the weighted linear equation can be write as follows:

A = 23024·C + 1.52×10^{-3}

For the linear regression, without taking into account weights, and considering the mean values A as dependent variables, we obtain:

$$A = 21650 \cdot C + 2.14 \times 10^{-3}$$

The results obtained from both equations are compared in Figure which also indicates the standard deviations of individual experimental points.



Substituting A = 1.1 in both equations, we get as follows:

for the weighted linear regression: $C_{(w)} = 4.77 \cdot 10^{-5}$ mol/dm^3

for the standard linear regression: $C = 5.07 \cdot 10^{-5}$ mol/dm^3

Taking the value calculated from the weighted linear regression equation as the exact value, the relative error is 6.3%.

In order to calculate standard deviations for the weighted regression coefficients, it is necessary to use Equations (2.27) and (2.28). If $w_i = 1$ (absolute weights), these equations get simplified:

$$s_{a_0} = s_y \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n \sum_{i=1}^{n} (x_i - \overline{x})^2}} \tag{2.35}$$

$$s_{a_1} = s_y \sqrt{\frac{1}{\sum_{i=1}^{n} (x_i - \overline{x})^2}} \tag{2.36}$$

To assess the fit of the regression function, a linear correlation coefficient (Pearson's *r*) can be applied, defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \qquad (2.37)$$

which is a measure of the strength of the linear relationship between variables *x* and *y*. The linear correlation coefficient (*r*) takes values between –1, +1. If *r* = 1, then the points lie exactly on a straight line not parallel to the axis *x*. If there is no linear relationship between the variables, *r* = 0, and variables *x* and *y* are uncorrelated. Examples of the correlation coefficient values are presented in Fig. 2.6.

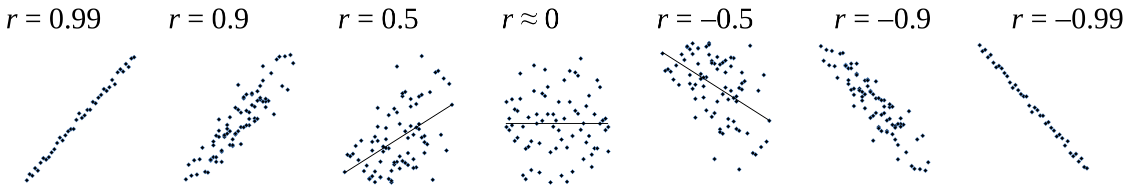


Fig. 2.6. Examples of the correlation coefficient values.

A more adequate measure of a model's goodness of fit to the observed (experimental) values is a squared correlation coefficient called *a coefficient of determination*. It sets out what proportion of the total variation (or what %) of variable *y* is explained by the linear regression model

EXAMPLE

Determine for the following set of data:

x	*y*
1	2
2	2.8
3	4
4	4.9
5	6

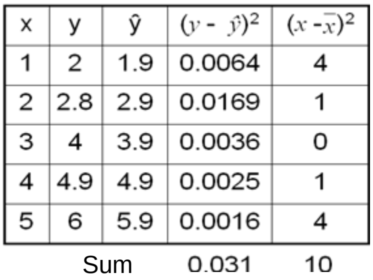
standard deviation of slope and intercept.

SOLUTION

The linear regression equation takes the following form:

$$y = 0.91 + 1.01 \cdot x$$

According to Equations (2.35) and (2.36), the average deviation from the regression line (s_y) must be first calculated with the Equation (2.26) to determine the s_{a_0} and s_{a_1} values. For this purpose it is necessary to determine the values of *y* estimated from the model ($\hat{y}_i$), the corresponding sum of squared deviations ($Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$) and the sum $\sum_{i=1}^{n}(x_i - \overline{x})^2$. The calculations are presented in the table below:



x	y	ŷ	$(y - \hat{y})^2$	$(x - \bar{x})^2$
1	2	1.9	0.0064	4
2	2.8	2.9	0.0169	1
3	4	3.9	0.0036	0
4	4.9	4.9	0.0025	1
5	6	5.9	0.0016	4
	Sum		0.031	10

Substituting the appropriate values into the equations, we obtain:

$$s_y = \sqrt{\frac{0.031}{5-2}} = 0.1016$$

and then:

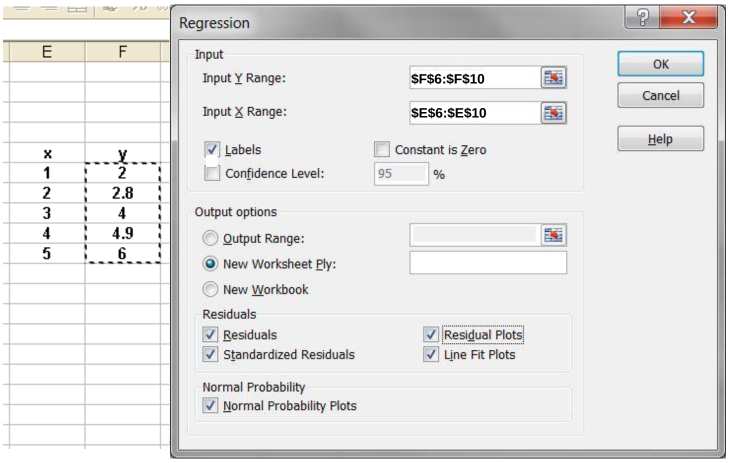
$$s_{a_1} = s_y \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 0.1016 \cdot \sqrt{\frac{1}{10}} = 0.0321 \approx 0.033$$

$$s_{a_0} = s_y \sqrt{\frac{\sum_{i=1}^{n} x_i^{\,2}}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 0.1016 \cdot \sqrt{\frac{55}{5 \cdot 10}} = 0.107 \approx 0.11$$

The calculated values of the coefficients and the standard deviations can be put as:

$$a_1 = 1.010 \pm 0.033$$
$$a_0 = 0.91 \pm 0.11$$

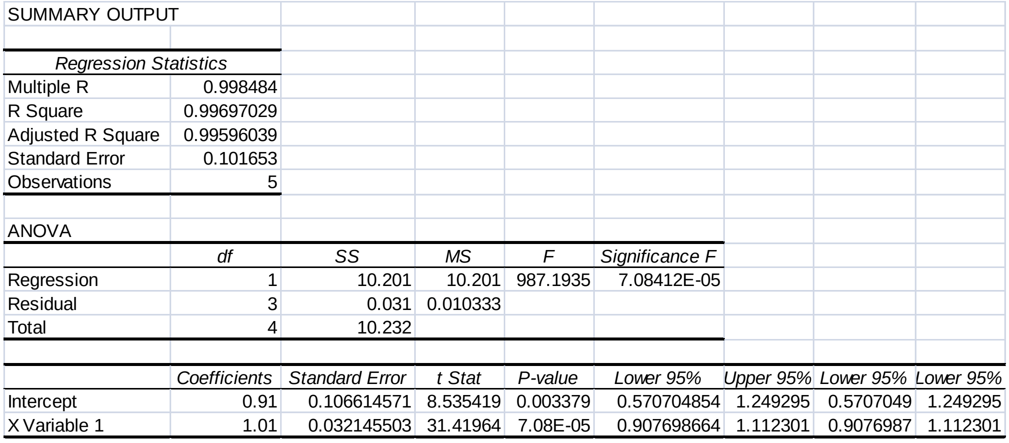
The same values can be obtained by using *Data analysis→Regression* in Excel spreadsheet. After selecting an input range *y* (1 column) and *x* (1 column):



x	y
1	2
2	2.8
3	4
4	4.9
5	6

```
Regression
Input
Input Y Range: $F$6:$F$10
Input X Range: $E$6:$E$10
☑ Labels          □ Constant is Zero
□ Confidence Level: 95 %
Output options
○ Output Range:
● New Worksheet Ply:
○ New Workbook
Residuals
☑ Residuals                ☑ Residual Plots
☑ Standardized Residuals   ☑ Line Fit Plots
Normal Probability
☑ Normal Probability Plots
```

a summary in the form of a table is displayed:



SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.998484							
R Square	0.99697029							
Adjusted R Square	0.99596039							
Standard Error	0.101653							
Observations	5							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	10.201	10.201	987.1935	7.08412E-05			
Residual	3	0.031	0.010333					
Total	4	10.232						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95%*	*Lower 95%*
Intercept	0.91	0.106614571	8.535419	0.003379	0.570704854	1.249295	0.5707049	1.249295
X Variable 1	1.01	0.032145503	31.41964	7.08E-05	0.907698664	1.112301	0.9076987	1.112301

where individual values can be read:

The coefficient of determination (R Square) indicating that 99.69% of the total variation of y is explained by the linear regression model.

a) mean deviation from the regression line (Standard error) $s_y = 0.1016 \approx 0.11$
b) standard deviation of the coefficient a_0 (Standard Error (Intercept)) s_{a_0} = $0.1066 \approx 0.11$
c) standard deviation of the coefficient a_1 (Standard Error (X1)) s_{a_1} = $0.0321 \approx 0.033$

According to the standard deviation of the coefficients, the regression equation can be put as follows:

$$y = 0.91(\pm 0.11) + 1.010(\pm 0.033) \cdot x$$
$$r^2 = 0.9969,\ s_y = 0.11$$

The regression equation calculated with the least squares method can be used to predict the values of y ($y_0 = f(x_0)$) for any values x_0 (point prediction of y). The standard error of prediction of the so obtained result can be calculated from Equation:

$$s_{y_0} = s_y \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}} \tag{2.38}$$

In this equation, expression $(x_0 - \bar{x})^2$ indicates that the farther the value (x_0) for which we make prediction from the sample mean, the less accurate the prediction.

In many cases (e.g. for large values n) the value of the root in the above equation is approximately equal to 1, so this formula gets simplified to:

$$s_{y_0} \approx s_y \tag{2.39}$$

EXAMPLE

Calculate the standard error of prediction of the value of y for $x_0 = 3.5$, using the regression equation and the data from the previous example.

SOLUTION

Substituting the value of $x_0 = 3.5$ in the regression equation ($y = 0.91 + 1.01 \cdot x$), the result we obtain is $y_0 = 4.445$. From Equation (2.38), we get:

$$s_{y_0} = 0.1016\sqrt{1+\frac{1}{5}+\frac{(3.5-3)^2}{10}} = 0.1124 \approx 0.12$$

According to the standard error of prediction, the final result can be put as follows:

$$y_0 = 4.44 \pm 0.12$$

Using the simplified equation (2.39), the result of calculations can be presented as:

$$y_0 = 4.44 \pm 0.11$$

slightly different from the previous result.

If the regression equation is used to predict the value x_0 for any value y_0 (point prediction of x), the standard deviation of the so designated number can be calculated from the equation:

$$s_{x_0} = \frac{s_y}{|a_1|}\sqrt{1+\frac{1}{n}+\frac{(y_0-\bar{y})^2}{a_1^2\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}} \tag{2.40}$$

This expression, like Equation (2.38), can be simplified, when the element is approximately equal to 1, then:

$$s_{x_0} \approx \frac{s_y}{|a_1|} \tag{2.41}$$

EXAMPLE

Calculate the standard error of prediction of the value of x_0 for $y_0 = 2.5$ using the regression equation and the data from the previous example.

SOLUTION

After substituting $y_0 = 2.5$ in the transformed regression equation ($x = (y - 0.91)/1.01$), we get $x_0 = 1.574$, and from Equation (2.40):

$$s_{x_0} = \frac{0.1016}{1.01}\sqrt{1+\frac{1}{5}+\frac{(2.5-3.94)^2}{(1.01)^2(10)^2}} = 0.1191 \approx 0.12$$

According to the standard error of prediction, the final result can be write as follows:

$$y_0 = 4.44 \pm 0.12$$

In the case of simplified equation (2.41), $s_{x_0} \approx 0.11$.

2.2. Analysis of residuals

The analysis of residuals is the primary method of detecting defects in fitting a model to experimental data. The residual for the i–th value of y_i is defined by:

$$e_i = y_i - \hat{y}_i \tag{2.42}$$

where y_i denote the observed (experimental) value of the dependent variable, and $\hat{y}_i$ – the value calculated using the fitting equation.

In a properly chosen model, residuals should exhibit a normal distribution and be randomly dispersed around a regression function. The distribution of residuals is normally estimated on the basis of the graph *e* with respect to an independent variable. Typical examples of correct and incorrect distributions of residuals are shown in Fig. 2.7.

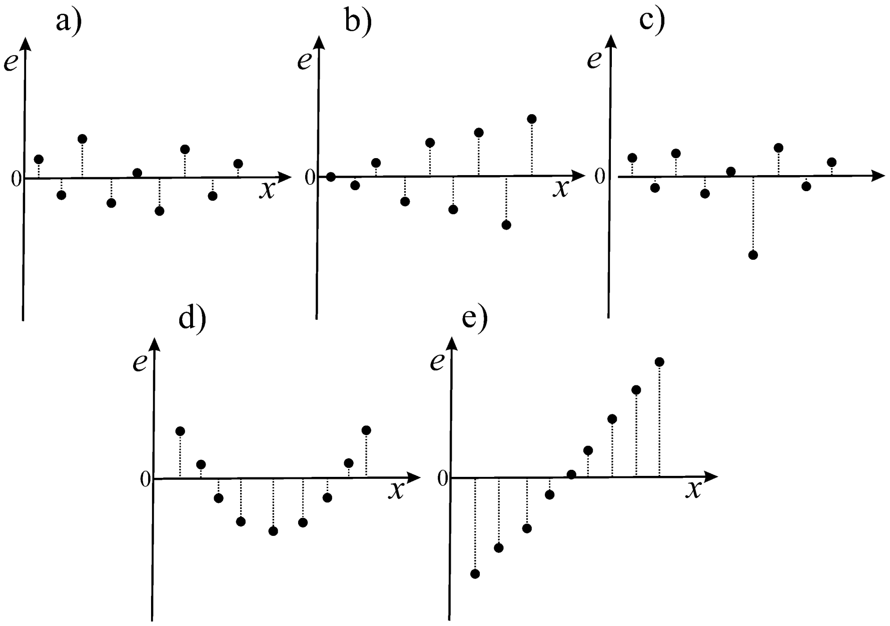


Fig. 2.7. Examples of correct (a–c) and incorrect (d–e) distributions of residuals.

The distribution of residuals shown in Figure 2.7a is correct (i.e. random) and shows no significant differences in the dispersion of results around the regression line. In Figure 2.7b, the increase in residuals with the increase of the independent variable may indicate the need to take account of these errors in the regression analysis and apply the weighted regression. The distribution of residuals shown in Figure 2.7c is correct from a theoretical point of view, however, it indicates the presence of an outlier, clearly deviating from the observed trend. When the model used in the calculation is incorrect, the distribution of residuals is inconsistent with the theoretical properties of *e* (Fig. 2.7d, e).

The equation (2.42) can take an expanded form:

$$e_i = (y_i - \hat{y}_i) = (y_i - \bar{y}) - (\hat{y}_i - \bar{y}) \tag{2.43}$$

The expression for the observed value deviation from its mean $(y_i - \bar{y})$ can therefore be written as follows:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \tag{2.44}$$

In this equation the first term $(\hat{y}_i - \bar{y})$ is a part of the total deviation of the variable *y* which was explained by the linear regression of *y* on *x*, the second term $(y_i - \hat{y}_i)$ is a part of the total variation which was not explained by regression (Fig. 2.8)

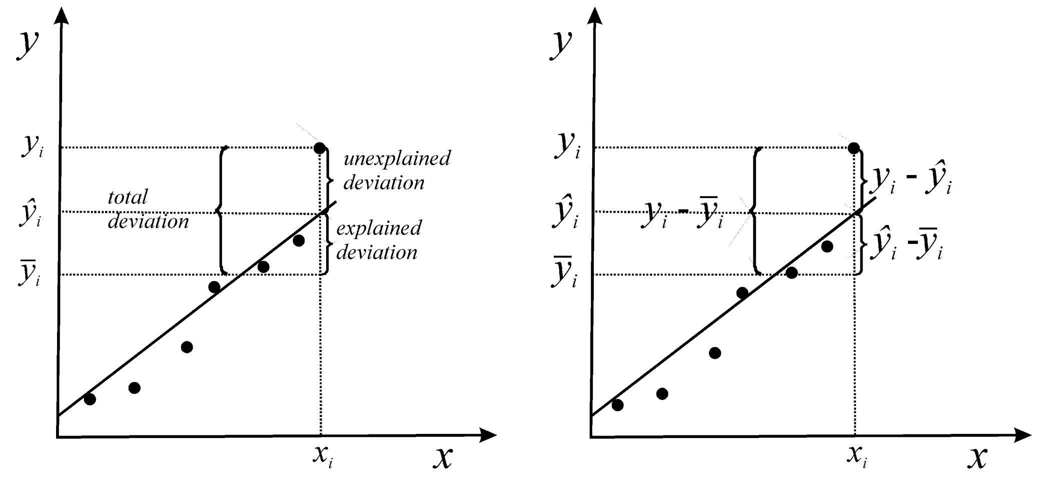


Fig. 2.8. Partitioning of variance in the ordinary least squares method.

By squaring Equation (2.44) and summing all values of *i*, we get:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum(y_i - \hat{y}_i)^2 \tag{2.45}$$

The above equation can be put in a simpler form:

$$Q_1 = Q_2 + Q_3 \tag{2.46}$$

where Q_1 denote the total sum of squares about the mean(the total variation), Q_2 – sum of squares explained by the regression model (the variation explained by the regression equation) and Q_3 – residual (error) sum for squares (the variation in *y* that cannot be explained by the regression equation). The source variation are compared in Tab. 2.1.

Tab. 2.1. The source of variation in the linear regression analysis

Source of variation	Degrees of freedom	Sum of squares
Total (Q_1)	$n - 1$	$\sum(y_i - \bar{y})^2$
Due to regression (Q_2)	1	$\sum(\hat{y}_i - \bar{y})^2$
About regression (Q_3)	$n - 2$	$\sum(y_i - \hat{y}_i)^2$

The measure of the estimated model's goodness of fit to the empirical data is the coefficient of determination, calculated as the ratio of Q_2 (explained variation) to Q_1 (total variation) from the equation:

$$r^2 = \frac{Q_2}{Q_1} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \tag{2.47}$$

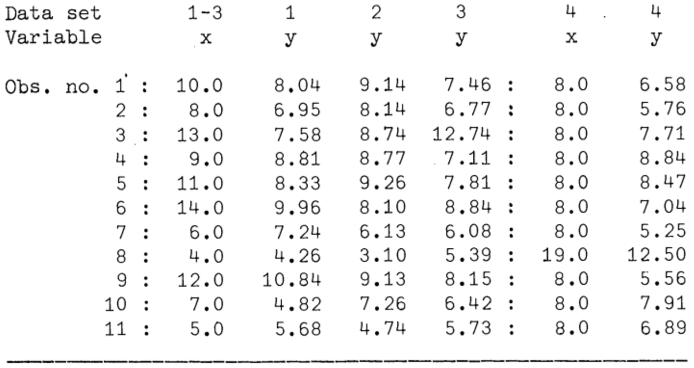
A determination coefficient determines what fraction (or %) of overall response variability is explained by the model.

The average deviation from the regression line (residual standard deviation, standard error of estimation) which measures the accuracy of approximation, can be determined from the equation:

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{Q_3}{n-2}} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n-2}} \tag{2.48}$$

The results of statistical calculations, such as s_y and r^2, should be interpreted with great caution, because it can lead to erroneous conclusions. An example is the following set of data [F.J. Anscombe, Graphs in statistical analysis, *Amer. Stat.*, 27 (1973) 17–21] which leads to the same regression equation:

$$y = 3 + 0.5 \cdot x$$
$$r^2 = 0.67,\ r = 0.818 \text{ and } s_y = 1.24$$



Data set	1-3	1	2	3	4	4
Variable	x	y	y	y	x	y
Obs. no. 1 :	10.0	8.04	9.14	7.46 :	8.0	6.58
2 :	8.0	6.95	8.14	6.77 :	8.0	5.76
3 :	13.0	7.58	8.74	12.74 :	8.0	7.71
4 :	9.0	8.81	8.77	7.11 :	8.0	8.84
5 :	11.0	8.33	9.26	7.81 :	8.0	8.47
6 :	14.0	9.96	8.10	8.84 :	8.0	7.04
7 :	6.0	7.24	6.13	6.08 :	8.0	5.25
8 :	4.0	4.26	3.10	5.39 :	19.0	12.50
9 :	12.0	10.84	9.13	8.15 :	8.0	5.56
10 :	7.0	4.82	7.26	6.42 :	8.0	7.91
11 :	5.0	5.68	4.74	5.73 :	8.0	6.89

Particular sets of data together with the regression line are shown in Fig. 2.9 a–d.

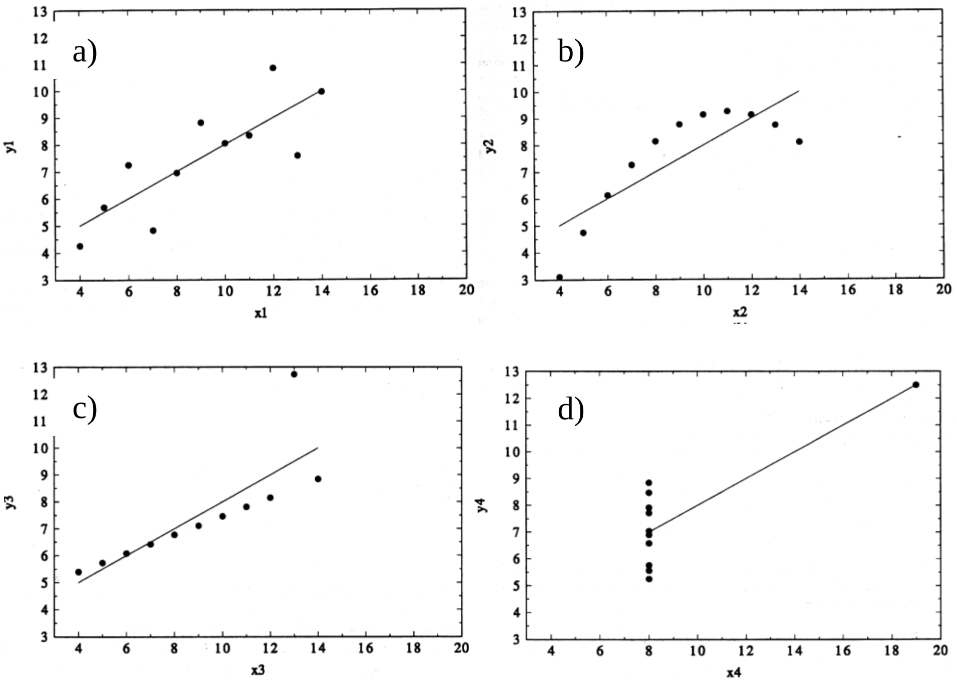


Fig. 2.9. Scatterplots of Anscombe's quartet.

For the data presented in Fig. 2.9 only in *a*) the linear model is adequate to the description of the experimental results, and the low value of r^2 is caused by large dispersion of experimental points. The case *b*) represents a different character of dependencies for which for example a second-degree polynomial model would be more appropriate. In other examples, the false conclusion about linear dependencies results from a random error in determining the values of one *y* (Fig. 2.9c)or the incorrectly chosen range of the independent variable *x* in the analysis of regression (Fig. 2.9.d).

2.3. Linearizing transformations

Experimental sciences often use non-linear equations which, after appropriate transformation of variables, can be put in a linear form (Fig. 2.10).

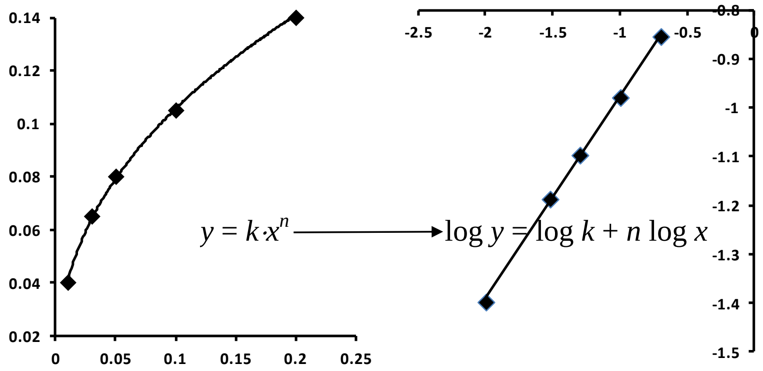


Fig. 2.10. Example of a non-linear function ($y = k{\cdot}x^n$) and its linearized form.

For the exponential function, presented in Fig. 2.10, the linearization can be obtained by taking the log of both sides of the equation:

$$y = k \cdot x^n \tag{2.49}$$

obtaining, after simple modifications, the following relationship:

$$\log y = \log k + n \log x \tag{2.50}$$

Denoting further:

$$Y^* = \log y \quad \text{and} \quad X^* = \log x \tag{2.51}$$

we obtain, a linear equation:

$$Y^* = a_0 + a_1 X^* \tag{2.52}$$

where:

$$a_0 = \log k \ \text{ and } \ a_1 = n \tag{2.53}$$

With the values of a_0 and a_1 evaluated using the method of least squares, we can calculate the values of k and n from the equations (2.53). In the case of the standard deviations of coefficients, we should remember about their appropriate transformation. In the analyzed dependency, we can write only for the coefficient n, that:

$$S_n = S_{a_1} \tag{2.54}$$

In order to determine the standard deviation of the coefficient k, we should use the equation (1.33) describing error propagation in the course of mathematical operations. Therefore, we obtain:

$$S_k = \sqrt{S_{a_0}^2 \left(\frac{\partial k}{\partial a_0}\right)^2} \tag{2.55}$$

The value of the partial derivative of the function $k = 10^{a_0}$ is:

$$\left(\frac{\partial k}{\partial a_0}\right) = 10^{a_0} \ln(10) \tag{2.56}$$

After substituting to the equation (2.55), the final equation takes the following form:

$$S_k = S_{a_0} \cdot 10^{a_0} \cdot \ln(10) \tag{2.57}$$

Typical non–linear functions and the appropriate linearizing substitutions are shown in Table 2.2.

Tab. 2.2. Selected non–linear functions and the linearizing substitutions

Non–linear functions	Linearizing substitutions
$y = a + \frac{b}{x}$	$Y^* = y$ $X^* = 1/x$
$\frac{1}{y} = a + b \cdot x$	$Y^* = 1/y$ $X^* = x$
$y = a \cdot b^x$	$Y^* = \log(y)$ $X^* = x$
$y = a \cdot x^b$	$Y^* = \log(y)$ $X^* = \log(x)$
$y = a \cdot e^x$	$Y^* = \ln(y)$ $X^* = x$
$y = a + b \cdot x^n$	$Y^* = y$ $X^* = x^n$
$y = \frac{a \cdot x}{b + x}$	$Y^* = x/y$ or $Y^* = 1/y$ $X^* = x$ $X^* = 1/x$

3. Non-linear regression – polynomial equation fitting

Linear regression is the simplest case of fitting a polynomial function to data, which can be put in the general form as:

$$f(x_i) = \sum_{j=0}^{m} a_j x_i^j \tag{3.1}$$

This equation is a general expression for the polynomial *m*, which for *m* = 1 corresponds to a simple linear regression ($f(x) = a_0 + a_1x$). Similarly to the line fitting case, the best fit is obtained for those values of the coefficients a_j for which the value of *Q* is minimal:

$$Q = \sum_i \left(y_i - \sum_j a_j x_i^j \right)^2 \tag{3.2}$$

To calculate the optimal values of a_j, the function (3.2) is differentiated with respect to each parameter (a_k) and equated to zero:

$$\frac{\partial Q}{\partial a_k} = \sum_i \left(y_i - \sum_j a_j x_i^j \right) x_i^k = 0 \tag{3.3}$$

The resulting system of m+1 equations with m+1 unknowns makes it possible to determine appropriate values of the coefficients a_j. Limiting considerations to the second–degree polynomial in the form:

$$f(x_i) = a_0 + a_1 x_i + a_2 x_i^2 \tag{3.4}$$

and using the equation (3.3), we get the following expression:

$$\sum_i \left(y_i - a_0 - a_1 x_i - a_2 x_i^2 \right) x_i^k = 0 \qquad k = 0, 1, 2 \tag{3.5}$$

After expansion, this formula can be put in the form of the following system of equations:

$$n \cdot a_0 + a_1 \sum_i x_i + a_2 \sum_i x_i^2 = \sum_i y_i \tag{3.6}$$

$$a_0 \sum_i x_i + a_1 \sum_i x_i^2 + a_2 \sum_i x_i^3 = \sum_i x_i y_i \tag{3.7}$$

$$a_0 \sum_i x_i^2 + a_1 \sum_i x_i^3 + a_2 \sum_i x_i^4 = \sum_i x_i^2 y_i \tag{3.8}$$

The equations (3.6) – (3.8) can be write in the form of matrix equation **AB = C,** where:

$$\mathbf{B} = \begin{pmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 \end{pmatrix} \mathbf{A} = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} \mathbf{C} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \end{pmatrix} \tag{3.9}$$

In order to solve this matrix equation (evaluation of the a_0, a_1, a_2 coefficients) we should use the equation:

$$\mathbf{A} = \mathbf{B}^{-1}\mathbf{C} = \mathbf{D}\ \mathbf{C} \tag{3.10}$$

in which $\mathbf{B}^{-1}$=**D** denotes the inverse matrix of **B**.
The average deviation from the regression line can be calculated from the equation:

$$s_{y=} \sqrt{s_y^2} = \sqrt{\frac{Q}{n-p-1}} \tag{3.11}$$

where p is the degree of the polynomial.
Standard deviations of individual coefficients a can be determined from the following equations:

$$s_{a_0} = \sqrt{d_{11} s_y^2} \quad s_{a_1} = \sqrt{d_{22} s_y^2} \quad s_{a_2} = \sqrt{d_{33} s_y^2} \tag{3.12}$$

where s_y^2 denotes the residual variance, and the d_{kk} values are the elements from the main diagonal of the inverse matrix **D**.

EXAMPLE

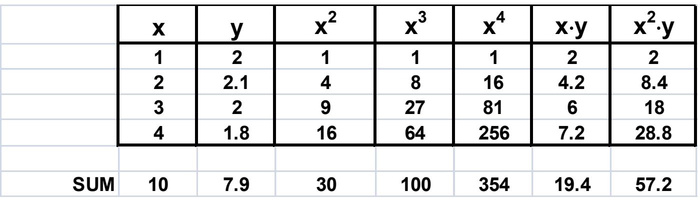
Determine for the following set of data:

x	*y*
1	2
2	2.1
3	2
4	1.8

the values of the second–degree polynomial coefficients and their standard deviations.

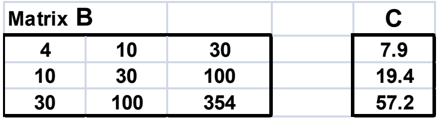
SOLUTION

In order to generate matrix **B** (Equation (3.9)), it is necessary to make calculations summarized in the following table:



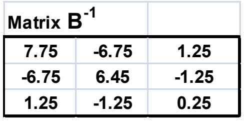
	x	y	x^2	x^3	x^4	x·y	x^2·y
	1	2	1	1	1	2	2
	2	2.1	4	8	16	4.2	8.4
	3	2	9	27	81	6	18
	4	1.8	16	64	256	7.2	28.8
SUM	10	7.9	30	100	354	19.4	57.2

According to the Equation (3.9) matrix **B** and matrix **C** can be written as:



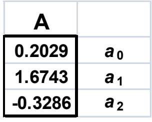
Matrix B				C
4	10	30		7.9
10	30	100		19.4
30	100	354		57.2

After calculating the inverse matrix of the matrix **B**, we get:



Matrix B^{-1}		
7.75	-6.75	1.25
-6.75	6.45	-1.25
1.25	-1.25	0.25

Multiplying the inverse matrix by a matrix **C**, we obtain the regression coefficients values (matrix **A)**:



A	
0.2029	a_0
1.6743	a_1
-0.3286	a_2

Standard deviations of the individual coefficients *a* can be determined from Equation (3.12), using the value of s_y:

$$s_y = \sqrt{s_y^2} = \sqrt{\frac{Q}{n-p-1}}$$

where *Q* is defined by the equation:

$$Q = Q_3 = \sum \left(y_i - \hat{y}_i \right)^2$$

and the elements from the main diagonal of the inverse matrix **D**.
Calculations for n = 4 and p = 2, lead to the following results:

$$s_y = 0.02236,\ \ s_y^2 = 0.0005$$

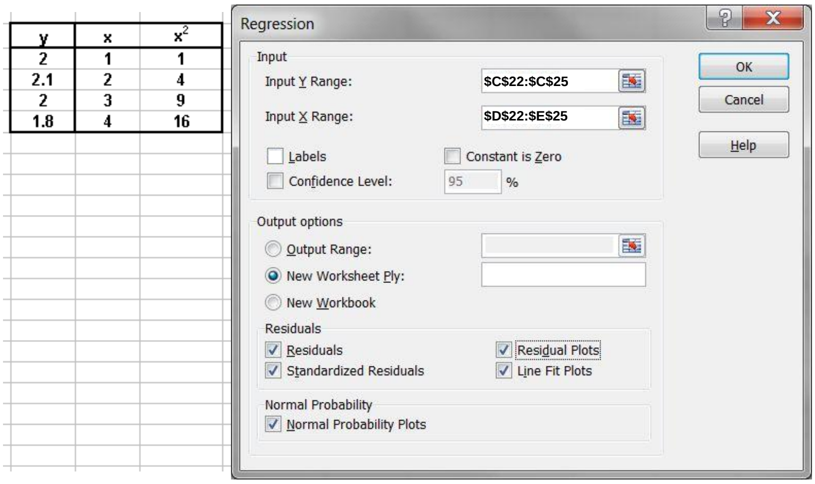
and:

$$s_{a_0} = \sqrt{7.75 \cdot 0.0005} \approx 0.063$$

$$s_{a_1} = \sqrt{6.45 \cdot 0.0005} \approx 0.057$$

$$s_{a_2} = \sqrt{0.25 \cdot 0.0005} \approx 0.012$$

Similar calculations can be carried out in the Excel spreadsheet using the add-in *Data Analysis→Regression*, after selecting the input range y (1 column) and x (2 columns):



y	x	x²
2	1	1
2.1	2	4
2	3	9
1.8	4	16

Regression

Input

Input Y Range: $C$22:$C$25

Input X Range: $D$22:$E$25

□ Labels □ Constant is Zero

□ Confidence Level: 95 %

Output options

○ Output Range:

● New Worksheet Ply:

○ New Workbook

Residuals

☑ Residuals ☑ Residual Plots

☑ Standardized Residuals ☑ Line Fit Plots

Normal Probability

☑ Normal Probability Plots

we can obtain a summary in the following form:

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.99472292					
R Square	0.98947368					
Adjusted R Square	0.96842105					
Standard Error	0.02236068					
Observations	4					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	0.047	0.0235	47	0.102597835	
Residual	1	0.0005	0.0005			
Total	3	0.0475				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	1.775	0.062249498	28.51429	0.022317	0.984045134	2.56595487
X Variable 1	0.305	0.056789083	5.370751	0.117193	-0.416573721	1.02657372
X Variable 2	-0.075	0.01118034	-6.7082	0.094208	-0.217059688	0.06705969

The final fit equation with standard deviations of individual coefficients can be put in the form:

$$y = 1.775(\pm 0.063) + 0.305(\pm 0.057)\cdot x - 0.075\ (\pm 0.012)\cdot x^2$$
$$r^2 = 98.95\% \qquad s_y = 0.023$$

4. Multiple linear regression analysis

In the more general multiple regression, *n* observations correspond to *p* independent variables (x_1, x_2, ..., x_p) and *n* values of dependent variables (y_i). In the simplest case, the model is reduced to a multiple linear regression and takes the form:

$$\hat{y}_i = a_0 + a_1 x_{1,i} + a_2 x_{2,i} + ... a_p x_{p,i} \tag{4.1}$$

where a_1, a_2, ..., a_p denotes the regression coefficients, and $\hat{y}_i$ is the predicted value of the *i*-th dependent variable.
If the variable *y* depends only on two independent variables *x*, the regression analysis leads to plane fitting, as shown in Fig. 4.1.

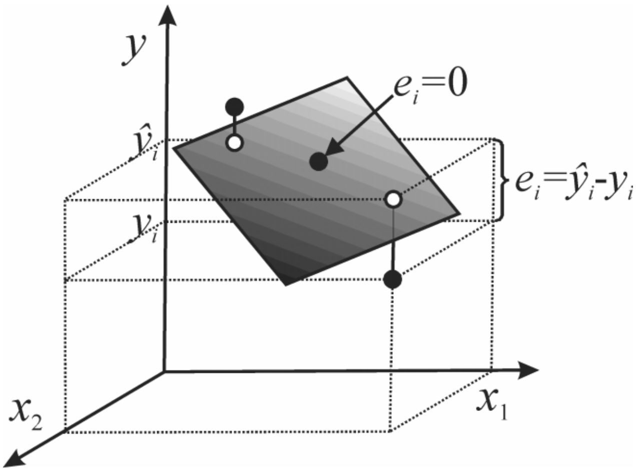


Fig. 4.1. Multiple linear regression for two independent variables x_1 and x_2.

4.1.Regression coefficients

Similarly to the method of the ordinary least squares (chapter 2), the best estimated regression coefficients (*a*) are those that lead to the minimum value of *Q* ($Q = \sum_{i=1}^{n} e^2$), and are a solution to the following system of equations:

$$\begin{aligned} &S_{1,1}a_1 + S_{1,2}a_2 + \cdots S_{1,p}a_p = S_{y,1} \\ &S_{2,1}a_1 + S_{2,2}a_2 + \cdots S_{2,p}a_p = S_{y,2} \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &S_{p,1}a_1 + S_{p,2}a_2 + \cdots S_{p,p} = S_{y,p} \end{aligned} \tag{4.2}$$

the sums of $S_{i,j}$ can be presented using the following expression:

$$S_{i,j} = \sum_{k=1}^{n} \left(x_{i,k} - \bar{x}_i\right)\left(x_{j,k} - \bar{x}_j\right) \quad i, j = 1, 2 \cdots p \tag{4.3}$$

$$S_{y,i} = \sum_{k=1}^{n} \left(y_k - \bar{y}\right)\left(x_{i,k} - \bar{x}_i\right) \quad i = 1, 2 \cdots p \tag{4.4}$$

where $\bar{x}_i$ and $\bar{y}_i$ are given by:

$$\bar{x}_i = \frac{1}{n}\sum_{k=1}^{n} x_{i,k} \qquad \text{and} \qquad \bar{y}_i = \frac{1}{n}\sum_{k=1}^{n} y_k \tag{4.5}$$

The values of regression coefficients *a* can be evaluated by solving the matrix equation:

$$\boldsymbol{a} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\,\mathbf{X}^{\mathbf{T}}\mathbf{Y} = \mathbf{B}^{-1}\mathbf{C} = \mathbf{D}\ \mathbf{C} \tag{4.6}$$

where:

$$\mathbf{X} = \begin{pmatrix} x_{0,1} & x_{1,1} & \cdots & x_{p,1} \\ x_{0,2} & x_{1,2} & \cdots & x_{p,2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{0,n} & x_{1,n} & \cdots & x_{p,n} \end{pmatrix} \qquad \mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \ \boldsymbol{a} = \begin{pmatrix} a_0 \\ a_1 \\ \cdots \\ a_p \end{pmatrix} \tag{4.7}$$

For thus calculated regression coefficients (a_i) we can calculate standard deviations according to the equation:

$$s_{a_j} = \sqrt{s_y^2 d_{j,j}} \tag{4.8}$$

In this equation $d_{j,j}$ is the elements from the main diagonal of the inverse matrix $(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$, called a dispersion matrix, while s_y^2 is a residual variance defined by:

$$s_y^2 = \frac{Q}{n-p-1} \tag{4.9}$$

where *n* denotes the number of observations (pairs *x*-*y*), *p* – number of independent variables (x_p), *Q* – the sum of squared residuals.

Multiple regression, like a simple linear regression, can be analysed through the sources of variation, i.e. variation caused by the regression model (Q_2) and the variation caused by random factors (error, Q_3). Sums of squares for these sources of variation and for the total variation (Q_1) with an appropriate number of degrees of freedom are summarized in Table 4.1.

Tab. 4.1. The source of variation in the multiple linear regression

Source of variation	Degrees of freedom	Sum of squares
Total (Q_1) Q_1=Q_2+Q_3	*n*-1	$Q_1 = \sum\left(y_i - \bar{y}\right)^2$
Due to regression (Q_2)	*p*	$Q_2 = \sum\left(\hat{y}_i - \bar{y}\right)^2$
About regression (Q_3≡Q)	*n*-*p*-1	$Q_3 = \sum\left(y_i - \hat{y}_i\right)^2$

The variance for each source of variation can be estimated by dividing the sums of squares by an adequate number of degrees of freedom. The measure of the goodness of fit (s_y) can be calculated from the equation:

$$s_y = \sqrt{\frac{\sum(y_i - \hat{y}_i)}{n-p-1}} = \sqrt{\frac{Q_3}{n-p-1}} \tag{4.10}$$

The measure of the model's goodness of fit to the experimental data (coefficient of determination) can be determined from the equation:

$$r^2 = 1 - \frac{Q_3}{Q_1} = \frac{Q_2}{Q_1} = \frac{\sum\left(\hat{y}_i - \bar{y}\right)^2}{\sum\left(y_i - \bar{y}\right)^2} \tag{4.11}$$

This coefficient determines what fraction of the total variation is explained by the adopted regression model.

4.2. Selecting variables – stepwise procedures

Considering the full (maximum) model in the following form:

$$\hat{y}_i = a_0 + a_1 x_{1,i} + a_2 x_{2,i} + \ldots a_p x_{p,i} \tag{4.12}$$

where $\hat{y}_i$ is the value of y_i estimated by the full model with p parameters, we should consider whether it is possible to generate as good a reduced model:

$$\hat{y}_i^* = a_0^* + a_1^* x_{1,i} + a_2^* x_{2,i} + \ldots a_q^* x_{q,i} \tag{4.13}$$

where $\hat{y}_i^*$ is the value of y_i estimated by the reduced model with q parameters ($q < p$).

In the case of the full model which took into account all the independent variables, we may find that the influence of some of them is negligible. To assess what variation a single variable brings to the general model, we can compare adequate sums of squared deviations for the full model:

$$Q = \sum\left(y_i - \hat{y}_i\right)^2 \tag{4.14}$$

and for the reduced model:

$$Q^* = \sum\left(y_i - \hat{y}_i^*\right)^2 \tag{4.15}$$

and calculate the difference between these values referred to as an extra sum of squares. The final evaluation of the models (the significance of additional variables $p - q$) can be carried out using the equation:

$$F = \frac{\left(Q^* - Q\right)\left(n - p - 1\right)}{\left(p - q\right)Q} \tag{4.16}$$

The value of F can be calculated for the full model and subsequently reduced models. The procedure consisting in tabulating F with respect to q is a standard method for models testing.

In order to find the best model, we can use the so called stepwise regression method using a equation (4.16). Stepwise regression includes two methods for selecting variables: progressive (forward) selection and backward selection. In the forward selection procedure, selected variables sequentially added to a model are retained or discarded, depending on an adopted criterion. In the case of backward selection, the analysis starts from a full model which is gradually reduced by one variable

EXAMPLE

Determine for the following set of data:

Independent variables			Dependent variable
x_1	x_2	x_3	y
0.071	28.79	–10.72	0.425
0.107	26.49	-10.24	0.779
0.15	26.4	-10.7	0.937
0.217	26.76	-10.15	0.646
0.295	26.78	-11.27	1.01
0.338	29.03	-11.34	0.485
0.361	26.42	-10.69	0.853
0.488	26.57	-11.67	1.144
0.538	27.13	-10.24	0.41
0.597	25.91	-11.08	1.015
0.636	26.72	-10.6	0.637
0.718	28.44	-11.03	0.349
0.746	28.84	-10.24	-0.073
0.823	26.95	-11.36	0.769
0.838	27.47	-10.77	0.415
0.852	26.42	-11	0.744
0.972	26.74	-11.15	0.656
1.052	26.46	-10.69	0.518
1.044	27.72	-11.65	0.595
1.133	27.76	-10.2	0.012

determine the best-fit regression model.

SOLUTION

First of all we should check whether models with just one independent variable are sufficient to describe this relationship. After using the add-in *Data analysis→Regression*, the following results of determination coefficients were obtained for particular equations:

x_1 i y	$r^2 = 0.16$
x_2 i y	$r^2 = 0.53$
x_3 i y	$r^2 = 0.26$

Since a satisfactory result was obtained in none of the cases, in the next step the possibility of data description by means of two independent variables was examined, yielding the following results:

model $y = f(x_1,x_2)$: $y = -\,0.36 \cdot x_1 - 0.2418 \cdot x_2 + 7.4$

$r = 0.82$ $r^2 = 0.672$ $s_y = 0.19$

model $y = f(x_2,x_3)$: $y = -\,0.242 \cdot x_2 - 0.323 \cdot x_3 + 3.7$

$r = 0.88$ $r^2 = 0.774$ $s_y = 0.16$

model $y = f(x_1,x_3)$: $y = -\,0.47 \cdot x_1 - 0.39 \cdot x_3 - 3.3$

$r = 0.71$ $r^2 = 0.504$ $s_y = 0.24$

Only considering all three independent variables leads to a model with the highest value of r^2 and lowest value of s_y:

model $y = f(x_1,x_2,x_3)$ $\qquad$ $y = -\,0.451\cdot x_1 - 0.237\cdot x_2 - 0.379\cdot x_3 + 3.23$

$r = 0.999$ $\; r^2 = 0.998$ $\; s_y = 0.012$

Analogous results can be obtained by solving Equation (4.6). For this purpose, we need to define matrices **X** and **Y**:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & \cdots & x_{p(1)} \\ 1 & x_{1(2)} & x_{2(2)} & \cdots & x_{p(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1(n)} & x_{2(n)} & \cdots & x_{p(n)} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_{(1)} \\ y_{(2)} \\ \vdots \\ y_{(n)} \end{bmatrix}$$

and solve the matrix equation $\boldsymbol{a} = (\mathbf{X}^{\mathbf{T}}\,\mathbf{X})^{-1}\,\mathbf{X}^{\mathbf{T}}\mathbf{Y}$:

$\mathbf{X^TX}$					**$\mathbf{X^TY}$**	
20.00	11.98	543.80	-216.79		12.33	
11.98	9.31	325.78	-130.35		6.58	
543.80	325.78	14802.53	-5894.36		331.07	
-216.79	-130.35	-5894.36	2354.27		-135.06	
$\mathbf{(X^TX)^{-1}}$					**a**	
73.2950	0.4929	-1.6675	2.6017		3.2282	a_0
0.4929	0.4831	-0.0051	0.0594		-0.4514	a_1
-1.6675	-0.0051	0.0603	-0.0029		-0.2372	a_2
2.6017	0.0594	-0.0029	0.2361		-0.3789	a_3

The sums of squared deviations can be calculated from the formulas summarized in Table 4.2, or from the corresponding matrix equations:

$$Q_2 = \sum(\hat{y}_i - \bar{y})^2 = \mathbf{a^TX^TY} - n\cdot\bar{y}^2$$

$$Q_3 = \sum(y_i - \hat{y}_i)^2 = \mathbf{Y^TY} - \boldsymbol{a}^{\mathbf{T}}\mathbf{X^TY}$$

$$Q_1 = \sum(y_i - \bar{y})^2 = \mathbf{Y^TY} - n\cdot\bar{y}^2$$

Standard deviations of individual coefficients can be calculated as the square root of the diagonal matrix values $\mathbf{(X^TX)^{-1}}$ multiplied by the residual variance (equation (4.9)):

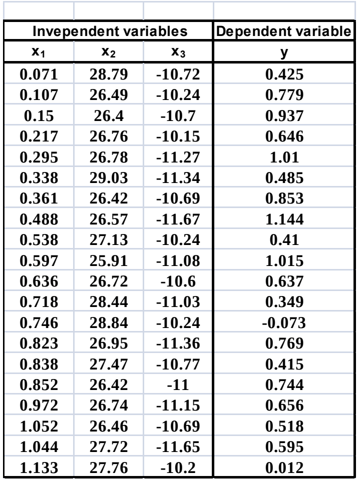
$(\mathbf{X^TX})^{-1}s_y^2 =$

0.010517	7.07E-05	-0.00024	0.000373
7.07E-05	**6.93E-05**	-7.3E-07	8.52E-06
-0.00024	-7.3E-07	**8.65E-06**	-4.1E-07
0.000373	8.52E-06	-4.1E-07	**3.39E-05**

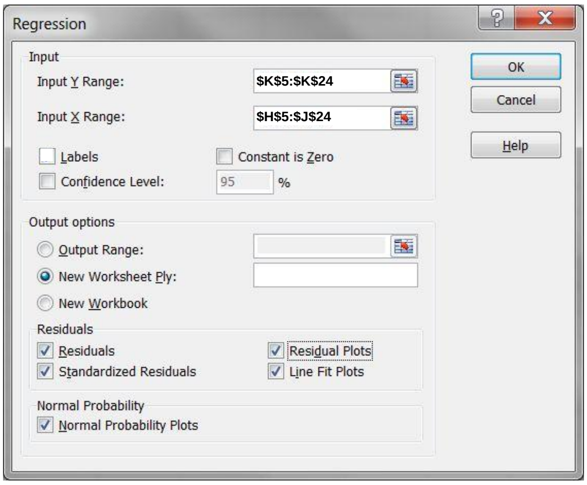
$$s_{a_0} = \sqrt{0.010517} = 0.1026 \qquad s_{a_1} = \sqrt{6.93\cdot 10^{-5}} = 0.00832$$

$$s_{a_2} = \sqrt{8.65\cdot 10^{-6}} = 0.00294 \qquad s_{a_3} = \sqrt{3.39\cdot 10^{-5}} = 0.00582$$

The calculations can also be made using the add-in *Data analysis*→Regression, after selecting the input range *y* (1 column) and *x* (3 columns):



Invependent variables			Dependent variable
x_1	x_2	x_3	y
0.071	28.79	-10.72	0.425
0.107	26.49	-10.24	0.779
0.15	26.4	-10.7	0.937
0.217	26.76	-10.15	0.646
0.295	26.78	-11.27	1.01
0.338	29.03	-11.34	0.485
0.361	26.42	-10.69	0.853
0.488	26.57	-11.67	1.144
0.538	27.13	-10.24	0.41
0.597	25.91	-11.08	1.015
0.636	26.72	-10.6	0.637
0.718	28.44	-11.03	0.349
0.746	28.84	-10.24	-0.073
0.823	26.95	-11.36	0.769
0.838	27.47	-10.77	0.415
0.852	26.42	-11	0.744
0.972	26.74	-11.15	0.656
1.052	26.46	-10.69	0.518
1.044	27.72	-11.65	0.595
1.133	27.76	-10.2	0.012



Regression

Input

Input Y Range: $K$5:$K$24

Input X Range: $H$5:$J$24

□ Labels □ Constant is Zero

□ Confidence Level: 95 %

Output options

○ Output Range:

● New Worksheet Ply:

○ New Workbook

Residuals

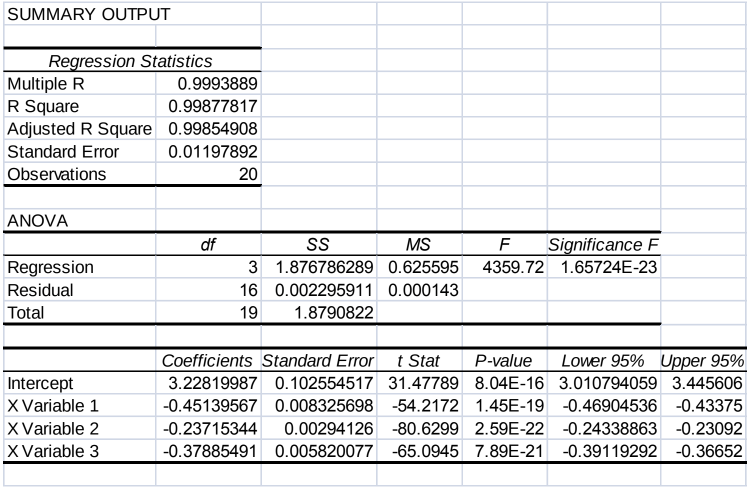
☑ Residuals ☑ Residual Plots

☑ Standardized Residuals ☑ Line Fit Plots

Normal Probability

☑ Normal Probability Plots

they lead to the following summary:



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9993889
R Square	0.99877817
Adjusted R Square	0.99854908
Standard Error	0.01197892
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.876786289	0.625595	4359.72	1.65724E-23
Residual	16	0.002295911	0.000143		
Total	19	1.8790822			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	3.22819987	0.102554517	31.47789	8.04E-16	3.010794059	3.445606
X Variable 1	-0.45139567	0.008325698	-54.2172	1.45E-19	-0.46904536	-0.43375
X Variable 2	-0.23715344	0.00294126	-80.6299	2.59E-22	-0.24338863	-0.23092
X Variable 3	-0.37885491	0.005820077	-65.0945	7.89E-21	-0.39119292	-0.36652

The established regression equation can be put in accordance with standard deviations:

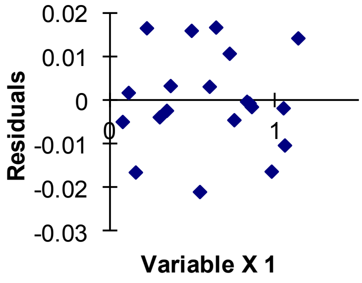
$$y = -\,0.4514(\pm 0.0084)\cdot x_1 - 0.2372(\pm 0.0030)\cdot x_2 - 0.3788(\pm 0.0059)\cdot x_3 + 3.23(\pm 0.11)$$

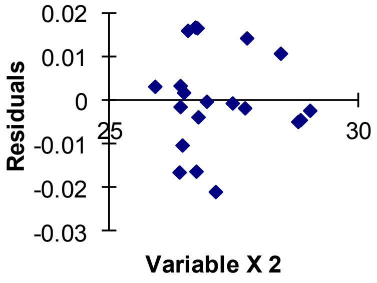
$$r^2 = 99.88\% \quad s_y \; = 0.12$$

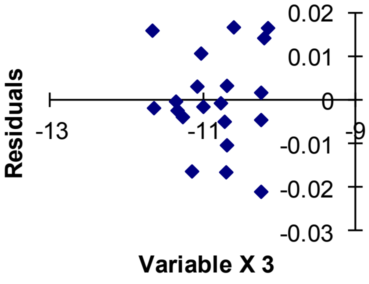
or confidence intervals:

$$y = -\,0.451(\pm 0.018)\cdot x_1 - 0.2372(\pm 0.0063)\cdot x_2 - 0.379(\pm 0.013)\cdot x_3 + 3.23(\pm 0.22)$$

In regression analysis we should also remember to check the distribution of residuals. Below are graphs of the distribution of residuals for each variable, which show random dispersion and are correct.







5. Numerical integration
5.1. Integral and its geometric interpretation

According to the definition, if $y = F(x)$ is a function of x, whose derivative is equal to $F'(x) = f(x)$, then the indefinite integral (antiderivative) of $f(x)$ function, with respect to x, can be written as:

$$\int F'(x)dx = F(x) + C \tag{5.1}$$

In Equation (5.1) C denote a constant of integration (any real number), $\int$ – symbol for integration, $F'(x)$ – integrand, dx – variable of integration. For instance, if $F(x) = x^3 - 2x$, then $F'(x) = 3x^2 - 2$ and indefinite integral of $3x^2 - 2$ is:

$$\int f(3x^2 - 2) = x^3 - 2x + C \tag{5.2}$$

Some examples of the indefinite integral formula are presented below:

$$\int x^n dx = \frac{x^{n+1}}{n+1} + C \quad (n \neq -1) \qquad \int e^x dx = e^x + C$$

$$\int a^x dx = \frac{a^x}{\ln(a)} + C \qquad \int \frac{1}{x} dx = \ln|x| + C$$

$$\int e^{ax} dx = \frac{e^{ax}}{a} + C \qquad \int \sin(x)dx = -\cos(x) + C$$

$$\int \sin(ax)dx = -\frac{1}{a}\cos(ax) + C \qquad \int \mathrm{tg}(x)dx = -\ln(\cos(x)) + C$$

For a function $y = f(x)$, continuous on $a \le x \le b$ interval, the area under the curve (Fig. 5.1) is equal to the definite integral, represented by:

$$S = \int_a^b f(x)dx = F(b) - F(a) \tag{5.3}$$

$F(a)$ and $F(b)$ denotes the lower and upper value of the integral, respectively. The distance between a and b is called also as the interval of integration.

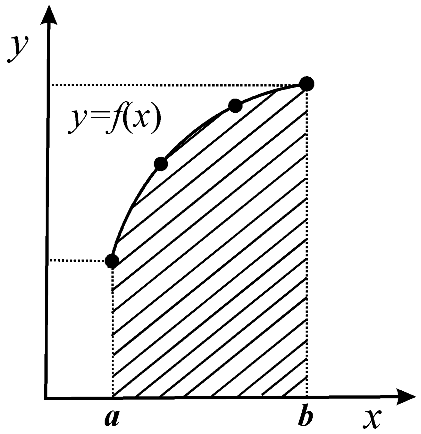


Fig. 5.1. Graphical interpretation of the definite integral (area under the curve).

In Equation (5.3), the difference between *F*(*b*) and *F*(*a*) is often denoted by the symbol $\left[F(x)\right]_a^b$.

EXAMPLE

Solve the definite integral of the function $f(x) = 3x^2 - 2$ if the lower and the upper limit are equal to *a* = 0 and *b* = 2.

SOLUTION

Using the equation for the indefinite integration of polynomial equations and the Eq. (5.4), we obtain:

$$\int_0^2 f(3x^2 - 2)dx = \left[x^3 - 2x\right]_0^2 = (8 - 4 + C) - (0 - 0 + C) = 4$$

The exact (analytical) method is applied only when the integral function is known or can be created using integration rules. If the analytical form of the function *F*(*x*) is not known (or it is too difficult to determine), is known but not integrable or the functional relationship exist as a collection of discrete data points (x_i, $f(x_i)$ values for *i* = 0, 1,.., *n*), the definite integral is calculated numerically.

5.2. Rectangle method

In the simplest method used for estimating the area under the curve (approximation of the definite integral), the interval $a \le x \le b$ is divided into *n* equal subintervals of length Δ*x*:

$$\Delta x = \frac{b - a}{n} \tag{5.4}$$

and separately calculated areas of the rectangles (Fig. 5.2) are added up.

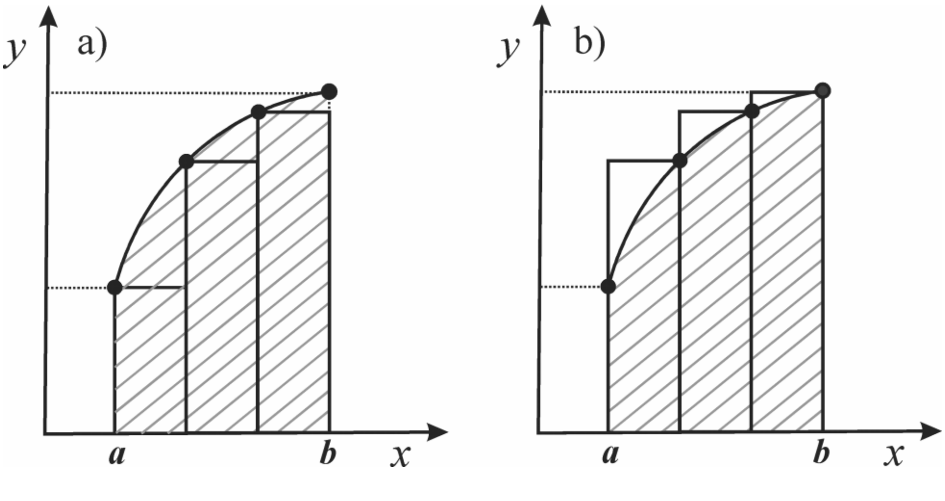


Fig. 5.2. Illustration of the different rectangle method: left–point (a) and right–point rule (b).

The approximation of the definite integral can be calculated from the equation:

$$I = \int_a^b f(x)dx \approx \Delta x \sum_{i=0}^{n-1} y_i \tag{5.5}$$

or:

$$I = \int_a^b f(x)dx \approx \Delta x \sum_{i=1}^{n} y_i \tag{5.6}$$

depending on the method (left–point or right–point rule) by which the rectangles area were calculated (Fig. 5.2 a or b).

5.3. Trapezoidal method

The area under some curve can also be estimated using the linear interpolating function for each of subinterval (lines between adjacent points, Fig. 5.3.).

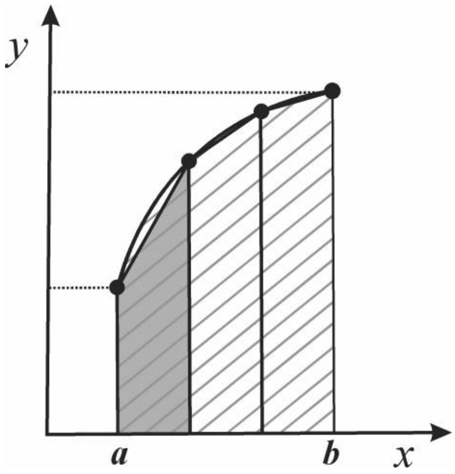


Fig. 5.3. Illustration of the trapezoidal method.

In each of the subintervals, the area of the trapezoid (P_i) can be calculated exactly from the equation which, for any interval from x_i to x_{i+1} is defined as:

$$P_i = \frac{(y_i + y_{i+1}) \cdot \Delta x}{2} \tag{5.7}$$

The sum of all areas of the trapezoids is equal to:

$$I = \int_a^b f(x)dx \approx \frac{\Delta x}{2}(y_0 + y_n + 2\sum_{i=1}^{n-1} y_i) \tag{5.8}$$

Despite the fact that the area of each trapezoid can be calculated exactly, it is only a reasonable approximation to the real area of the interval. The difference between the real area and the trapezoid's area is called the truncation error. For small values of Δx it can be assumed that the function $f(x)$ (integrand) is almost linear and the error related with trapezoid method is equal to:

$$\varepsilon \approx -\frac{\Delta x^2}{12}(y'_n - y'_0) \tag{5.9}$$

in which y'_0 and y'_n denotes the first derivatives of $f(x)$ function at the ends of the interval.

Equation (5.9) indicate, that dividing the subinterval size (Δx) in half results in a four–fold reduction of the numerical integration error:

5.4. Simpson’s rule method

The value of an integral calculated numerically can be closer to the actual value if we use a piecewise parabolic interpolation when approximating the integrand. For the three points (Fig. 5.4) there is exactly one second-degree polynomial (parabola) whose graph passes through these points (see the chapter on interpolation).

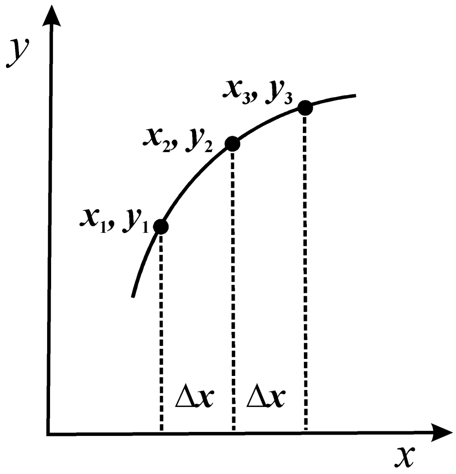


Fig. 5.4. Interpolation with a secondo–degree polynomial.

The subintervals length is equal to Δx and the interpolation polynomial passing through these three points is represented by:

$$y = a_0 + a_1 x + a_2 x^2 \tag{5.10}$$

In this case, the area under the curie between x_1 and x_3 can be calculated form the equation:

$$I_p = \int_{x_1}^{x_3} y dx = \int_{x_2-\Delta x}^{x_2+\Delta x} (a_0 + a_1 x + a_2 x^2) dx \tag{5.11}$$

Substituting the expression for y_2 in the form:

$$y_2 = a_0 + a_1 x_2 + a_2 x_2^2 \tag{5.12}$$

into Equation (5.11) after solving the equation and algebraic transformations, we obtain:

$$I_p = 2y_2 \Delta x + \frac{2}{3} a_0 \Delta x^3 \tag{5.13}$$

It can be show that:

$$y_1 + y_3 = 2y_2 + 2a_0 \Delta x^2 \tag{5.14}$$

and use this formula to eliminate a_0 from Equation (5.13). The final equation which can be used to calculate the area I_p from the values of y_1, y_2 and y_3 as well as the length of the subinterval Δx, takes the following form:

$$I_p = \frac{1}{3}\Delta x\left(y_1 + 4y_2 + y_3\right) \tag{5.15}$$

In order to calculate the area within the boundaries of any interval $\langle a, b\rangle$, it is divided into n equal parts, according to Equation (5.4). If Equation (5.15) is applied to each pair of subintervals (n must be even), then the summing of all the areas leads to the general equation:

$$I = \frac{\Delta x}{3}\left(f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4)\cdots 4f(x_{n-1}) + f(x_n)\right) \tag{5.16}$$

called the Simpson's rule. This equation can be written in the following form:

$$I = \frac{\Delta x}{3}\left(\sum(\text{two outermost points}) + 4\cdot\sum(\text{points with odd index}) + 2\cdot\sum(\text{points with even index})\right) \tag{5.17}$$

Simpson's numerical integration method therefore involves the $n + 1$ – fold determination of the integrand values and calculation of the sum of individual values of the function $f(x)$ multiplied by constant coefficients equal to 1, 4 and 2 respectively.

The error analysis related to the Simpson's method, i.e. the difference between the actual value of an integral and its approximation:

$$\int_{x_1}^{x_3} f(x)dx - \frac{\Delta x}{3}(y_1 + 4\cdot y_2 + y_3) \tag{5.18}$$

can be conducted assuming that the function $f(x)$ can be represented as a Taylor series around the point x_2, y_2:

$$f(x) = y_2 + xy_2' + \frac{x^2}{2!}y_2'' + \frac{x^3}{3!}y_2''' + \frac{x^4}{4!}y_2^{(4)} \tag{5.19}$$

In this case, a factor other than zero is the expression:

$$\frac{x^4}{4!}y_2^{(4)} \tag{5.20}$$

which after entering the Equation (5.18) allows the derivation of a formula for the error in the form:

$$e_4 \approx y^{iv}(x)\cdot(\Delta x)^5 \tag{5.21}$$

According to this formula, the error caused by the Simpson's method is directly proportional to the fourth derivative of the function multiplied by the width of the interval to the power of five. We can therefore expect that the integrals of the function for which the fourth derivative is small will be approximated with a small error. While the functions whose fourth derivative has a significant value (sinusoidal and exponential functions) will not lead to satisfactory results.

If you double the number of subintervals (n), according to Equation (5.21), the error will decrease:

$$n^5 \big/ (0.5n)^5 = 2^5 = 32 \quad \text{times}$$

In all numerical algorithms for integration, after each calculation stage the difference between integral values is analysed before and after dividing an interval. The gradual splitting of the interval is automatic. Iterations are repeated as long as two successive estimates of the integral are identical within the assumed (e.g. relative) error:

$$\frac{\left|I_k - I_{k-1}\right|}{I_k} \langle \varepsilon \tag{5.22}$$

ε is the assumed accuracy of calculations, i.e. a criterion specified by a user.

EXAMPLE

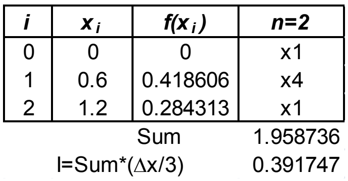
Determine the definite integral:

$$I = \int_0^{1.2} x e^{-x^2} dx$$

using Simpson’s rule with *n*= 2, 4, 6 and 12.

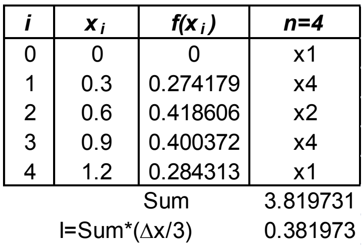
SOLUTION

For *n* = 2 the integration interval is divided into two parts. To use the Equation (5.16), we need to have the values of *x* and *y* = *f*(*x*) (integrand $f(x) = xe^{-x^2}$). The table below shows relevant calculations for $x_0 = 0$, $x_1 = 0.6$ and $x_2 = 1.2$. The last column includes *f*(*x*) value multiplier for different values of the function (according to Equation (5.16)):

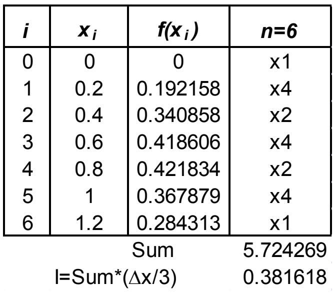


i	x_i	$f(x_i)$	*n=2*
0	0	0	x1
1	0.6	0.418606	x4
2	1.2	0.284313	x1
		Sum	1.958736
		I=Sum*(Δx/3)	0.391747

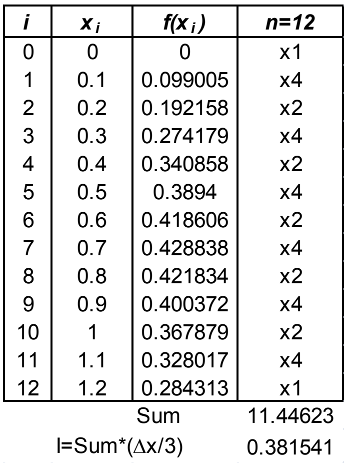
Proceeding by analogy, the following integrals can be obtained for *n*=4, 6 and 12:



i	x_i	$f(x_i)$	*n=4*
0	0	0	x1
1	0.3	0.274179	x4
2	0.6	0.418606	x2
3	0.9	0.400372	x4
4	1.2	0.284313	x1
		Sum	3.819731
		I=Sum*(Δx/3)	0.381973



i	x_i	$f(x_i)$	*n=6*
0	0	0	x1
1	0.2	0.192158	x4
2	0.4	0.340858	x2
3	0.6	0.418606	x4
4	0.8	0.421834	x2
5	1	0.367879	x4
6	1.2	0.284313	x1
		Sum	5.724269
		I=Sum*(Δx/3)	0.381618



i	*x_i*	*f(x_i)*	*n=12*
0	0	0	x1
1	0.1	0.099005	x4
2	0.2	0.192158	x2
3	0.3	0.274179	x4
4	0.4	0.340858	x2
5	0.5	0.3894	x4
6	0.6	0.418606	x2
7	0.7	0.428838	x4
8	0.8	0.421834	x2
9	0.9	0.400372	x4
10	1	0.367879	x2
11	1.1	0.328017	x4
12	1.2	0.284313	x1
		Sum	11.44623
	I=Sum*(Δx/3)		0.381541

In this example, the calculated values of integrals can be compared with the exact value calculated analytically (*I*=0.381536). The relative errors are: 2.68, 0.11, 0.021 and 0.0013 % for *n* = 2, 4, 6 i 12, respectively.

EXAMPLE

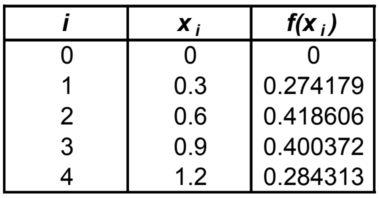
Approximate the definite integral:

$$I = \int_{0}^{1.2} xe^{-x^2} dx$$

using rectangle, trapezoid and Simpson's rule with *n*= 4.

SOLUTION

Using the results calculated in the previous example:



i	*x_i*	*f(x_i)*
0	0	0
1	0.3	0.274179
2	0.6	0.418606
3	0.9	0.400372
4	1.2	0.284313

and substituting the values into the appropriate equations, we obtain:

for the rectangle method:

$$I = 0.3\cdot(0.2742+0.4186+0.4004+0.2843)$$
$$I = 0.41325$$
Relative error = 8.3 %

for the trapezoid method:

$$I = (0.3/2)\cdot(0+0.2843+2\cdot(0.2742+0.4186+0.4004))$$
$$I = 0.3706$$
Relative error = 2.9 %

for the Simpson's method:

$$I = (0.3/3)\cdot(0 + 0.2843 + 4\cdot0.2742 + 2\cdot0.4186 + 4\cdot0.4004))$$
$$I = 0.38197$$
Relative error = 0.11 %

5.5. Gauss–Legendre method

The methods presented so far relate to the integration of function in which values of x_i are equidistant (intervals Δx of identical size). In Gaussian quadratures these points are selected so as to achieve the greatest possible accuracy for a given interpolation formula. Therefore, they are not equidistant. The general formula for these integration methods can be as follows:

$$\int_a^b w(x) f(x) dx \approx \sum_{k=1}^{n} A_k f(x_k) \tag{5.23}$$

In this equation, the integrand is the product of $w(x)$ (of the weighting function) and the general function $f(x)$. The values A_k are called weighting coefficients (factors), n is the number of elements subject to summation, (the number of internal boundaries between the intervals), while x_k stands for values of the independent variable with which the value $f(x)$ is to be estimated.

Of the many methods used to calculate integrals of various kinds, the *Gauss-Legendre method* is the simplest. In the case of this method, $a = -1$, $b = 1$ and $w(x) = 1$, so the equation (5.23) gets simplified to the form:

$$\int_{-1}^{1} f(x) dx \approx \sum_{k=1}^{n} A_k f(x_k) \tag{5.24}$$

This equation can be applied to the integration of any function $f(x)$ with a prior transformation of the integration limits a and b to –1 and +1. This requires a linear transformation of x to t, which can be carried out using the following formulas:

$$t = \frac{2x - (a+b)}{b-a} \tag{5.25}$$

$$x = \frac{1}{2}(b-a) \cdot t + \frac{1}{2}(b+a) \tag{5.26}$$

According to these equations, the integration of x ranging from a to b is equivalent to the integration of t ranging from –1 to +1:

$$\int_a^b f(x) dx = \int_{-1}^{1} g(t) dt \tag{5.27}$$

$$g(t) = f\left[\frac{1}{2}(b-a) \cdot t + \frac{1}{2}(b+a)\right] \tag{5.28}$$

Thus, the final equation takes the following form:

$$\int_a^b f(x) dx = \frac{b-a}{2} \sum_{k=1}^{n} f(x_k) \tag{5.29}$$

Appropriate values of the coefficients A_k used in the Gauss-Legendre method can be found in many textbooks on mathematics and numerical methods [*P.J. Davis, I. Polonsky, Numerical Interpolation, Differentiation and Integration in Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, M. Abramovitz, I.A. Stegun (Ed.), National Bureau of Standards Applied Mathematics Series, No 55, Washington, DC, 1964, sect. 25, pp. 875–924, T. E. Shoup, Applied numerical methods for the microcomputer, Prentice-Hall, Inc. 1984,* etc.]. Values of weight coordinates A_k and values x_k are also included in advanced computer programmes or libraries of numerical procedures.

EXAMPLE

Approximate the definite integral:

$$I = \int_0^{1.2} xe^{-x^2} dx$$

using Gauss–Legendre method with n= 4 (four–point summation).

SOLUTION

Tabulated values for the four roots(t_k) are equal to:

$$t_1 = -t_4 = 0.861136116,\ t_2 = -t_3 = 0.3399810436$$

and the four values of the coefficients (A_k) are:

$$A_1 = A_4 = 0.3478548451,\ A_2 = A_3 = 0.3478548451$$

Transforming the integration variable from x to t we obtain:

$$x = 0.6 \cdot (1 + t)$$

After inserting values for four nodes t_k to the formula for x, we get:

$$x_1 = 0.6 \cdot (1 + 0.861136116) = 1.116681787$$
$$x_2 = 0.803988626, \quad x_3 = 0.396011374, \ x_4 = 0.083318213$$

In the next step we calculate the $f(x_k) = x_k e^{-x_k{}^2}$ for four values x_k, i.e.:

$$f(x_1) = 1.116681787 \cdot e^{-(1.116681787)^2} = 0.320902925$$
$$f(x_2) = 0.421233542 \ \ f(x_3) = 0.338531764 \ \ f(x_4) = 0.082741827$$

In the last step we calculate the integral (I) from the equation:

$$I = \frac{1.2 - 0}{2} \sum_{k=1}^{4} A_k f(x_k) = 0.381532227$$

where A_k denote tabulated values of the coefficients mentioned above. The exact value of this integral is 0.381536, so the relative error is 0.000989 % and is smaller in comparison to the Simpson’s method for 12 intervals

6. Fundamentals of numerical solving of differential equations

A differential equation is an equation that contains derivatives. A first–order ordinary differential equation can be put in the following form:

$$y'(x) = \frac{dy}{dx} = f(x,y) \tag{6.1}$$

the solution is a function $y(x)$ (or a family of functions) which fulfils this equation and one of the initial conditions, usually $y(x_0)=y_0$. For example, the differential equation:

$$\frac{dy}{dx} = 3x^2 \tag{6.2}$$

has a solution:

$$y = x^3 + c \tag{6.3}$$

where c is some constant. The equations of this type are solved after separation of variables:

$$dy = 3x^2 dx \tag{6.4}$$

by integrating both sides of the equation:

$$\int dy = 3\int x^2 dx \tag{6.5}$$

The general solution of the equation (6.2) thus takes the form defined by the equation (6.3).

If an analytic solution to a differential equation is not possible, we should use numerical methods, based on the stepwise procedure during which the solution of the equation is tracked. These methods can be divided into single–step and multistep methods depending on whether for the next step of calculations we use function values and the solution from a previous step (single–step method), or a number of the immediately preceding steps (multistep method).

In contrast to the analytical solution of a differential equation, in the case of numerical methods it is necessary to specify the initial conditions (x_0, y_0).

Most of the numerical methods for solving differential equations involves expanding a function into a Taylor series. Given the value $y(x_0) = y_0$ of the function $y(x)$ at point $x = x_0,$ the value of this function at neighbouring points $x_0 + \Delta x$ can be presented in the following formula:

$$y(x_1) = y(x_0 + \Delta x) = y(x_0) + \Delta x y'(x_0) + \frac{(\Delta x)^2}{2!} y''(x_0) + \frac{(\Delta x)^3}{3!} y'''(x_0) + \cdots \tag{6.6}$$

6.1. Euler method

The Euler method, considered as the simplest, uses function expansion into a Taylor series (6.6) with only two first terms:

$$y(x_1) = y(x_0 + \Delta x) = y(x_0) + \Delta x y'(x_0) \tag{6.7}$$

Since the value of $f(x_0, y_0)$ equals the slope of the function representing a solution at a given point ($y'(x_0)$), approximate value of y_1 can be calculated from the equation:

$$y_1 = y_0 + \Delta y = y_0 + (\Delta x) \times f(x_0, y_0) \tag{6.8}$$

The value of the function at $x+\Delta x$ is therefore estimated by extrapolation as shown in Figure 6.1.

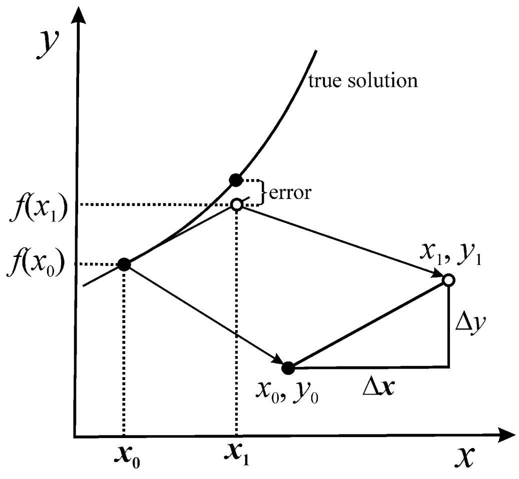


Fig. 6.1. Geometric interpretation of Euler method.

After calculating $y_1(x_1)$, the next value (y_2) is calculated from:

$$y_2 = y_1 + \Delta y = y_1 + (\Delta x) \times f(x_1, y_1) \tag{6.9}$$

etc.

The general formula for the Euler method can be put in the following form:

$$y_{n+1} = y_n + \Delta y = y_n + (\Delta x) \cdot f(x_n, y_n) \tag{6.10}$$

After repeated application fo the recursive Equation (6.10), the results obtained take the form of a set of values from (x_0, y_0) to the last value (x_k, y_k) of the calculations

Since this method involves a stepwise transition from one interval Δx to another, real and numerical solutions show an increasing divergence. Errors committed in individual steps are accumulated, which leads to a difference characteristic of the iterative method for solving differential equations (Fig. 6.2).

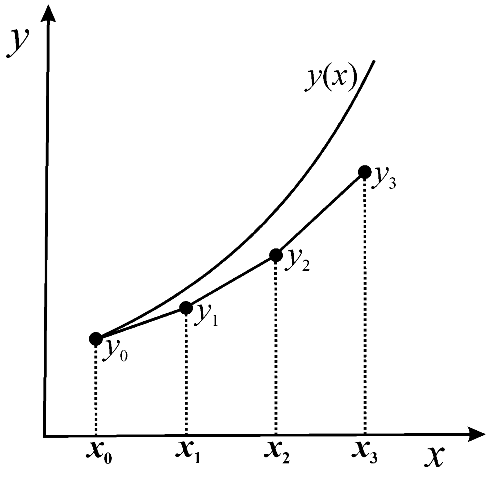


Fig. 6.2. Accumulation of errors in the Euler method.

The Euler method is considered to be among the first–order methods, as in a Taylor series all expressions with powers Δx higher than the first one have been omitted. Therefore, an error per a single step is on the order of $(\Delta x)^2$, while the method error per all steps is on the order of Δx. The solution quality obviously depends on the size of Δx – reduction of a step by half the length reduces by four times an error per a single step.

Scheme of algorithm for Euler method is presented in Fig. 6.3.

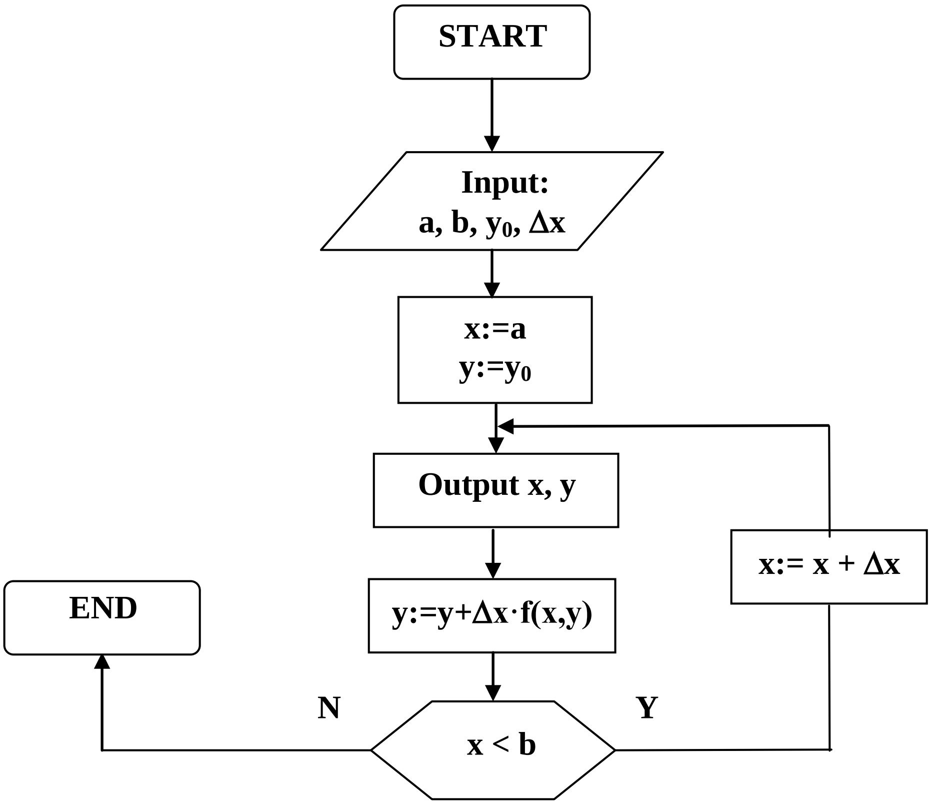


Fig. 6.3. Scheme of algorithm for Euler method.

EXAMPLE

For the following differential equation:

$$y'(x) = y + 1,$$

and the initial condition:

$$y(0) = 1 \ (x_0 = 0, y(x_0) = 1)$$

calculate the approximate value of $y(1)$, using the Euler method with steps $\Delta x = 0.2$ and $\Delta x = 0.12$.

SOLUTION

Using the Equations (6.8) and (6.9), we obtain:

$$y_1 = y_0 + \Delta y = y_0 + (\Delta x) \times f(x_0, y_0) = 1 + 0.2 \cdot (1+1) = 1.4$$

$$y_2 = y_1 + \Delta y = y_1 + (\Delta x) \times f(x_1, y_1) = 1.4 + 0.2 \cdot (1.4+1) = 1.88$$

etc.

Continuing the calculations for other values of x_n and performing similar calculations for $\Delta x = 0.1$, the results can be put in the following table which also shows the exact values (Ex) and the relative error (Err).

Δx	n	x_n	y_n	Ex	Err	Δx	n	x_n	y_n	Ex	Error
0.2	0	0.0	1.0000	1.0000	0.00	0.1	0	0.0	1.0000	1.0000	0.00
							1	0.1	1.2000	1.2103	0.85
	1	0.2	1.4000	1.4428	2.97		2	0.2	1.4200	1.4428	1.58
							3	0.3	1.6620	1.6997	2.22
	2	0.4	1.8800	1.9836	5.22		4	0.4	1.9282	1.9836	2.79
							5	0.5	2.2210	2.2974	3.32
	3	0.6	2.4560	2.6442	7.12		6	0.6	2.5431	2.6442	3.82
							7	0.7	2.8974	3.0275	4.30
	4	0.8	3.1472	3.4511	8.81		8	0.8	3.2872	3.4511	4.75
							9	0.9	3.7159	3.9192	5.19
	5	1.0	3.9766	4.4366	10.37		10	1.0	4.1875	4.4366	5.61

The above example shows that the quality of the solution of a differential equation with the numerical method depends on Δ*x*. As Δ*x* is reduced, the differences between the numerical and analytical (exact) solutions decrease. Theoretically there is nothing to prevent a reduction of Δ*x*, but then the amount of calculations also increases (calculation time increases). In practice, preliminary tests are carried out to determine a quantity Δ*x* necessary to achieve sufficiently accurate results

6.2. Runge–Kutta method

Runge and Kutta developed a number of methods for stepwise transition through the interval from x_0 to $x_0 + \Delta x$ of an increasing accuracy of calculations. The first–order R–K method is the same as Euler method. In the R–K methods, some especially selected intermediate points lying close to a solution $f(x, y)$ were calculated.

The second–order R–K method uses an approximate value of a slope (c_2) in the middle of a point–connecting interval (x_n, y_n) and (x_{n+1}, y_{n+1}):

$$c_2 = f(x_n + \frac{1}{2}\Delta x,\ y_n + \frac{1}{2}\Delta x c_1) \tag{6.11}$$

where c_1:

$$c_1 = f(x_n, y_n) \tag{6.12}$$

denotes the value of the slope at the starting point of the interval (x_n, y_n).

Geometric interpretation of the second–order Runge–Kutta method is presented in Fig. 6.4.

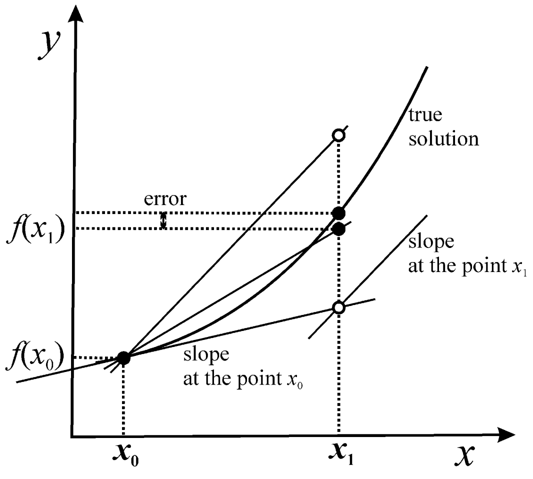


Fig. 6.4. Geometric interpretation of the second–order Runge–Kutta method.

The widely used fourth–order Runge–Kutta method, is defined by the equations:

$$y_{n+1} = y_n + \frac{1}{6}\Delta x\left(c_1 + 2c_2 + 2c_3 + c_4\right) \tag{6.13}$$

$$c_1 = f(x_i, y_i) \tag{6.14}$$

$$c_2 = f(x_n + \frac{1}{2}\Delta x,\ y_n + \frac{1}{2}\Delta x c_1) \tag{6.15}$$

$$c_3 = f(x_n + \frac{1}{2}\Delta x,\ y_n + \frac{1}{2}\Delta x c_2) \tag{6.16}$$

$$c_4 = f(x_n + \Delta x,\ y_n + \Delta x\, c_3) \tag{6.17}$$

where c_1 denote the value of the slope of the solution function at the starting point of the interval x=x_0; c_2 and c_3 – the values of the slope at the midpoint of the interval:

$$x = x_0 + \frac{1}{2}\Delta x \tag{6.18}$$

c_4 the value of the slope at the end of the interval:

$$x = x_0 + \Delta x \tag{6.19}$$

The equations (6.13)–(6.17) show that every step of calculations requires designating four function values in points c_1, c_2, c_3 and c_4.

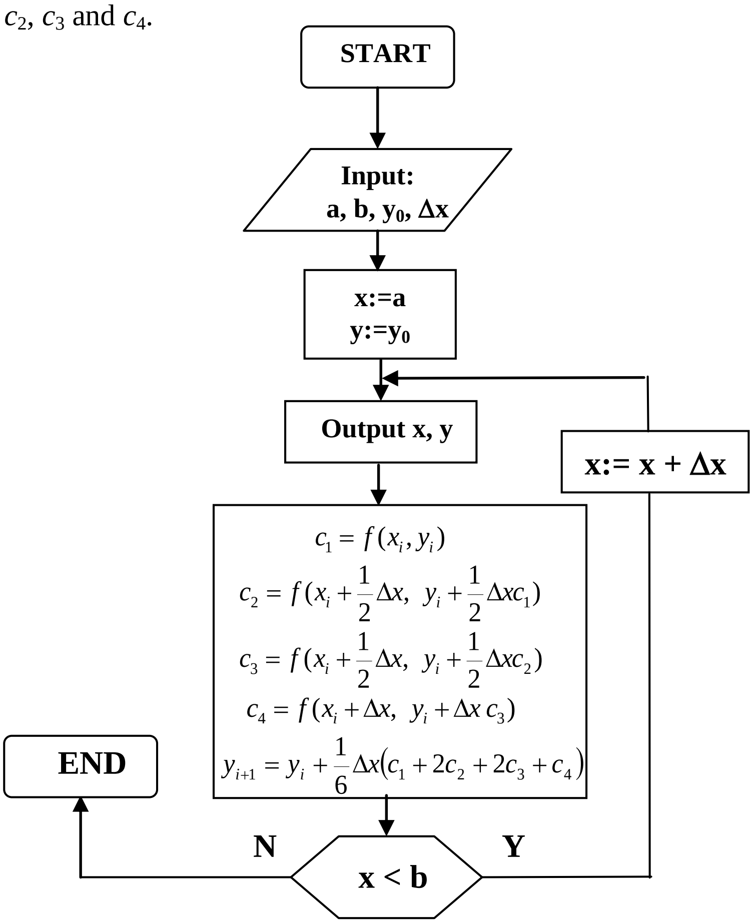

START
Input:
a, b, y_0, Δx
x:=a
y:=y_0
Output x, y
x:= x + Δx
$c_1 = f(x_i, y_i)$
$c_2 = f(x_i + \frac{1}{2}\Delta x,\ y_i + \frac{1}{2}\Delta x c_1)$
$c_3 = f(x_i + \frac{1}{2}\Delta x,\ y_i + \frac{1}{2}\Delta x c_2)$
$c_4 = f(x_i + \Delta x,\ y_i + \Delta x\, c_3)$
$y_{i+1} = y_i + \frac{1}{6}\Delta x(c_1 + 2c_2 + 2c_3 + c_4)$
END
N
Y
x < b

Fig. 6.5. Scheme of algorithm for fourth–order Runge–Kutta method.

The fourth–order Runge–Kutta method is the most commonly used numerical method to solve the system of ordinary differential equations with initial conditions.This method is simple to implement and provides a high accuracy solution. Professional programmes using the fourth–order R–K method have a mechanism for automatic selection of the length of an integration step, which speeds up the computation time while maintaining the assumed accuracy.

The R–K method takes into account the factor $(\Delta x)^4$ which occurs in a Taylor series (Eq. (6.6)). The first factor omitted is $(\Delta x)^5$, which indicates that the error made in estimating Δy is inversely proportional to the fifth power of Δx, and at the same time a small decrease in Δx causes an enormous increase in the accuracy of calculations.

Like the Euler method, the R–K methods are self–starting methods in which a single starting point (x_0, y_0) is sufficient to start calculations.

EXAMPLE

Find an approximate value of $y(1)$ for the equation $y'(x) = y + 1$ and the initial condition $y(0) = 1$, using the fourth–order R–K method.

SOLUTION

For x_0, $y(x_0) = 1$, using Equations (6.13) – (6.17) we get:

$$c_1 = y+1 = 1+1 = 2$$

$$c_2 = y+1+\frac{1}{2}\Delta x c_1 = 2+\frac{1}{2}\cdot 0.2\cdot 2 = 2.2$$

$$c_3 = y+1+\frac{1}{2}\cdot 0.2\cdot 2.2 = 2.22$$

$$c_4 = y+1+0.2\cdot 2.22 = 2.444$$

$$y_1 = y_0+\frac{1}{6}\Delta x\left(c_1+2c_2+2c_3+c_4\right) = 1+\frac{1}{2}\cdot 0.2\cdot(2+2\cdot 2.2+2\cdot 2.22+2.444))$$

etc.

The calculated values are summarized in the following table:

Δx	n	x_n	y_n	Ex	Err	c_1	c_2	c_3	c_4
0.2	0	0	1	1.0000	0.0000				
	1	0.2	1.4428	1.4428	0.0000	2	2.2	2.22	2.444
	2	0.4	1.98364	1.9836	0.0018	2.443	2.687	2.711508	2.9851
	3	0.6	2.64421	2.6442	0.0005	2.984	3.282	3.311836	3.646
	4	0.8	3.45104	3.4511	0.0017	3.644	4.009	4.045076	4.45323
	5	1	4.4365	4.4366	0.0022	4.451	4.896	4.940656	5.43917

The results indicate that the fourth–order Runge–Kutta method with the step Δx=0.2 gives a much better accuracy than the Euler method with Δx=0.1.

The fourth–order R–K method can be used to solve a system of coupled first–order differential equations which take the following form:

$$
\begin{aligned}
\frac{dy_1}{dx} &= f_1(x, y_1, y_2 \cdots y_n) \\
\frac{dy_2}{dx} &= f_1(x, y_1, y_2 \cdots y_n) \\
&\vdots \\
\frac{dy_n}{dx} &= f_1(x, y_1, y_2 \cdots y_n)
\end{aligned}
\tag{6.20}
$$

The initial values are determined by point $x = x_0$, which means that the values $y_1(x_0), y_2(x_0), \cdots, y_n(x_0)$ are known. A typical example of a system of coupled first-order differential equations are equations describing the kinetics of a consecutive reaction:

$$A \underset{k_2}{\overset{k_1}{\rightleftarrows}} B \underset{k_4}{\overset{k_3}{\rightleftarrows}} C$$

in the following form:

$$
\begin{aligned}
\frac{d[A]}{dt} &= -k_1[A] + k_2[B] \\
\frac{d[B]}{dt} &= -(k_3 + k_2)[B] + k_1[A] + k_4[C] \\
\frac{d[C]}{dt} &= -k_4[C] + k_3[B]
\end{aligned}
\tag{6.21}
$$

where [A], [B] i [C] de notes the concentration [mol/dm^3], k – reaction rate constant [s^{-1}]. The equations are coupled, as the products of one reaction are the substrates of the next one.

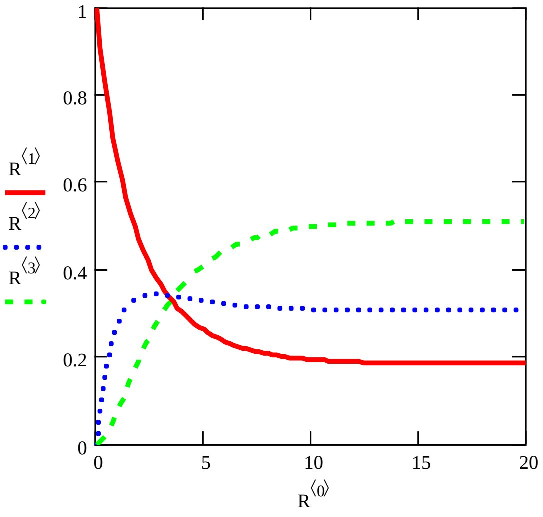
The solution involves determining the concentration changes [A], [B] and [C] depending on time ([A]=f(t), [B]=f(t) and [C]=f(t)) for specific reaction rate constants. The initial conditions for the reaction are assumed to be the initial values of the integration procedure (eg. $[A]_0 = 1$, $[B]_0 = 0$ i $[C]_0 = 0$).

Below is a sample solution of this system of equations in Mathcad. After entering the vector containing the initial concentration values (C), reaction rate constants k_1–k_4, the vector of the first derivatives of the functions sought (P(t,C)) and the function (rkfixed) which solves the system of differential equations (6.21), we can obtain a graph of the functions that are the solution.

$$C := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad k_1 := 0.5 \qquad k_2 := 0.3 \qquad k_4 := k_2 \qquad k_3 := k_1$$

$$P(t, C) := \begin{pmatrix} -k_1 \cdot C_0 + k_2 \cdot C_1 \\ k_1 \cdot C_0 - k_2 \cdot C_1 - k_3 \cdot C_1 + k_4 \cdot C_2 \\ k_3 \cdot C_1 - k_4 \cdot C_2 \end{pmatrix}$$

$$R := \mathrm{rkfixed}(C, 0, 20, 100, P)$$



A numerical solution of differential equations in Mathcad allows for tracking changes in the course of individual curves (concentration changes) depending on the reaction rate constants and initial concentration of substance *A*.

6.3. Milne method (predictor-corrector)

An alternative method for a single–step transition through the interval Δx while solving differential equations involves the use of more than one starting point in calculations. These methods, called multistep methods, use approximate values calculated in several consecutive, immediately preceding steps to perform one step of computations. Due to the fact that the initial condition is given only at one point, calculations begin with any self–starting (one–step) method. After generating a necessary number of initial values, the calculations can be continued with any multistep method.

Overall, multistep methods use extrapolation from point $y_0 = y(x_0)$ to a new point $y_1 = y(x_1)$ using a prediction step and then a correction step. The prediction step is performed by fitting a polynomial to point y'_0 and the two previous points, y'_{-1} and y'_{-2} (Fig. 6.6).

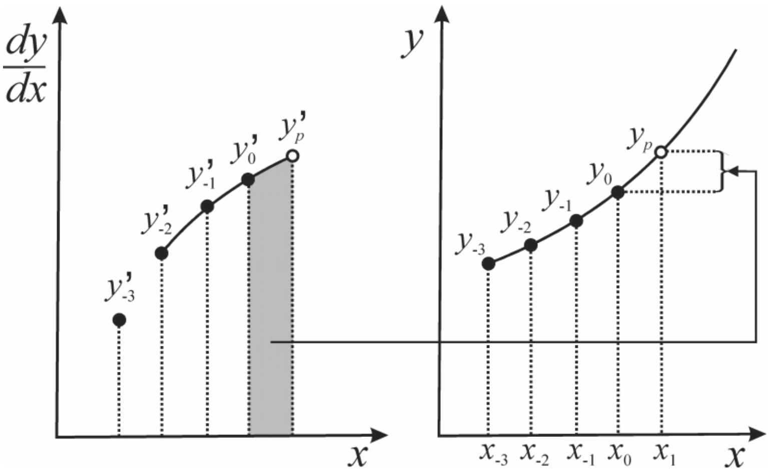


Fig. 6.6. The principle of the Milne method.

A parabola fitted to these points is extrapolated through the interval Δx and allows for the calculation of the area under the parabola based on the previous points y'_k. The calculated surface area is added to the value of *y* to calculate the predicted value of y_1 (called y_{1p}):

$$y_{1,p} = y_{-3} + \frac{4}{3}\Delta x\left(2y'_{-2} - y'_{-1} + 2y'_0\right) \tag{6.22}$$

The differential equation is then used to correct the first calculated value of y_{1p} by substituting this value and x_1 to the equation:

$$y'_1 = f(x_1, y_{1p}) \tag{6.23}$$

and calculating a derivative at point x_1.
The estimated value y'_1 along with the two previous points (y'_0 i y'_{-1}) are used to calculate a new, better parabola applied to calculate a new, corrected y_{1c}:

$$y_{1c} = y_{-1} + \frac{\Delta x}{3}\left(y'_{-1} + 4y'_0 + y'_1\right) \tag{6.24}$$

and a corrected derivative:

$$y_1' = f(x_1, y_{1c}) \tag{6.25}$$

The application of the method can be traced on the example presented below.

EXAMPLE

Find an approximate value of $y(1)$ for the equation $y'(x) = y + 1$ and the initial condition $y(0) = 1$, using the Milne method.

SOLUTION

To perform calculations with this method, we need to have the following values:

$$\begin{array}{cccc} y_0, & y_{-1}, & y_{-2}, & y_{-3} \\ y_0', & y_{-1}', & y_{-2}', & y_{-3}' \end{array}$$

which can be calculated using the fourth–order Runge–Kutta method (see previous example). Further calculations are carried out according to the equations (6.22)–(6.25), i.e.:

$$y_{1,p} = y_{-3} + \frac{4}{3}\Delta x\left(2y_{-2}' - y_{-1}' + 2y_0'\right) = 1 + \frac{4}{3}\cdot 0.2\cdot\left(2\cdot 2.4428 - 2.9836 + 2\cdot 3.6442\right) = 3.450771$$

$$y_1' = f(x_1, y_{1p}) = y_{1p} + 1 = 3.450771 + 1 = 4.450771$$

$$y_{1c} = y_{-1} + \frac{\Delta x}{3}\left(y_{-1}' + 4y_0' + y_1'\right) = 1.98364 + \frac{0.2}{3}(2.9836 + 4\cdot 3.6442 + 4.450771) = 3.45105$$

$$y_1' = f(x_1, y_{1c}) = y_{1c} + 1 = 3.45105 + 1 = 4.45105$$

etc.
The results of the calculations are shown in the table below:

Δx	n	x_n	y_n	y'_n	y_{1p}	y'_1	y_{1c}	y'_1	Ex	Err
0.2	-3	0	1	2.0000					1.0000	0.0000
	-2	0.2	1.4428	2.4428					1.4428	0.0000
	-1	0.4	1.98364	2.9836					1.9836	0.0018
	0	0.6	2.64421	3.6442					2.6442	0.0005
	1	0.8	3.45105	4.4511	3.450771	4.450771	3.45105	4.45105	3.4511	0.0014
	2	1	4.43652	5.4365	4.436177	5.436177	4.43652	5.43652	4.4366	0.0018

In this example, after the two–step calculations, the error obtained is by over 12% smaller than the one obtained with the R–K method.

An undeniable advantage of such methods is the ability to track the accuracy of numerical solutions, because the difference between values y_{1p} and y_{1c} provides information whether the adopted interval is appropriate. The Milne method, like the R–K method, has an error proportional to the fifth power of Δx.

7. Methods for solving algebraic equations

An algebraic equation (not containing derivatives or integrals) with one unknown can be written as:

$$f(x) = 0 \tag{7.1}$$

Solutions of this equation are values of x for which the above equality is fulfilled (roots of function $f(x)$).

Among the different numerical root–finding methods this chapter will discuss:

- bisection method
- secant method (regula falsi)
- tangent method (Newton–Raphson)

7.1. Bisection method

If the function is continuous at intervals $\langle x_1, x_2 \rangle$ and change the sign, i.e. $f(x_1)\cdot f(x_2)<0$, this indicate, that there is at least a one root between x_1 and x_2.

According to the definition of a function's continuity, if $f(x_1)\cdot f(x_2)<0$, then the interval $\langle x_1, x_2 \rangle$ contains at least one such point at which $f(x)=0$. In the bisection method, to determine the approximate zero of a function, the interval $\langle x_1, x_2 \rangle$ gradually decreases so as to contain the element sought. The starting point in this method is two argument values for which the function $f(x)$ changes its sign (Fig. 7.1).

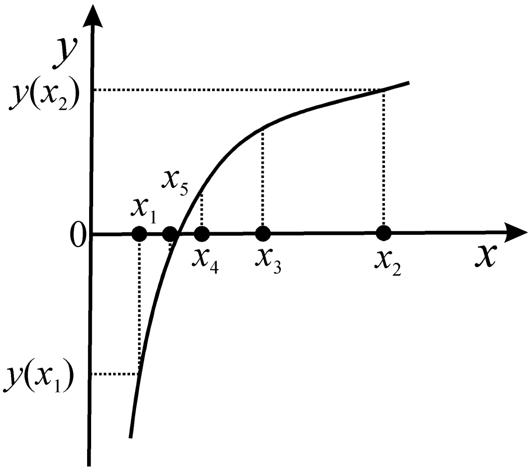


Fig. 7.1. Graph illustrating the bisection method.

In the first step we calculate $f(x_3)$ at the midpoint of the interval:

$$x_3 = ½\cdot(x_1+x_2) \tag{7.2}$$

As a result, we get two intervals twice smaller than the initial interval. If $f(x_3)>0$, then the solution is between points x_1 and x_3:

$$x_4 = ½\cdot(x_1+x_3) \tag{7.3}$$

Otherwise (for $f(x_3)<0$) the zero of the function is between points x_2 and x_3:

$$x_4 = ½\cdot(x_2+x_3) \tag{7.4}$$

The calculations are repeated several times until a sufficiently good estimate of zero is obtained. In practice, the iterative calculations end after fulfilling any of the following conditions:

$$\left|x_{n+1} - x_n\right| < \varepsilon \tag{7.5}$$

which means that the difference between successive approximations is small enough, or:

$$|f(x_n)| < \varepsilon \tag{7.6}$$

i.e. the value of the function at the designated point is close to 0 (lower than ε). In these equations, ε is the assumed accuracy of calculations (criterion specified by a user). These equations ((7.5) and (7.6)) are also used in the secant and tangent methods.

Simplicity is an important advantage of the bisection method. The fundamental disadvantages include a slow convergence of the iterative process, and problems if many zeros of a function are concentrated in a very narrow interval (Fig. 7.2).

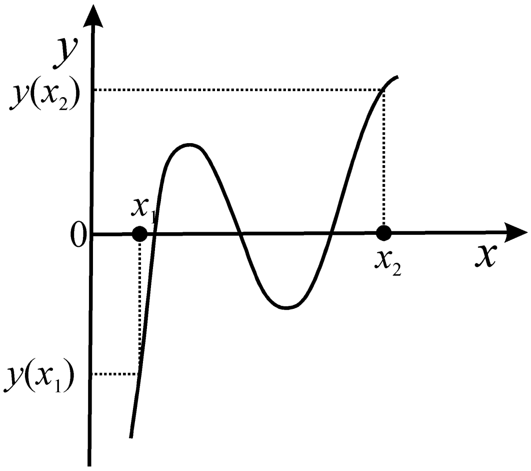


Fig. 7.2. The location of many zeros within a narrow interval $\langle x_1, x_2 \rangle$.

EXAMPLE

Find the root for the function:

$$f(x) = 0.1 \cdot x^3 - x^2 + 1$$

between $x_1 = 9$ and $x_2 = 10$ using bisection method.

SOLUTION

The values of the function at the points $x_1 = 9$ and $x_2 = 10$ are respectively:

$$f(x_1) = -7.1 \text{ and } f(x_2) = 1$$

Using Equation (7.2), we can determine the x_3 value and the value of the function ($f(x_3)$):

$$x_3 = 9.5 \text{ and } f(x_3) = -3.51$$

Analysing the results, it can be stated that the zero of the function is between $x_3 = 9.5$ and $x_2 = 10$. Repeating the calculations for these two points, we obtain:

$$x_4 = 9.75 \text{ and } f(x_4) = -1.376$$

Therefore the roots of the function is located between $x_4 = 9.75$ and $x_2 = 10$. Further calculations lead to the following results:

$$x_5 = 9.875 \text{ and } f(x_5) = -0.219$$

indicating the presence of the zero in the interval ⟨$x_5 = 9.875$, $x_2 = 10$⟩. In the next step the following values are obtained:

$$x_6 = 9.9375 \text{ and } f(x_6) = 0.3828$$

that lead to the next reduction of the interval (⟨$x_5 = 9.875$, $x_6 = 9.9375$⟩). In the step last analysed in the example, the calculated values are:

$$x_7 = 9.90625 \text{ and } f(x_7) = 0.07999$$

The final result (x_7) can be compared with the exact value of the zero of the function ($f(x) = 0$ for $x =$ 9.89793) by calculating the relative error which is 0.084%.
The algorithm diagram for this method is presented in Fig. 7.3.

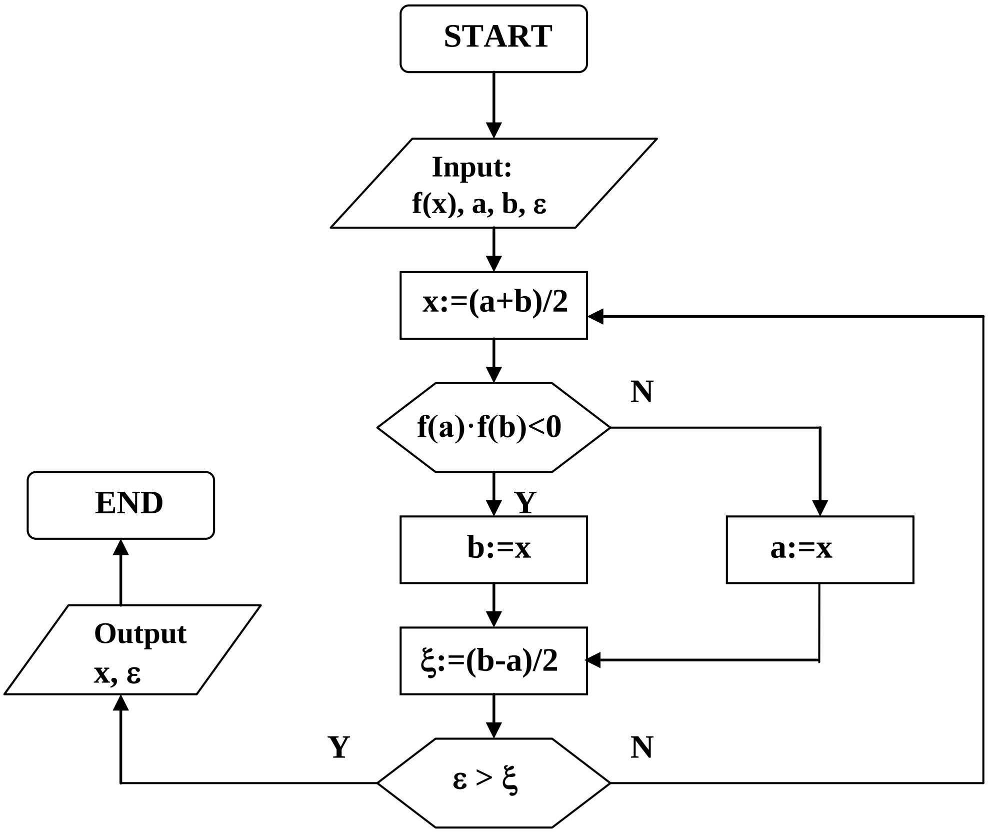


Fig. 7.3. The algorithm diagram for bisection method.

7.2. Secant method (regula falsi)

In this method, also called the false position method, a chord is drawn through points x_1 and x_2, for which the function $f(x)$ changes its sign, with the following equation:

$$y - f(x_1) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}(x - x_1) \tag{7.7}$$

The abscissa x_3 of the point at which the fixed chord AB intersects the axis OX (Fig. 7.4), is assumed as the first approximation of the desired zero location.

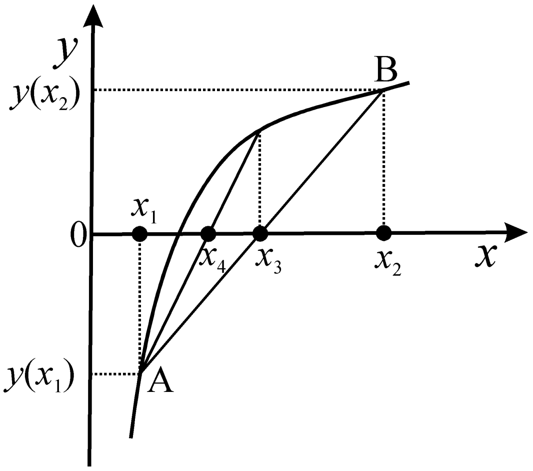


Fig. 7.4. Graph illustrating the secant method.

The value of the point x_3 can be calculated from the equation:

$$x_3 = x_2 - f(x_2)\frac{x_2 - x_1}{f(x_2) - f(x_1)} \tag{7.8}$$

As in the bisection method, the calculations are continued until a sufficiently good estimate of the roots is obtained. The general recursive equation for this method can be as follows:

$$x_{(k+2)} = x_{(k+1)} - f(x_{(k+1)})\frac{x_{(k+1)} - x_k}{f(x_{(k+1)}) - f(x_k)} \tag{7.9}$$

where $k = 1, 2, \ldots$

EXAMPLE

Find the root for the function:

$$f(x) = 0.1 \cdot x^3 - x^2 + 1$$

between $x_1 = 9$ and $x_2 = 10$ using secant method.

SOLUTION

Just as in the example analyzed previously, the values of the function at points $x_1 = 9$ and $x_2 = 10$ are respectively:

$$f(x_1) = -7.1 \text{ and } f(x_2) = 1$$

Substituting the numerical values in Equation (7.8), we can calculate the value of the argument at point x_3 and the corresponding value of the function ($f(x_3)$):

$$x_3 = 9.87653 \quad \text{and} \quad f(x_3) = -0.20427$$

In the next step, using Equation (7.9) in the form:

$$x_4 = x_3 - f(x_3)\frac{x_3 - x_2}{f(x_3) - f(x_2)}$$

the following results are obtained:

$$x_4 = 9.89748 \quad \text{and} \quad f(x_4) = -0.00425$$

Comparing the final result (x_4) with the exact value of the root ($f(x) = 0$ for $x = 9.89793$), we obtain a relative error equal to 0.0045%. In comparison to the bisection method discussed earlier, after two approximations made with the secant method, the error of estimation is almost 20 times smaller

The algorithm diagram for this method is presented in Fig. 7.5.

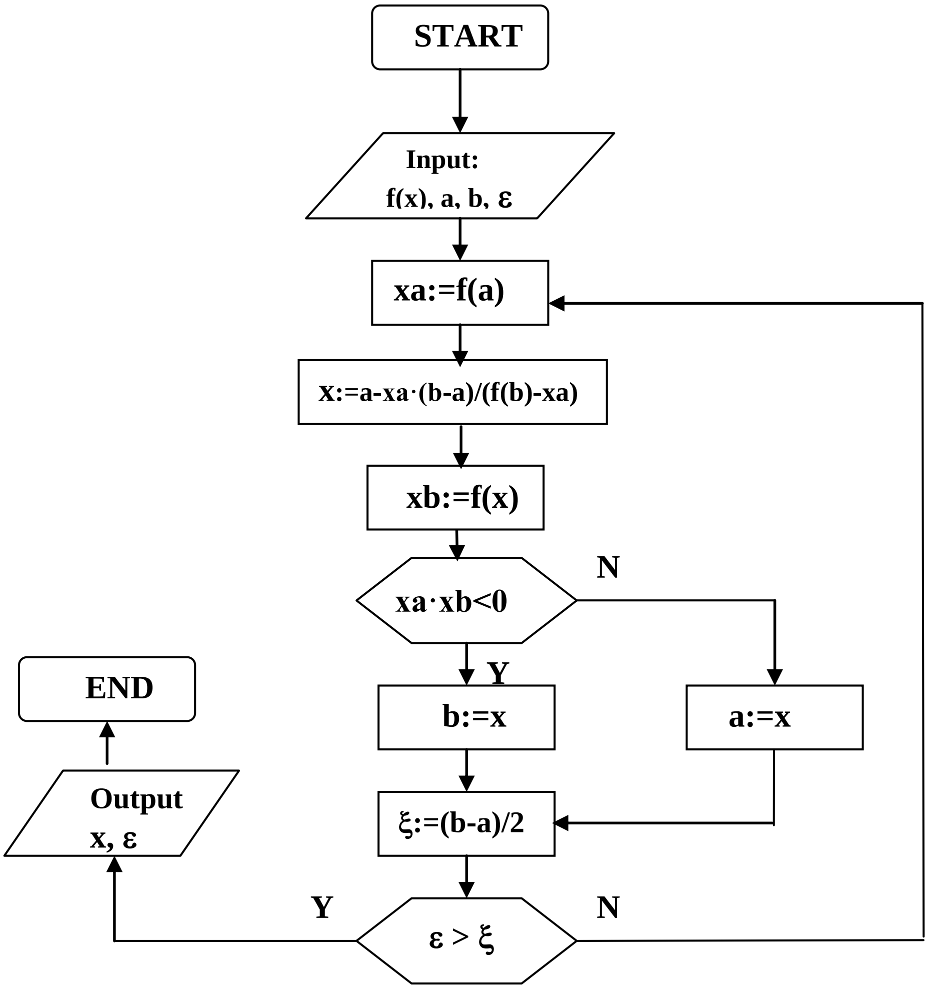


Fig. 7.5. The algorithm diagram for secant method.

7.3. Tangent method (Newton-Raphson)

In this method, which is the most common method for determining zeros of functions, it is necessary to know the function $f(x)$ and its derivative $f'(x)$. According to this method (Fig. 7.6), the slope of the tangent to the graph at point x_2 can be calculated from the expression:

$$f'(x_2) = \frac{f(x_2)}{x_2 - x_3} \tag{7.10}$$

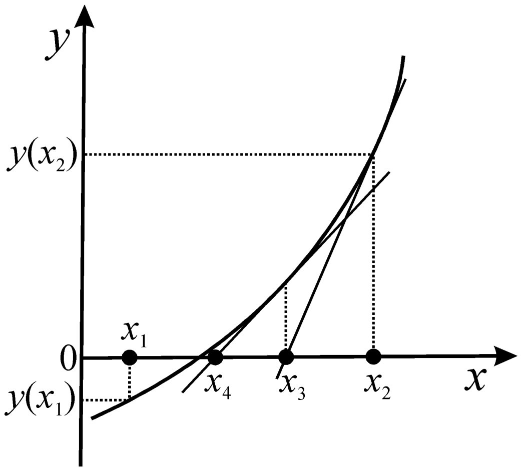


Fig. 7.6. The Newton–Raphson method.

Therefore, the first approximation of the root (x_3) can be calculated from the equation:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} \tag{7.11}$$

The general recursive equation can be written as:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \tag{7.12}$$

EXAMPLE

Find the root for the function:

$$f(x) = 0.1\cdot x^3 - x^2 + 1$$

between $x_1 = 9$ and $x_2 = 10$ using tangent method.

SOLUTION

To apply Equation (7.11) in calculations, we need to have the value of the function at point x_2 ($f(x_2) = 1$) and its derivative at point x_2. The derivative of the function is:

$$f'(x) = 0.3\cdot x^2 - 2\cdot x$$

Thus the derivative at the point is $f'(x_2) = 10$. Substituting appropriate values in the Equation (7.11) gives the following results:

$$x_3 = 9.9,\ f(x_3) = 0.0199 \text{ and } f'(x_3) = 9.603$$

In the second step of the calculations, using the Equation (7.12) in the form:

$$x_4 = x_3 - \frac{f(x_3)}{f'(x_3)}$$

we get the following results:

$$x_4 = 9.897928 \text{ and } f(x_4) = 8.46\cdot10^{-6}$$

The relative error for this method, after two steps of calculations, is only $7.4\cdot10^{-6}$ % and is over 600 times smaller than the relative error calculated with the secant method.
The algorithm diagram for the Newton–Raphson method is presented in Fig. 7.7.

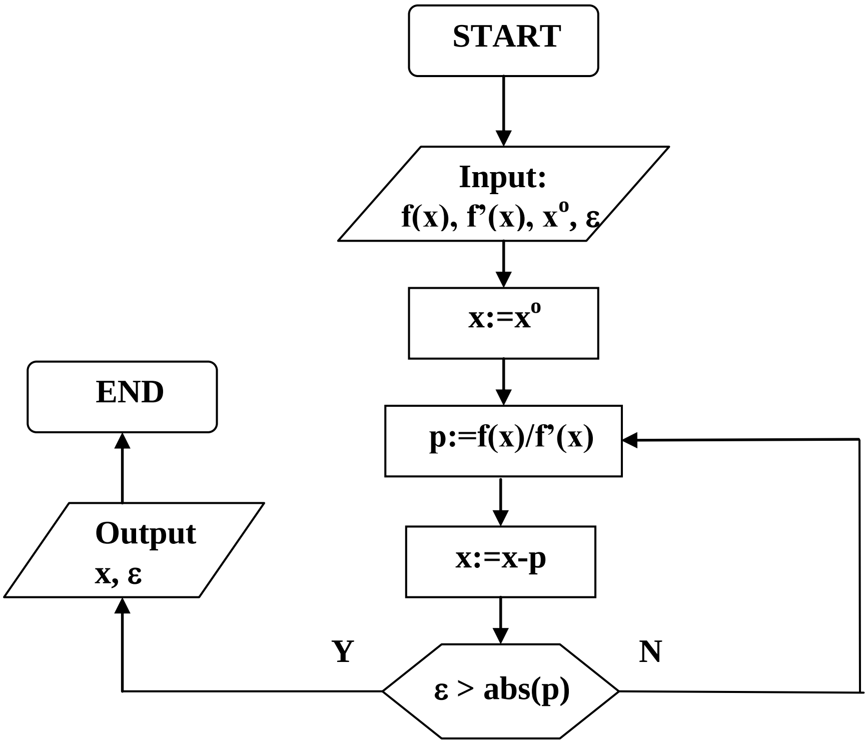


Fig. 7.7. The Newton–Raphson algorithm.

8. Methods for solving systems of linear equations

8.1. Matrix calculus - fundamentals

Systems of linear equations are often used to describe numerous chemical phenomena such as a multiple correlation analysis, a study of equilibria in multicomponent systems or a spectrophotometric analysis of mixtures. A linear equation can be presented in the general form:

$$y = a_0 + a_1x_1 + a_2x_2 \cdots a_nx_n \tag{8.1}$$

As methods for solving systems of linear equations usually use the matrix calculus, the basic concepts in this field are presented below.

A matrix is an object that consists of $m\times n$ elements located in an array built of parenthesized m rows and n columns (dimension of matrix is $m\times n$).
Any matrix **A** (indicated in bold letter) consisting of mn elements can be written as follows:

$$\mathbf{A} = a_{ij} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mn} \end{bmatrix} \tag{8.2}$$

Each element of a matrix (a_{ij}) is in an adequate *i*-th row ($i = 1, 2, \ldots, m$) and *j*-th column ($j = 1, 2, \ldots, n$). The matrix calculus is used to perform many elementary algebraic operations. If we add a matrix **A** to **B**, a matrix **C** is obtained:

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \tag{8.3}$$

The sum of matrices can be calculated only if the matrices have the same dimensions and it consists in adding together the elements of matrices **A** and **B** with identical indices:

$$c_{ij} = a_{ij}+b_{ij} \tag{8.4}$$

for all $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.
In adding matrices, as in the case of simple addition, the principle of alternation prevails, namely:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A} \tag{8.5}$$

$$\text{and} \quad (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \tag{8.6}$$

The product of a matrix **A** and a scalar (number) *c* is obtained by multiplying each element of the matrix **A** by the constant *c.*

$$c\,\mathbf{A} = (ca_{ij}) \tag{8.7}$$

The product of a matrix **A** and a matrix **B** to obtain a matrix **C**:

$$\mathbf{A}\,\mathbf{B} = \mathbf{C} \tag{8.8}$$

can be obtained only when the number of columns in **A** equals the number of rows in **B**. In this case, *i,j*–th element in the matrix **C** is the sum of the products of element pairs in the *i*-th row of the matrix **A** and the *j*-th column of the matrix **B**:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} \tag{8.9}$$

If **A** is a matrix of *m*×*n* size, and **B** is a matrix of *n*×*p* size, the matrix **C** has the same number of rows as **A** and the same number of columns as **B**, i.e. its dimension is *m*×*p*:

$$\mathbf{A}\ (\text{matrix } m{\times}n)\ \mathbf{B}\ (\text{matrix n×p}) = \mathbf{C}\ (\text{matrix } m{\times}p) \tag{8.10}$$

At the same time, the alternation principle does not apply in matrix multiplication, that is, not for all cases:

$$\mathbf{A}\,\mathbf{B} = \mathbf{B}\,\mathbf{A} \text{ or } \mathbf{A}(\mathbf{B}\,\mathbf{C}) = \mathbf{A}\,\mathbf{B}(\mathbf{C}) \tag{8.11}$$

A matrix **A** is the inverse matrix of **B** (and *vice versa*) and it is a square matrix (number of rows equals the number of columns)which fulfils the equation:

$$\mathbf{A}\,\mathbf{B} = \mathbf{I} = \mathbf{B}\,\mathbf{A} \tag{8.12}$$

where **I** is an identity matrix whose diagonal elements are equal to 1 ($a_{ii} = 1$), while the rest are equal to 0. For a square matrix **A**, the inverse is written as $\mathbf{A}^{-1}$, therefore:

$$\mathbf{A}\,\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I} \tag{8.13}$$

A matrix **A** has an inverse if and only if its determinant is not zero. In this case, the matrix **A** is called a nonsingular matrix

Calculating an invertible matrix is one of the main operations of matrix algebra, applied in solving a system of linear equations.

The system of linear equations can be written in the following way:

$$\begin{aligned} a_{11}x_1 + a_{12} + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22} + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2} + \cdots + a_{mn}x_n &= b_n \end{aligned} \tag{8.14}$$

According to the principles of matrix algebra, this system of linear equations can be written as:

$$\mathbf{A}\,\mathbf{X} = \mathbf{B} \tag{8.15}$$

In this equation **A** is a matrix of coefficients *a* with dimensions $m\times n$ (Equation (8.2)), **X** is a vector (column matrix, dimension $n\times1$) containing unknowns or the solutions of the system of equations (x_n). The matrix **B** is a constant (b_n) column matrix of size $n\times1$.

When we start to solve the above system of equations (determine the matrix elements **X**), we need to have a square matrix of coefficients (**A**), i.e. as many equations as unknowns occurring in them (fulfil the condition $m = n$). If the matrix **A** is nonsingular, then the system of equations described by the formula (8.15) has exactly one solution.

8.2. Cramer method

Systems of linear equations of the general form defined by the equations (8.14) can be solved using Cramer’s formulas. The solution of the system of equations where $m=n$ (dimension $n\times n$) is defined by the ratio of two determinants. In this quotient, the denominator is the determinant of the coefficient matrix (*a*), while the numerator is the same determinant where the *i*-th column has been replaced with the column of constants appearing in the equations (i.e. *b*). Using the Cramer method to solve the system of two linear equations of the form:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \\ a_{21}x_1 + a_{22}x_2 &= b_2 \end{aligned} \tag{8.16}$$

we obtain the following equations:

$$x_1 = \frac{D_1}{D} \quad x_2 = \frac{D_2}{D} \tag{8.17}$$

where:

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \quad D_1 = \begin{vmatrix} b_1 & a_{12} \\ b_2 & a_{22} \end{vmatrix} \quad D_2 = \begin{vmatrix} a_{11} & b_1 \\ a_{21} & b_2 \end{vmatrix} \tag{8.18}$$

The 2×2 determinant of the coefficient matrix (*D*) can be solved as follows:

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \tag{8.19}$$

The above formulas provide a solution to the system of two linear equations by performing two multiplications and one subtraction.
With an increase in the size of the determinant, the number of necessary calculations also increases. For a 3×3 determinant it is necessary to perform the following operations:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{33}a_{21}a_{12} \tag{8.20}$$

On the right side of this equation there are $n!$=6 factors which are the product of n numbers, and thus the calculations require (n)(n!)=18 unit operations. Determinants of higher dimensions are reduced to the size of 3×3 using appropriate coefficients. It is possible to develop an algorithm that calculates nxn determinants and makes 2n! multiplications. For large values of n, however, this method becomes impractical, because, for example, calculating a 20×20 determinant requires approximately 10^{18} mathematical unit operations.

EXAMPLE

Solve the following system of linear equations:

$$\begin{aligned} x + y + z &= 6 \\ x + 2y + 3z &= 14 \\ x + 4y + 9z &= 36 \end{aligned}$$

by Cramer method.

ROZWIAZANIE

Using Equation (8.20) calculate the determinant of the coefficient matrix (D):

$$D = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{vmatrix} = 1\cdot 2\cdot 9 + 1\cdot 3\cdot 1 + 1\cdot 1\cdot 4 - 1\cdot 2\cdot 1 - 4\cdot 3\cdot 1 - 9\cdot 1\cdot 1 = 2$$

Thus, for D = 2, the subsequent determinants are:

$$x = \frac{1}{2}\begin{vmatrix} 6 & 1 & 1 \\ 14 & 2 & 3 \\ 36 & 4 & 9 \end{vmatrix} = 6\cdot 2\cdot 9 + 1\cdot 3\cdot 36 + 1\cdot 14\cdot 4 - 36\cdot 2\cdot 1 - 4\cdot 3\cdot 6 - 9\cdot 14\cdot 1 = 1$$

$$y = \frac{1}{2}\begin{vmatrix} 1 & 6 & 1 \\ 1 & 14 & 3 \\ 1 & 36 & 9 \end{vmatrix} = 1\cdot 14\cdot 9 + 6\cdot 3\cdot 1 + 1\cdot 1\cdot 36 - 1\cdot 14\cdot 1 - 36\cdot 3\cdot 1 - 9\cdot 1\cdot 6 = 2$$

$$z = \frac{1}{2}\begin{vmatrix} 1 & 1 & 6 \\ 1 & 2 & 14 \\ 1 & 4 & 36 \end{vmatrix} = 1\cdot 2\cdot 36 + 1\cdot 14\cdot 1 + 6\cdot 1\cdot 4 - 1\cdot 2\cdot 6 - 4\cdot 14\cdot 1 - 36\cdot 1\cdot 1 = 3$$

Solution of this system of equations are the values of" x = 1, y = 2 and z = 3.

8.3. Gauss–Seidel method

A system of linear equations can be solved using an iterative method which in each step of calculations gets progressively closer to the value of solution. One of the methods based on this algorithm is the Gauss–Seidel method which can be used for any system of equations (including nonlinear) given as follows:

$$P_i\left(x_1, x_2, \cdots x_n\right) = 0 \qquad i = 1, 2 \cdots n \tag{8.21}$$

which can be rearranged to the form:

$$x_i = f_i\left(x_1, x_2, \cdots x_n\right) = 0 \qquad i = 1, 2 \cdots n \tag{8.22}$$

In this method, it is necessary to know the initial approximate solution for all unknowns, i.e. the values:

$$x_1^{(0)}, x_2^{(0)}, \cdots x_n^{(0)} \tag{8.23}$$

According to this method, the first approximation of x_i is obtained from:

$$x_i^{(1)} = f_i\left(x_1^{(0)}, x_2^{(0)}, \cdots x_n^{(0)}\right) = 0 \qquad i = 1, 2 \cdots n \tag{8.24}$$

Similarly, successive approximations are obtained using the following recursive equation:

$$x_i^{(k+1)} = f_i\left(x_1^{(k)}, x_2^{(k)}, \cdots x_n^{(k)}\right) = 0 \qquad \begin{array}{l} i = 1, 2 \cdots n \\ k = 1, 2, \cdots \end{array} \tag{8.25}$$

If iteration convergence conditions are met, a group of numbers:

$$x_j^{(i)} \qquad j = 1, 2 \cdots n \tag{8.26}$$

from sequences:

$$x_j^{(1)} x_j^{(2)} x_j^{(3)} \cdots x_j^{(i)} \qquad j = 1, 2 \cdots n \tag{8.27}$$

with a sufficiently large *i* is arbitrarily accurate approximation of the solution of the system.
For instance, according to the general equation (8.14), the system of equations (dimension 3×3) can be written as:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3 \end{aligned} \tag{8.28}$$

and further rearranged to the form:

$$\begin{aligned}x_1 &= \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3)\\ x_2 &= \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3)\\ x_3 &= \frac{1}{a_{33}}(b_3 - a_{31}x_1 - a_{32}x_2)\end{aligned} \tag{8.29}$$

It is generally assumed that $x_i^{(0)} = b_i$. Using the estimated initial values $x_2^{(0)}$ and $x_3^{(0)}$, we can calculate (inserting them into the equation for x_1) the first approximate value of $x_1^{(1)}$. In the next step, the evaluated values of $x_1^{(1)}$ and $x_3^{(0)}$ are substituted in the equation for x to calculate the first approximate value of $x_2^{(1)}$. Then $x_1^{(1)}$ and $x_2^{(1)}$ calculated in this manner are used to calculate $x_3^{(1)}$ from the third equation. This procedure is repeated as long as a satisfactory accuracy of solutions is achieved. In this method, each cycle of calculations requires only n^2 multiplications.

The general equation for k+1 approximation of the i-th variable which allows solving any n×n system of equations can be put as follows:

$$x_i^{(k+1)} = \frac{1}{a_{ii}}\left[b_i - \sum_{j}^{i-1} a_{ij}x_j^{k+1} - \sum_{j=i+1}^{n} a_{ij}x_j^{k}\right] \tag{8.30}$$

EXAMPLE

Solve the following system of linear equations:

$$\begin{aligned}x + y + z &= 6\\ x + 2y + 3z &= 14\\ x + 4y + 9z &= 36\end{aligned}$$

by Gauss–Seidel method.

SOLUTION

According to the algorithm for the calculations discussed above, appropriate absolute terms are taken as the initial approximate solutions of equations, i.e.:

$x^{(0)} = b_1 = 6$, $y^{(0)} = b_2 = 7$ (because 2y = 14) and $z^{(0)} = b_3 = 4$ (because 9z = 36),

Substituting the values $y^{(0)}$ and $z^{(0)}$ into the rearranged first equation, the first approximate value of x is obtained:

$$x^{(1)} = 6 - y^{(0)} - z^{(0)} = 6 - 7 - 4 = -5$$

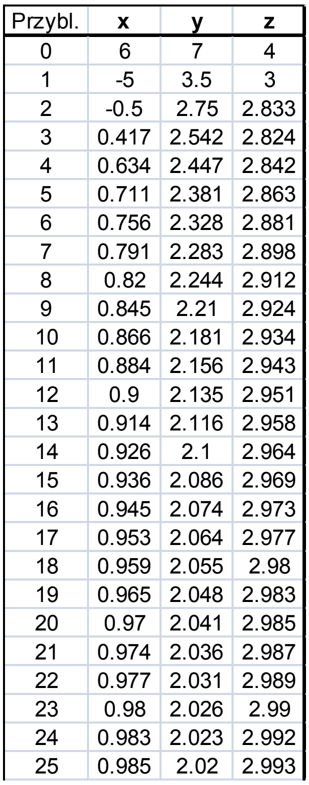
So determined values of $x^{(1)}$ and $z^{(0)}$ are inserted into the transformed second equation to calculate the first approximation of $y^{(1)}$:

$$y^{(1)} = \frac{(14 - x^{(1)} - 3z^{(0)})}{2} = \frac{(14 + 5 - 3\cdot 4)}{2} = 3.5$$

then the values for $x^{(1)}$ and $y^{(1)}$ are substituted into the rearranged third equation to obtain the first approximation of z:

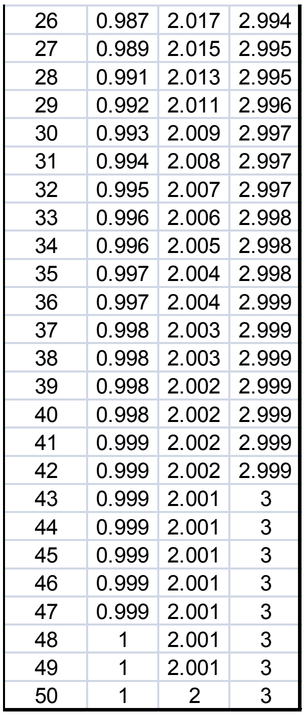
$$y^{(1)} = \frac{(36 - x^{(1)} - 4y^{(1)})}{9} = \frac{(36 + 5 - 4 \cdot 3.5)}{9} = 3$$

etc. The results are summarized below in the table.



Przybl.	x	y	z
0	6	7	4
1	-5	3.5	3
2	-0.5	2.75	2.833
3	0.417	2.542	2.824
4	0.634	2.447	2.842
5	0.711	2.381	2.863
6	0.756	2.328	2.881
7	0.791	2.283	2.898
8	0.82	2.244	2.912
9	0.845	2.21	2.924
10	0.866	2.181	2.934
11	0.884	2.156	2.943
12	0.9	2.135	2.951
13	0.914	2.116	2.958
14	0.926	2.1	2.964
15	0.936	2.086	2.969
16	0.945	2.074	2.973
17	0.953	2.064	2.977
18	0.959	2.055	2.98
19	0.965	2.048	2.983
20	0.97	2.041	2.985
21	0.974	2.036	2.987
22	0.977	2.031	2.989
23	0.98	2.026	2.99
24	0.983	2.023	2.992
25	0.985	2.02	2.993

cont.



26	0.987	2.017	2.994
27	0.989	2.015	2.995
28	0.991	2.013	2.995
29	0.992	2.011	2.996
30	0.993	2.009	2.997
31	0.994	2.008	2.997
32	0.995	2.007	2.997
33	0.996	2.006	2.998
34	0.996	2.005	2.998
35	0.997	2.004	2.998
36	0.997	2.004	2.999
37	0.998	2.003	2.999
38	0.998	2.003	2.999
39	0.998	2.002	2.999
40	0.998	2.002	2.999
41	0.999	2.002	2.999
42	0.999	2.002	2.999
43	0.999	2.001	3
44	0.999	2.001	3
45	0.999	2.001	3
46	0.999	2.001	3
47	0.999	2.001	3
48	1	2.001	3
49	1	2.001	3
50	1	2	3

The calculations indicate that only after 50 steps of iterative calculations a satisfactory result was obtained which was the solution of the above system of linear equations.
In practice, the number of iterations is chosen so that the difference between successive estimates was lower than the assumed error (ε):

$$\left| x_i^{(m)} - x_i^{(m-1)} \right| \leq \varepsilon \tag{8.31}$$

A very important disadvantage of this method is a very common lack of convergence which makes it impossible to achieve a correct solution .

8.4. Gauss–Jordan elimination method

Another way to solve systems of linear equations is the use of the Gauss–Jordan elimination method. It involves a gradual transformation of a augmented ("double") matrix of coefficients and an identity matrix (**A**|**I**) to **I**|**A**$^{-1}$ (identity matrix|inverse matrix) by means of operations on rows which include:

- replacement of any two rows,
- multiplying elements of a row by a constant different from zero,
- adding results of any row multiplication to another row.

Operations on rows of a coefficient matrix **A** (to transform **A** into **I**) are made parallel to the operations on a matrix **I**. As a result, the matrix **A** is transformed to **I** and **I** is transformed to **A**$^{-1}$, according to the expression:

$$\mathbf{A\,I} \rightarrow \mathbf{I\,A^{-1}} \tag{8.32}$$

After determining $\mathbf{A^{-1}}$ with the given method, the values of x can now be calculated simply by multiplying the matrix:

$$\mathbf{X} = \mathbf{A^{-1}B} \tag{8.33}$$

It is not necessary to calculate the invertible matrix and the solution of the matrix equation (8.33), if the matrix **A** expands to include the column matrix **B** and a matrix with the following structure is formed:

$$\mathbf{A}\big|\mathbf{B} = \begin{bmatrix} a_{11} & a_{12} & & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} & b_n \end{bmatrix} \tag{8.34}$$

Then after transforming the matrix **A** to the identity matrix **I**, we get the following matrix (**I**|**X**):

$$\begin{bmatrix} 1 & 0 & \cdots & 0 & x_1 \\ 0 & 1 & \cdots & 0 & x_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & x_n \end{bmatrix} \tag{8.35}$$

where the last column includes the required values of x_i.

EXAMPLE

Solve the following system of linear equations:

$$\begin{aligned} x + y + z &= 6 \\ x + 2y + 3z &= 14 \\ x + 4y + 9z &= 36 \end{aligned}$$

by Gauss–Jordan elimination method.

SOLUTION

In order to transform the matrix **A**|**B** into **I**|**X:**

$$\mathbf{A}\big|\mathbf{B} = \begin{vmatrix} 1 & 1 & 1 & 6 \\ 1 & 2 & 3 & 14 \\ 1 & 4 & 9 & 36 \end{vmatrix} \rightarrow \begin{vmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \end{vmatrix}$$

the following elementary row operations on the augmented matrix can be perform:

1) Subtracting elements of row 1 from row 2:

$$\begin{vmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 2 & 8 \\ 1 & 4 & 9 & 36 \end{vmatrix}$$

2) Subtracting elements of row 1 from row 3:

$$\begin{vmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 2 & 8 \\ 0 & 3 & 8 & 30 \end{vmatrix}$$

3) Subtracting elements of row 2, multiplied by 3, from row 3:

$$\begin{vmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 2 & 6 \end{vmatrix}$$

4) Dividing row 3 by 2:

$$\begin{vmatrix} 1 & 1 & 1 & 6 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$

5) Subtracting elements of row 2 from row 1:

$$\begin{vmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$

6) Adding elements of row 1 to row 3:

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 8 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$

7) Subtracting elements of row 2 from elements of row 3 multiplied by 2:

$$\begin{vmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$

The final results are therefore $x = 1$, $y = 2$ and $z = 3$ and are the solution of the above system of linear equations.

8.5. Newton–Raphson metod for nonlinear algebraic equations

The Newton–Raphson method can be used to seek approximate solutions of nonlinear algebraic equations. The general algorithm of this method is shown below.

In the Newton–Raphson method, for a system of n nonlinear algebraic equations of the form:

$$\phi_i\left(x_1, x_2 \cdots x_n\right) = b_i \qquad i = 1, 2 \cdots n \tag{8.36}$$

the initial approximations of unknowns values (x) are required:

$$x_1^{(0)}, x_2^{(0)} \cdots x_n^{(0)} \tag{8.37}$$

Assuming that:

$$\Delta x_1, \Delta x_2 \cdots \Delta x_n \tag{8.38}$$

are necessary corrections for the calculation of the first approximations, i.e.:

$$x_i^{(1)} = x_i^{(0)} + \Delta x_i \tag{8.39}$$

system of equations (8.36) can be written as follows:

$$\phi_i\left(x_1^{(0)}+\Delta x_1, x_2^{(0)}+\Delta x_2\cdots x_n^{(0)}+\Delta x_n\right)=b_i \tag{8.40}$$

Expanding a function φ_i into a Taylor series we obtain:

$$\phi_i\left(x_1^{(0)}, x_2^{(0)}\cdots x_n^{(0)}\right)+\frac{\delta\phi_1}{\delta x_1}\Big|_0\Delta x_1+\frac{\delta\phi_2}{\delta x_2}\Big|_0\Delta x_2+\cdots\frac{\delta\phi_n}{\delta x_n}\Big|_0\Delta x_n\cdots \tag{8.41}$$

Taking into account only the first two terms in the Taylor series, the following system of *n* equations can be obtained:

$$\phi_i\left(x_1^{(0)}, x_2^{(0)}\cdots x_n^{(0)}\right)+\sum_{i=1}^{n}\frac{\delta\phi_i}{\delta x_i}\Big|_0\Delta x_i=b_i \tag{8.42}$$

In matrix notation equation (8.42) can be expressed as:

$$\begin{bmatrix} \left.\frac{\delta\phi_1}{\delta x_1}\right|_0 & \left.\frac{\delta\phi_2}{\delta x_2}\right|_0 & \cdots & \left.\frac{\delta\phi_n}{\delta x_n}\right|_0 \\ \vdots\left.\frac{\delta\phi_2}{\delta x_1}\right|_0 & \left.\frac{\delta\phi_2}{\delta x_2}\right|_0 & \cdots & \left.\frac{\delta\phi_2}{\delta x_n}\right|_0 \\ \vdots & \vdots & \vdots & \vdots \\ \left.\frac{\delta\phi_n}{\delta x_1}\right|_0 & \left.\frac{\delta\phi_n}{\delta x_2}\right|_0 & \cdots| & \left.\frac{\delta\phi_n}{\delta x_n}\right|_0 \end{bmatrix}\begin{bmatrix}\Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n\end{bmatrix}=\begin{bmatrix} b_1-\phi_1\left(x_1^{(0)}, x_2^{(0)}\cdots x_n^{(0)}\right) \\ b_2-\phi_2\left(x_1^{(0)}, x_2^{(0)}\cdots x_n^{(0)}\right) \\ \vdots \\ b_3-\phi_3\left(x_1^{(0)}, x_2^{(0)}\cdots x_n^{(0)}\right)\end{bmatrix} \tag{8.43}$$

The above equations in matrix form can be written in the form:

$$\mathbf{AX} = \mathbf{B} \tag{8.44}$$

The solution of this matrix equation with respect to Δx enables to find the successive approximations of the solution to a nonlinear system of equations (8.36).

EXAMPLE

Solve the following system of equations by Newton–Raphson method:

$$\begin{aligned} f &= y^2+8x-16=0 \\ g &= 2y+x-4=0 \end{aligned}$$

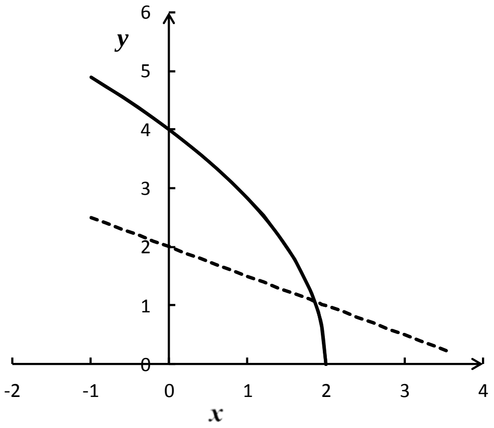


Fig. 8.1. Graph of the system of equations.

SOLUTION

According to the presented figure, as the initial approximations the values:

$$x^{(0)} = 2, \quad y^{(0)} = 2$$

were applied.

In the first step of calculations it is necessary to calculate the values of the functions and their derivatives (with respect to each variable) at the starting point:

$$f\left(x^{(0)}, y^{(0)}\right) = y^2 + 8x - 16 = 4 + 16 - 16 = 4$$
$$g\left(x^{(0)}, y^{(0)}\right) = 2y + x - 4 = 4 + 2 - 4 = 2$$
$$\left.\frac{\delta f}{\delta x}\right|_{x(0)} = 8 \quad \left.\frac{\delta f}{\delta y}\right|_{y(0)} = 2y = 4$$
$$\left.\frac{\delta g}{\delta x}\right|_{x(0)} = 1 \quad \left.\frac{\delta g}{\delta y}\right|_{y(0)} = \quad 2$$

According to the general equation (8.43), the matrix equation for the analysed case can be given in the form:

$$\begin{bmatrix} 8 & 4 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -4 \\ -2 \end{bmatrix}$$

Solving this equation with the Cramer's method allows the calculation of desired corrections:

$$D = 16 - 4 = 12$$
$$\Delta x = \frac{1}{12} \begin{vmatrix} -4 & 4 \\ -2 & 2 \end{vmatrix} = 0$$
$$\Delta y = \frac{1}{12} \begin{vmatrix} 8 & -4 \\ 1 & -2 \end{vmatrix} = -1$$

The calculated corrections ($\Delta x = 0$, $\Delta y = -1$) are used to determine the first approximations of the desired variables *x*:

$$x^{(1)} = x^{(0)} + \Delta x = 2$$
$$y^{(1)} = y^{(0)} + \Delta y = 1$$

In the second step, similar calculations lead to:

$$f\left(x^{(1)}, y^{(1)}\right) = 1 + 16 - 16 = 1$$
$$g\left(x^{(1)}, y^{(1)}\right) = 2 + 2 - 4 = 0$$
$$\left.\frac{\delta f}{\delta x}\right|_{x(1)} = 8 \quad \left.\frac{\delta f}{\delta y}\right|_{y(1)} = 2$$
$$\left.\frac{\delta g}{\delta x}\right|_{x(1)} = 1 \quad \left.\frac{\delta g}{\delta y}\right|_{y(1)} = 2$$

while the matrix equation can be expressed as:

$$\begin{bmatrix} 8 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

The second corrections calculated with the Cramer's method (Δx = –0.1428, Δy = 0.0714) are used to determine the second approximations of the desired variables *x*:

$$x^{(2)} = x^{(1)} + \Delta x = 1.8572$$
$$y^{(2)} = y^{(1)} + \Delta y = 1.0714$$

In the third step, the similar calculation algorithm leads to:

$$f\left(x^{(1)}, y^{(1)}\right) = 1 + 16 - 16 = 0.051$$
$$g\left(x^{(1)}, y^{(1)}\right) = 2 + 2 - 4 = 0$$
$$\left.\frac{\delta f}{\delta x}\right|_{x(1)} = 8 \quad \left.\frac{\delta f}{\delta y}\right|_{y(1)} = 2.143$$
$$\left.\frac{\delta g}{\delta x}\right|_{x(1)} = 1 \quad \left.\frac{\delta g}{\delta y}\right|_{y(1)} = 2$$

the matrix equation in the form:

$$\begin{bmatrix} 8 & 2.143 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} -0.051 \\ 0 \end{bmatrix}$$

the third corrections:

$$\Delta x = -0.00074, \ \Delta y = 0.000368$$

the third approximations of the desired variables *x*:

$$x^{(3)} = x^{(2)} + \Delta x = 1.8564$$
$$y^{(3)} = y^{(2)} + \Delta y = 1.0717$$

This procedure is repeated until a required approximation is achieved, i.e. the condition described by the equation (8.31) is fulfilled.

9. Interpolation

Interpolation is a process of determining values of a function $f(x)$ anywhere in an interval (x_0, x_n), when values of the function at points x_0, x_1, x_2,... x_n are known. The overall objective of interpolation is therefore to find a function which allows approximate determination of intermediate points between (x_0, y_0) and (x_n, y_n). The interpolation equations can also be used for numerical calculation of integrals and differential equations.

Polynomial interpolation is one of the simplest and most commonly used methods. If we know (n+1) the values of the function y_i at (n+1) the nodes of interpolation (x_0, x_1, x_2,... x_n), we can construct (by analogy to the problem of approximation) an adequate interpolation polynomial in the form:

$$f(x) = a_0 + a_1 x + a_2 x^2 \cdots a_n x^n \tag{9.1}$$

The polynomial described by the equation (9.1) at the nodes of interpolation takes values equal to the precisely known values of y_i, unlike the approximating function for which the estimated value of $\hat{y}_i$ is subject to an error arising from the method of least squares applied. At the same time, for the given (n+1) points there is exactly one at most n-th degree polynomial whose graph passes through these points. For example, exactly one straight line (n=1) passes through every two points, etc.

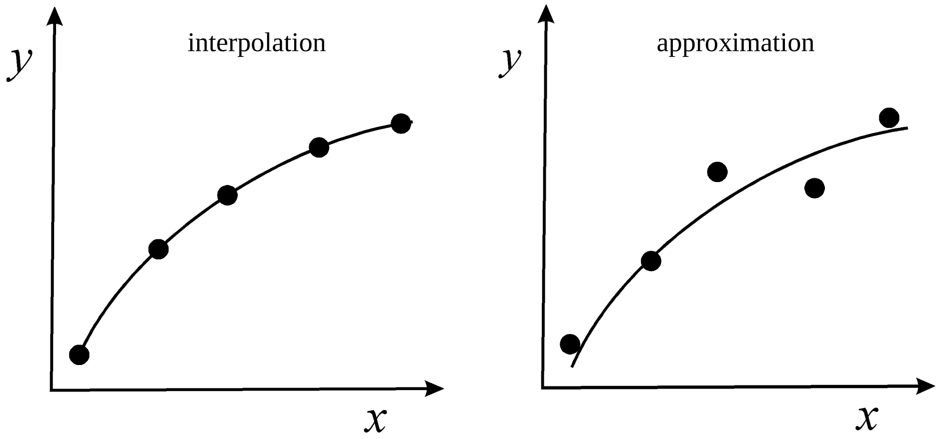


Fig. 9.1. Difference between interpolation and approximation.

9.1. Lagrange interpolation formula

Assuming that we know n+1 values of the function $y = f(x)$ ($f(x_0)$, $f(x_1)$,..., $f(x_n)$), we can develop n + 1 algebraic equations in the following way:

$$\begin{array}{l} f(x_0) = a_0 + a_1 x_0 + a_2 x_0^2 + \cdots a_n x_0^n \\ f(x_1) = a_0 + a_1 x_1 + a_2 x_1^2 + \cdots a_n x_1^n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ f(x_n) = a_0 + a_1 x_n + a_2 x_n^2 + \cdots a_n x_n^n \end{array} \tag{9.2}$$

In the above system of equations, there are n+1 unknowns (a_0, a_1,..., a_n coefficients of the interpolating polynomial), while relations between the coefficients a_i and the values of x and y always satisfy the general relation resulting from the formula for interpolation polynomial. This system can be put in the form of determinant (equal zero), which allows derivation of any interpolation polynomial:

$$\begin{vmatrix} f(x) & 1 & x & x^2 & \cdots & x^n \\ f(x_0) & 1 & x_0 & x_0^2 & \cdots & x_0^n \\ f(x_1) & 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ f(x_n) & 1 & x_n & x_n^2 & \cdots & x_n^n \end{vmatrix} = 0 \qquad (9.3)$$

Any interpolation polynomial can be determined using the following Lagrange interpolation formula

$$f(x) = \sum_{j=0}^{n} f(x_j) \frac{\prod_{\substack{i=0 \\ i\neq j}}^{n} (x - x_i)}{\prod_{\substack{i=0 \\ i\neq j}}^{n} (x_j - x_i)} \qquad (9.4)$$

For the simplest case (first-degree polynomial), that is a straight line strictly passing through two interpolation nodes (points (x_0, y_0) and (x_1, y_1) Fig. 9.2), a corresponding linear interpolation polynomial is given by:

$$f(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) \qquad (9.5)$$

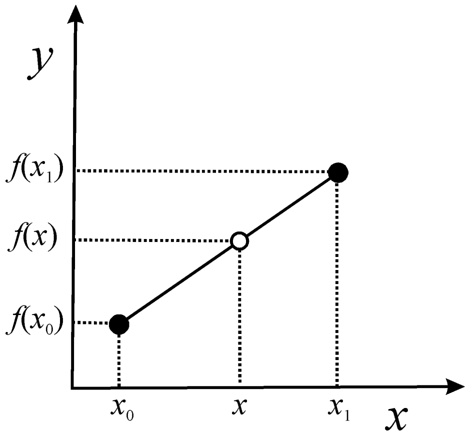


Fig. 9.2. Linear interpolation.

For the second-degree polynomial interpolation, from Equation (9.4) we get:

$$f(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f(x_1) + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f(x_2) \qquad (9.6)$$

that is, the expression for the parabolic interpolation (Fig. 9.3).

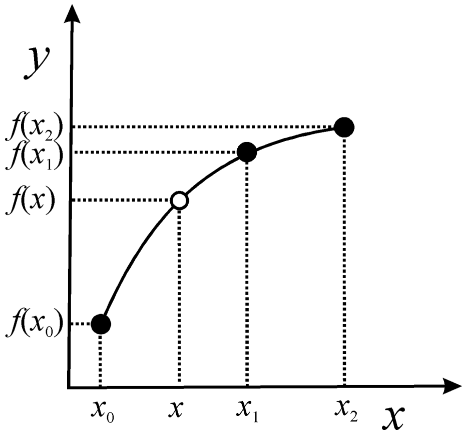


Fig. 9.3. Parabolic interpolation.

The calculations are significantly simplified for equidistant interpolation nodes on the interval:

$$x_{j+1} - x_j = h \tag{9.7}$$

for which, a new variable defined by the formula:

$$x = x_m = x_0 + hm \qquad (m = 1, 2, \cdots n) \tag{9.8}$$

can be introduced.
In this case:

$$\begin{aligned} &x - x_0 = hm \\ &x - x_1 = h(m-1) \\ &x - x_2 = h(m-2) \\ &etc. \end{aligned} \tag{9.9}$$

The corresponding interpolation equations (Equation (9.5) and (9.6)) takes the following form:
for n = 1:

$$f(x + hm) \equiv f(m) = \frac{h(m-1)}{-h} f(x_0) + \frac{hm}{h} f(x_1) = (1-m) f(x_0) + mf(x_1) \tag{9.10}$$

and for n = 2:

$$\begin{aligned} &f(m) = \frac{h^2(m-1)(m-2)}{(-h)(-2h)} f(x_0) + \frac{h^2(m)(m-2)}{h(-h)} f(x_1) + \frac{h^2(m)(m-1)}{(2h)(h)} f(x_2) = \\ &\frac{1}{2}(m-1)(m-2) f(x_0) - m(m-2) f(x_1) + \frac{1}{2} m(m-1) f(x_2) \end{aligned} \tag{9.11}$$

EXAMPLE

Calculate the interpolation polynomial (using the Lagrange interpolation formula) for the data presented in the table:

x	y
1	1
2	3
3	7

SOLUTION

Substituting the numerical values of x and y into Equation (9.6) we get:

$$f(x)=\frac{(x-2)\cdot(x-3)}{(1-2)\cdot(1-3)}\cdot 1+\frac{(x-1)\cdot(x-3)}{(2-1)\cdot(2-3)}\cdot 3+\frac{(x-1)\cdot(x-2)}{(3-1)\cdot(3-2)}\cdot 7$$

After multiplication:

$$f(x)=\frac{7}{2}x^2-\frac{21}{2}x+7+\left(-3x^2+12x-9\right)+\frac{1}{2}x^2-\frac{5}{2}x+3$$

Which leads to a interpolation polynomial in the form:

$$f(x)=x^2-x+1$$

This task can be calculated using the simplified equation (9.11) which after inserting the numerialc values takes the following form:

$$f(m)=\frac{1}{2}(m-1)(m-2)\cdot 1-m(m-2)\cdot 3+\frac{1}{2}m(m-1)\cdot 7$$

After performing the appropriate mathematical operations:

$$f(m)=\frac{1}{2}m^2-\frac{3}{2}m+1-3m^2+6m+\frac{7}{2}m^2-\frac{7}{2}m$$

leads to the equation in the form:

$$f(m)=m^2+m+1$$

In view of the fact that for the presented example, $x-x_0=hm$ and *h* = 1, the relationship between *x* and *m* is as follows:

$$x-1=m$$

Substituting this equation into the final equation for *f*(*m*), we obtain:

$$f(x)=x^2-x+1$$

which is identical with the equation derived from Equation (9.6).

9.2. Differences and divided differences

Another method for creating interpolation polynomials uses the concept of progressive (forward) differences or divided differences.
In the case of equidistant interpolation nodes, corresponding differences ($\varDelta y$) between function values *f*(*x*):

$$\varDelta y_0=f\left(x_1\right)-f\left(x_0\right),\ \varDelta y_1=f\left(x_2\right)-f\left(x_1\right),\cdots \varDelta y_{n-1}=f\left(x_n\right)-f\left(x_{n-1}\right) \qquad (9.12)$$

are called first order progressive (forward) differences of a function $f(x)$. Similarly, higher–order progressive differences are defined as:

$$\mathit{\Delta}^n y = \mathit{\Delta}(\mathit{\Delta}^{n-1} y) \tag{9.13}$$

Ways of building the progressive differences and relationships between them are shown in Table 9.1. The relationship between differences of a certain degree can be presented by the equations:

$$\begin{aligned}
&\mathit{\Delta} y_0 = f(x_1) - f(x_0) \\
&\mathit{\Delta}^2 y_0 = \mathit{\Delta} y_1 - \mathit{\Delta} y_0 = f(x_2) - 2f(x_1) + f(x_0) \\
&\mathit{\Delta}^3 y_0 = \mathit{\Delta}^2 y_1 - \mathit{\Delta}^2 y_0 = f(x_3) - 3f(x_2) + 3f(x_1) - f(x_0) \\
&\mathit{\Delta}^4 y_0 = \mathit{\Delta}^3 y_1 - \mathit{\Delta}^3 y_0 = f(x_4) - 4f(x_3) + 6f(x_2) - 4f(x_1) + f(x_0)
\end{aligned} \tag{9.14}$$

Tab. 9.1. Progressive differences and relationships between them

x	$f(x)$	First forward difference	Second forward difference	Third forward difference	Fourth forward difference
x_0	$f(x_0)$				
		$\mathit{\Delta} y_0$			
x_1	$f(x_1)$		$\mathit{\Delta}^2 y_0$		
		$\mathit{\Delta} y_1$		$\mathit{\Delta}^3 y_0$	
x_2	$f(x_2)$		$\mathit{\Delta}^2 y_1$		$\mathit{\Delta}^4 y_0$
		$\mathit{\Delta} y_2$		$\mathit{\Delta}^3 y_1$	
x_3	$f(x_3)$		$\mathit{\Delta}^2 y_2$		
		$\mathit{\Delta} y_3$			
x_4	$f(x_4)$				

In the general case, when x takes on any value, or the differences defined by Equation (9.12) are not fixed, we should use divided differences.

First–order divided differences of a function $f(x)$ are the expressions defined as follows:

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \qquad f[x_1, x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \quad f[x_2, x_3] = \frac{f(x_3) - f(x_2)}{x_3 - x_2} \tag{9.15}$$

Second–order divided differences are defined by analogy:

$$f\left[x_{0,} x_1, x_2\right] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}, \quad f\left[x_{1,} x_2, x_3\right] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1} \tag{9.16}$$

Overall, a n–order divided difference is created from an n–1 –order divided difference using the following recursive formula:

$$f\left[x_{0,} x_1, \cdots, x_{n-1}, x_n\right] = \frac{f[x_1, x_2, \cdots, x_n] - f[x_0, x_1, \cdots, x_{n-1}]}{x_n - x_0} \tag{9.17}$$

Comparing Equations (9.12) for first–order progressive differences with first–order divided differences for equidistant interpolation nodes (Equations (9.15)), we obtain the following relations:

$$f\left[x_0, x_1\right] = \frac{\Delta y_0}{h}, \quad f\left[x_1, x_2\right] = \frac{\Delta y_1}{h}, \cdots, f\left[x_{n-1}, x_n\right] = \frac{\Delta y_{n-1}}{h} \tag{9.18}$$

where *h* is the distance between the *x* points, defined by Equation (9.7).

9.3. Newton's interpolation formula

The interpolation polynomial (Equation (9.5)) can be written in the form:

$$f(x) = f(x_0) + (x - x_0)\left(\frac{y_1 - y_0}{x_1 - x_0}\right) \tag{9.20}$$

Substituting into this formula the expression for the first–order divided difference (Equation (9.15)) we get:

$$f(x) = f(x_0) + (x - x_0) f[x_0, x_1] \tag{9.21}$$

A similar transformation of Equation (9.16) for the second–order divided difference, leads to the relationship:

$$f(x) = f(x_0) + (x - x_0) f[x_0, x_1] + (x - x_0)(x - x_1) f[x_0, x_1, x_2] \tag{9.22}$$

In general, the expression:

$$\begin{aligned} & f_n(x) = f(x_0) + (x - x_0) f[x_0, x_1] \\ & +(x - x_0)(x - x_1) f[x_0, x_1, x_2] \\ & +(x - x_0)(x - x_1)(x - x_2) f[x_0, x_1, x_2, x_3] \\ & \vdots \\ & +(x - x_0)(x - x_1)\cdots(x - x_{n-1}) f[x_0, x_1, \cdots, x_n] \end{aligned} \tag{9.23}$$

is a interpolation polynomial *f*(*x*) for *n* + 1 interpolation nodes.This formula is called the *Newton's interpolation formula for unequal intervals* or *Newton's divided difference interpolation formula*.

EXAMPLE

Calculate the interpolation polynomial (using the Newton's divided difference interpolation formula) for the data presented in the table:

x	*y*
1	1
2	3
3	7

SOLUTION

In order to use Equation (9.22), two of the first–order divided differences should be calculated:

$$f[x_0,x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{3-1}{2-1} = 2$$

$$f[x_1,x_2] = \frac{f(x_2) - f(x_1)}{x_2 - x_1} = \frac{7-3}{3-2} = 4$$

and second–order divided difference:

$$f\left[x_{0,}x_1,x_2\right] = \frac{f[x_1,x_2] - f[x_0,x_1]}{x_2 - x_0} = \frac{4-2}{3-1} = 1$$

Substituting the calculated values into Equation (9.22), we obtain:

$$f(x) = 1 + (x-1)\cdot 2 + (x-1)(x-2)\cdot 1$$

and after multiplication:

$$f(x) = x^2 - x + 1$$

EXAMPLE

Given the values of $x_0 = 0$, $y_0 = 4$ and $x_1 = 2$, $y_1 = 6$ (Fig. 9.4), calculate the interpolated value for $x = 1$, using the Newton's interpolation formula.

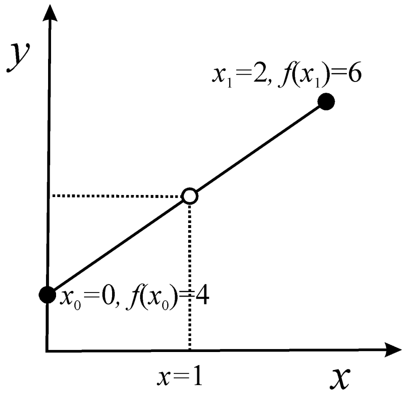


Fig. 9.4. Linear interpolation for presented example.

SOLUTION

The first–order divided difference is equal to:

$$f[x_0,x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{6-4}{2-0} = 1$$

Substituting the values in the Newton’s interpolation formula (Equation (9.22)), we obtain:

$$f(x) = 4 + (1-0)\cdot 1 = 5$$

The interpolated value for x = 1 is equal to 5.

EXAMPLE

Given the values of ($x_0 = 0$, $y_0 = 4$), ($x_1 = 2$, $y_1 = 6$) and ($x_2 = 4$, $y_2 = 6.5$) (Fig. 9.5), calculate the interpolated value for $x = 1$, using the Newton's interpolation formula.

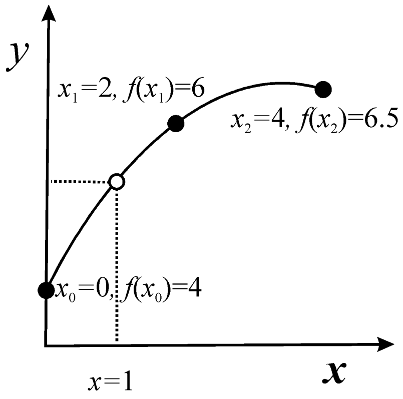


Fig. 9.5. Parabolic interpolation for presented example.

SOLUTION

In this example, the first-order divided differences:

$$f\left[x_0,x_1\right]=\frac{f(x_1)-f(x_0)}{x_1-x_0}=\frac{6-4}{2-0}=1$$

$$f\left[x_1,x_2\right]=\frac{f(x_2)-f(x_1)}{x_2-x_1}=\frac{6.5-6}{4-2}=0.25$$

and second-order divided difference:

$$f\left[x_{0,}x_1,x_2\right]=\frac{f\left[x_1,x_2\right]-f\left[x_0,x_1\right]}{x_2-x_0}=\frac{4-2}{3-1}=1$$

should be calculated.

Substituting the appropriate values in the Newton’s interpolation formula (Equation (9.22)), we obtain:

$$f(x)=4+(1-0)\cdot 1+(1-0)(1-2)\cdot(-0.1875)=5.1875$$

Therefore the interpolated value for $x = 1$ is equal to 5.1875.

9.4. Numerical differentiation

Interpolation functions can be used for numerical differentiation. This is especially useful when the function is given in the form of a table for a set of certain values of an independent variable, or if the analytical solution function is too complex.

Differentiating the Lagrange interpolation formula with respect to m in the form:

$$f(m)=\frac{1}{2}(m-1)(m-2)f(x_0)-(m)(m-2)f(x_1)+\frac{1}{2}m(m-1)f(x_2) \qquad (9.24)$$

the following equation is obtaining:

$$\frac{df}{dm}=\frac{1}{2}(2m-3)f(x_0)-2(m-1)f(x_1)+\frac{1}{2}(2m-1)f(x_2) \qquad (9.25)$$

Because for equidistant interpolation nodes can be written that:

$$x = x_m = x_0 + hm \qquad (m = 1, 2, \cdots n) \tag{9.26}$$

therefore:

$$\frac{df}{dx} = \frac{df}{dm}\frac{dm}{dx} = \frac{1}{h}\frac{df}{dm} \tag{9.27}$$

Substituting Equation (9.25) into Equation (9.27), the final expression for the numerical differentiation is obtained:

$$\frac{df}{dx} = \frac{1}{h}\left[\frac{1}{2}(2m-3)\,f(x_0) - 2(m-1)\,f(x_1) + \frac{1}{2}(2m-1)\,f(x_2)\right] \tag{9.28}$$

EXAMPLE

Given the values of ($x_0 = 0$, $y_0 = 4$), ($x_1 = 2$, $y_1 = 6$) and ($x_2 = 4$, $y_2 = 6.5$) (Fig. 9.6), calculate the derivatives at the points x_0, x_1 i x_2.

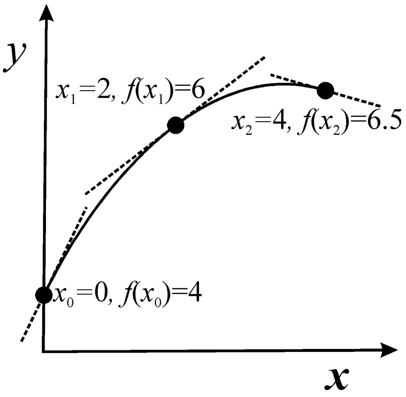


Fig. 9.6. Geometric interpretation of the differentiation of a parabola.

SOLUTION

In order to calculate the derivatives at the interpolation nodes, Equation (9.28) can be used. The performed calculation for the appropriate points are as follows:

a) x_0 = 0, y_0 = 4, m = 0, h = 2

$$\left(\frac{df}{dx}\right)_{x_0} = \frac{1}{h}\left[\frac{1}{2}(-3)\,f(x_0) + 2\,f(x_1) - \frac{1}{2}\,f(x_2)\right] = \quad \frac{1}{2}(-6+12-3.25) = 1.375$$

b) x_1 = 2, y_1 = 6, m =1, h = 2

$$\left(\frac{df}{dx}\right)_{x_1} = \frac{1}{h}\left[-\frac{1}{2}\,f(x_0) + \frac{1}{2}\,f(x_2)\right] = \quad \frac{1}{2}(-2+3.25) = 0.625$$

c) x_2 = 4, y_2=6.5, m = 2, h = 2

$$\left(\frac{df}{dx}\right)_{x_2} = \frac{1}{h}\left[\frac{1}{2}\,f(x_0) - 2\,f(x_1) + \frac{3}{2}\,f(x_2)\right] = \quad \frac{1}{2}(2-12+9.375) = -0.125$$

A similar values can be obtained using Newton's formula for equidistant interpolation nodes in the form:

$$f(m) = f(x_0) + m\Delta y_0 + \frac{1}{2}m(m-1)\Delta^2 y_0 \tag{9.29}$$

which leads to the following equation:

$$\frac{df}{dx} = \frac{1}{h}\frac{df}{dm} = \frac{1}{h}\left[\Delta y_0 + \frac{1}{2}(2m-1)\Delta^2 y_0)\right] \tag{9.30}$$

In practical applications, the presented numerical differentiation methods can be used only if the function values at the interpolation nodes are absolutely precise, i.e. an approximate polynomial must take a definite value for a given *x*. Therefore, the degree of the interpolation polynomial must be no less and no greater than the number of the nodes minus 1. Interpolation is usually used for a small number of measurement points. For the analysis of many experimental results, a simpler method is to determine a regression equation (e.g. polynomial) which then can be easily differentiated.

10. Optimization methods

Among many numerical methods used in optimization, this chapter briefly characterizes the methods of changing a single parameter and random walk. Because of the most common practical application, special attention is devoted to the grid search method (factorial designs) and the simplex method.

10.1. Method of changing a single parameter

One of the simplest optimization methods is to study changes in experimental response to the studied phenomenon, involving a change of a single parameter when the other variables are constant. After the extremum of a function in a given direction is found (e.g. for variable x_1), a constant value for this variable is taken, and the study continues in the direction of another variable (e.g. x_2). Once these steps are taken for all variables, we return to the first, optimized parameter (variable) and the whole procedure is carried out again. This method can give good results if variables are not correlated

10.2. Random walk method

In this method, every parameter is chosen at random, so the response surface is sampled extensively. This creates an opportunity to find the area of the extremum, however, it is also considered costly and time consuming due to the need for a large number of experiments.

10.3. Grid search method (factorial design)

The method consists in constructing a grid (design) that uses many different sets of values chosen for the optimization of variables. The grid covers a large part of the response surface and is used for grid testing. After the relevant experiments are done, close to the optimum values, we can make another design to clarify the optimal parameter values. Multiple repetition of this procedure leads to gradually better estimates of optimal values.

Factorial designs can also be used to determine a regression equation (regression model) describing the optimum area and allowing analytical calculation of the extreme point. Despite the need to conduct many experiments the number of which grows rapidly with increasing number of the variables under consideration, this method leads to good results.

10.3.1. Rules for creating a regression model

In order to find an equation describing a phenomenon, various kinds of empirical regression models may be used. In contrast to models fully specified for which the mathematical form of a function resulting for example from physical chemistry is known, in the case of an empirical model the functional dependence is not known. In the simplest case a multiple linear regression model can be used to describe a phenomenon:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots + a_p \cdot x_p \tag{10.1}$$

where y is a variable which is the experimental response of the studied phenomenon, x_i – independent (or explanatory) variable, a_i – coefficients that must be evaluated, while p is the number of independent variables.

If the linear model is insufficient to describe a response of an object, it is necessary to use more complex models such as a linear model with interaction terms between the factors, i.e. the products of various independent variables. For the two independent variables, this model takes the form:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_{12} \cdot x_1 \cdot x_2 \tag{10.2}$$

For the three independent variables, this model can be written as:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_3 + a_{12} \cdot x_1 \cdot x_2 + a_{13} \cdot x_1 \cdot x_3 + a_{23} \cdot x_2 \cdot x_3 \tag{10.3}$$

and in the general case as:

$$y = a_0 + \sum_{i=1}^{p} a_i \cdot x_i + \sum_{i=1}^{p} \sum_{j>n} a_{ij} \cdot x_i \cdot x_j \tag{10.4}$$

The equation (10.4) contains $1 + p + \frac{p(p-1)}{2}$ coefficients, so we should have a number at least equal to this value in order to estimate the coefficients a. For the linear model (Equation (10.1)), the number of measurements sufficient to determine the coefficients is just $p + 1$.

The linear model can be expanded by including square terms. In this case, a simplified square model can be put in the form:

$$y = a_0 + \sum_{i=1}^{p} a_i x_i + \sum_{i=1}^{p} a_{ii} x_i^2 \tag{10.5}$$

and determination of $2p + 1$ coefficients in this equation requires at least the same number of measurements, i.e. $n \geq 2p + 1$.

An extension of the simplified square model (Equation (10.5)) is a full square model which covers all possible interaction terms. Its use is limited due to a very large number of coefficients ($2^p + p$).

An interactive square model which is a combination of models (10.4) with (10.5) and contains only $2p + \frac{p(p-1)}{2} + 1$ coefficients is much more often used:

$$y = a_0 + \sum_{i=1}^{p} a_i x_i + \sum_{i=1}^{p} \sum_{j>n} a_{ij} \cdot x_i \cdot x_j + \sum_{i=1}^{p} a_{ii} x_i^2 \tag{10.6}$$

To identify the model, i.e. determine the coefficients, the minimum number of measurements to be performed is equal to the number of coefficients. Table 10.1 lists the minimum number of measurements necessary to identify the model.

Tab. 10.1. Minimum number of measurements necessary to identify the model

Model	Linear	Linear with interaction terms	Square	Interactive square	Full square
	p+1	1+p+p(p-1)/2	2p+1	1+2p+p(p-1)/2	2^p+p
p = 1	2	2	3	3	3
2	3	4	5	6	6
3	4	7	7	10	11
4	5	11	9	15	20
5	6	16	11	21	37
6	7	22	13	28	70
7	8	29	15	36	135

In practical calculations, the number of measurements should be much larger (at least 4–5 times) than the minimum values. A simple linear model with four independent variables requires at least 20−25 measurements. By careful selection of measurement points, in accordance with the principles of chemometrics, it is possible to obtain sufficiently good estimates of model coefficients for a much smaller number of measurements. A careful planning of these points is therefore required in this case.

10.3.2. Experimental design

A development of a model describing a mathematical correlation between results and the values of various parameters (i.e. close to the optimum) requires a proper planning. To do this, we can use a variety of experimental designs. Due to the fact that in experiments we usually deal with a variety of variables falling within various ranges, the so−called coded variables (u_i) rather than real variables (x_i) are used. Coded variables can be converted to real variables (and vice versa) using the following formulas:

$$u_i = \frac{2x_i - \left(x_{i,\max} + x_{i,\min}\right)}{\left(x_{i,\max} - x_{i,\min}\right)} \tag{10.7}$$

$$x_i = \frac{u_i\left(x_{i,\max} - x_{i,\min}\right) + \left(x_{i,\max} + x_{i,\min}\right)}{2} \tag{10.8}$$

where $x_{i,min}$ and $x_{i,max}$ denote the minimum and maximum values of the independent variable (x_i). These are also pre−established, acceptable ranges of variables. For most experimental designs, coded variables range from −1 to +1.

The so called optimality is one of the most important properties that an experimental design should have. The optimality criterion adopted in chemometrics is the possibility of predicting values of a dependent variable at different points in a design space. Therefore, an optimal design is a design which at a given number of experiments n guarantees the best prediction of a dependent variable in a design space. To obtain an optimal design, it is necessary to create such a design matrix **U** (containing coded variables u_{ij}) so that the elements of the main diagonal of a dispersion matrix (**D**):

$$\mathbf{D} = (\mathbf{U}^{\mathbf{T}}\mathbf{U})^{-1} \tag{10.9}$$

were as small as possible. The design matrix **U** contains a set of design coordinates presented by the coded variables. The lower elements of the main diagonal of the dispersion matrix, the greater the determinant of the so−called information matrix ($\mathbf{U}^{\mathbf{T}}\mathbf{U}$):

$$\det(\mathbf{U}^{\mathbf{T}}\mathbf{U}) = |\mathbf{U}^{\mathbf{T}}\mathbf{U}| \tag{10.10}$$

The determinant of an information matrix is large when we use orthogonal design variables, i.e. variables that are not correlated. At the same time, the use of orthogonal variables causes that a product of any two variables is also orthogonal to all other independent variables. Additionally, the estimated in this case coefficients of the model are independent of each other and thus removing any, statistically insignificant term of the model does not result in the need to re−calculate all the coefficients of the model.

The so called full factorial designs 2^m are among the simplest designs that meet the optimality criterion. The design includes *m* coded (explanatory) variables, each of which takes only two different values, i.e. two levels (e.g. −1, +1). These types of designs are used to describe relations (regression coefficients evaluation) for linear models and linear models with interaction terms. Examples of measurement point distribution for the full factorial design 2^2 and 2^3 are shown in Figure 10.1.

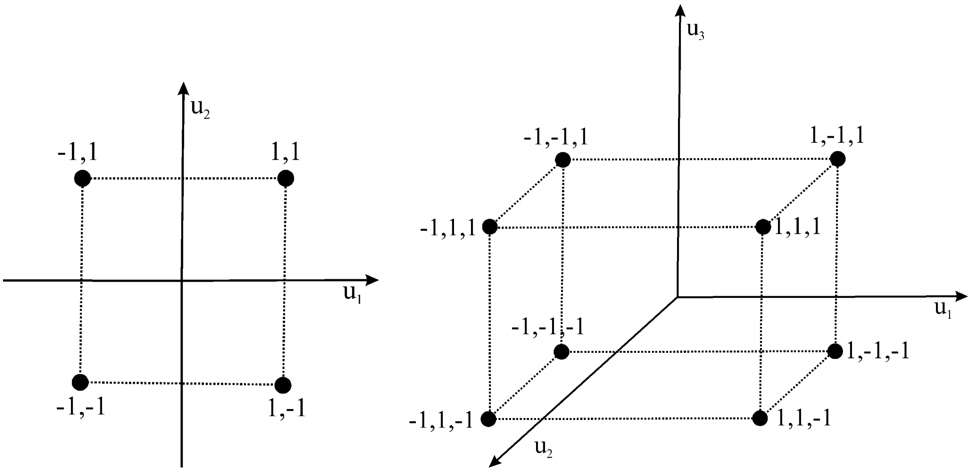


Fig. 10.1. Examples of measurement point distribution for the full factorial design 2^2 (2 levels and 2 number of factors) and 2^3 (2 levels and 3 number of factors).

In order to generate a full factorial design, we can use a very simple iterative method. In the first step, starting from the rightmost column, which corresponds to the *p*–th variable, two values are entered (−1 and +1), because it is a two−level design:

u_p
1
−1

In this way, the 2^1 design has been generated. In the second step, to include another variable, the generated design is copied in the bottom lines of the column u_p, and in the preceding column ($u_{p\text{-}1}$) the values of 1 are placed in the first half of the rows, while − 1 in the second:

$u_{p\text{-}1}$	u_p
1	1
1	−1
−1	1
−1	−1

If p> 2, then the above procedure shall be applied as long as m–1 has the value of 1. For 2^3 design, this leads to the following form:

$u_{p\text{-}2}$	$u_{p\text{-}1}$	u_p
1	1	1
1	1	–1
1	–1	1
1	–1	–1
–1	1	1
–1	1	–1
–1	–1	1
–1	–1	–1

To complete the process of creating a design (for p = 3), we should add to the left of the design a column of coefficients u_0 corresponding to the absolute term (intercept) in the linear equation:

u_0	u_1	u_2	u_3
1	1	1	1
1	1	1	–1
1	1	–1	1
1	1	–1	–1
1	–1	1	1
1	–1	1	–1
1	–1	–1	1
1	–1	–1	–1

The generated full factorial design (2^3) allows for identification of the coefficients a_i in the following linear model:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 \tag{10.11}$$

To create a full factorial design, allowing determination of coefficients in the following square model:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2 + a_3 \cdot x_1^2 + a_4 \cdot x_2^2 \tag{10.12}$$

it is necessary to use a 3–level design (3 levels for each factor: –1, 0, +1 for two variables). The way of creating a design, analogous to the above-discussed case, results in the matrix presented below.

Factorial designs 3^m for p≥4 require a large number of experiments. For instance, for p = 4 the number of points (experiments) in the full factorial design is equal to 81. A possible solution to this problem is to reduce a number of measurement points by using fractional factorial designs (e.g. $3^{m\text{-}k}$),, widely described in the book [14].

u_0	u_1	u_2	u_1^2	u_2^2
1	1	1	1	1
1	1	0	1	0
1	1	–1	1	1
1	0	1	0	1
1	0	0	0	0
1	0	–1	0	1
1	–1	1	1	1
1	–1	0	1	0
1	–1	–1	1	1

EXAMPLE

Determine the maximum efficiency of a reaction (*W* [%]) dependent on temperature (*T* [°C]) and substrate concentration (*c* [mol/m^3]). Minimum and maximum values of *T* and *c* are equal to:10 and 40 °C as well as 4 and 14 mol/m^3. From preliminary studies we know that the desired relationship is not linear.

SOLUTION

The general model can be put as follows:

$$W = a_0+a_1\cdot c+a_2\cdot T+a_3\cdot c^2+a_4\cdot T^2$$

In the first step, as in the procedure described in this chapter, we need to create an accurate factorial design, expressed in design variables 3^2 (2 variables at 3 levels) and real variables:

u_0	u_1	u_2
1	1	1
1	1	0
1	1	–1
1	0	1
1	0	0
1	0	–1
1	–1	1
1	–1	0
1	–1	–1

a_0	c	T
1	14	40
1	14	25
1	14	10
1	9	40
1	9	25
1	9	10
1	4	40
1	4	25
1	4	10

In order to convert the coded variables to real variables, Equation (10.8) can be applied.

The next step is to determine reaction efficiency for 9 experiments, in accordance with the generated full factorial design:

Experiment	*c* mol/m^3	*T* °C	*W* %
1	14	40	42
2	14	25	94
3	14	10	84
4	9	40	45
5	9	25	96
6	9	10	86
7	4	40	40
8	4	25	91
9	4	10	81

To perform a regression analysis using an Excel spreadsheet, data must be adequately prepared. For this purpose, it is best to present experimental results in the form:

a_0	c	T	c^2	T^2	W
1	14	40	196	1600	42
1	14	25	196	625	94
1	14	10	196	100	84
1	9	40	81	1600	45
1	9	25	81	625	96
1	9	10	81	100	86
1	4	40	16	1600	40
1	4	25	16	625	91
1	4	10	16	100	81

The first column (a_0) is used to determine an absolute term in a model (intercept), however, it is not necessary to select this column, if we use *Data analysis* (*Regression*). This column should be selected in calculations that use a matrix equation (Equation (4.6)). After calling *Data analysis→Regression* and selecting four columns (c, T, c^2, T^2) as the *input range X* and choosing appropriate analysis options, coefficients of a regression model are obtained. After taking into account standard deviations, the regression equation can be put as follows:

$$W = 31.09(\pm 0.86) + 2.91(\pm 0.18)\cdot c + 5.437(\pm 0.054)\cdot T - 0.1467(\pm 0.0095)\cdot c^2 - 0.1363(\pm 0.0011)\cdot T^2$$
$$r^2 = 0.9999,\ s_y = 0.34,\ n = 9,\ F = 10082$$

Calculating partial derivatives with respect to c and T and comparing them to 0, we get:

$$\partial \mathrm{W}/\partial \mathrm{c} = 2.91 - 0.293\cdot c = 0 \text{ and } c = 9.93 \text{ mol/m}^3$$
$$\partial \mathrm{W}/\partial \mathrm{T} = 5.437 - 0.273\cdot \mathrm{T} = 0 \text{ and } T = 19.93\ ^{\mathrm{o}}\mathrm{C}$$

Better determination of the maximum efficiency requires generating the next full factorial design close to the estimated maximum, and repeating some or all of the experiments for a new set of variables.

10.4. The simplex method

In contrast to the random walk method, random variations in parameter values in the simplex method have been replaced by an appropriate sequential algorithm. The optimization process begins with the creation of p+1 dimensional polyhedron (simplex) with p+1 vertices which correspond to a particular set of parameter values. An example can be a one–dimensional simplex which corresponds to a line segment, two–dimensional triangle and three–dimensional one, i.e. tetrahedron (Fig. 10.2).

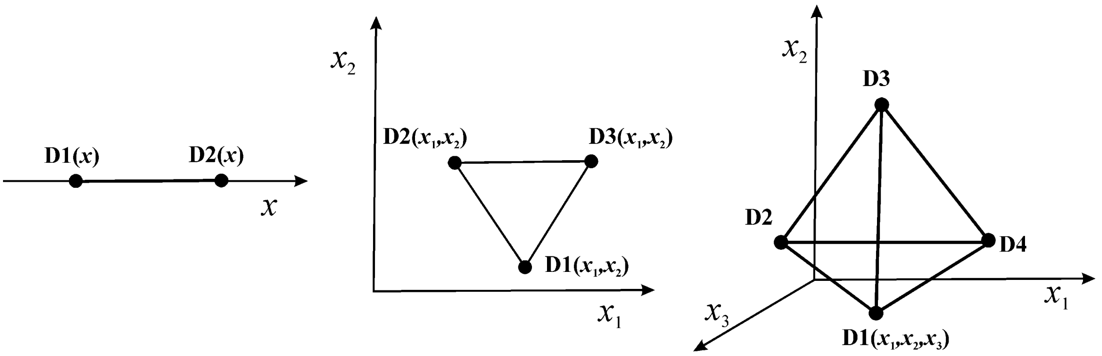


Fig. 10.2. Examples of simplexes in the of one–, two–and three–dimensional space.

After determining the value of a system response (after proper experiments) at various points a point is rejected for which the value of the object response is the smallest (if we search for the maximum).

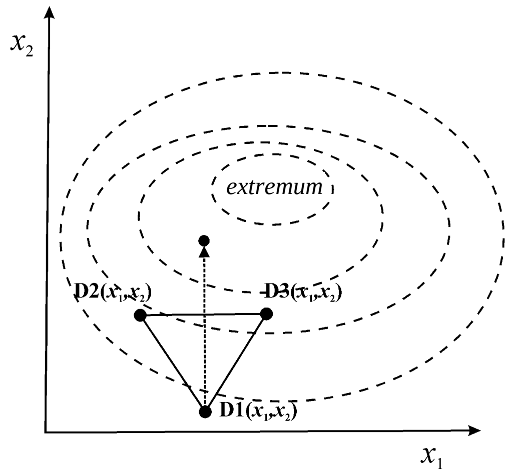


Fig. 10.3. The principle of the simplex method.

After rejecting this point, we determine a new set of parameters through the mirror image of the point rejected with respect to the segment joining the remaining points and we re-set the system response. The principle of the simplex method for two parameters is shown in Figure 10.3, where the dotted lines indicate schematically the response surface. Such a procedure is repeated several times until we reach the area of optimum (extremum) and the simplex moves in a zigzag line (Fig. 10.4).

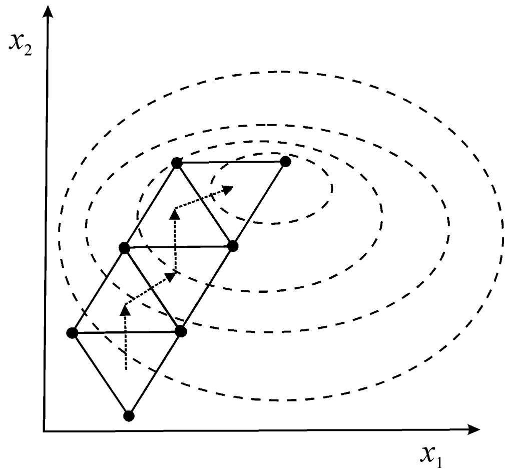


Fig. 10.4. Movements of a simplex in two dimensional space.

General rules of conduct in the simplex method are thus as follows:

Rule 1

The simplex moves after each p + 2 observation.

Rule 2

By rejecting the vertex (a set of parameters), characterized by the worst result (worst object response), followed by its mirror image with respect to the opposite edge of the simplex, we get a new simplex. Figure 10.5 shows the generated, new simplex adjacent to the existing one; it has been created by removing the point (D1) and adding the vertex (D4).

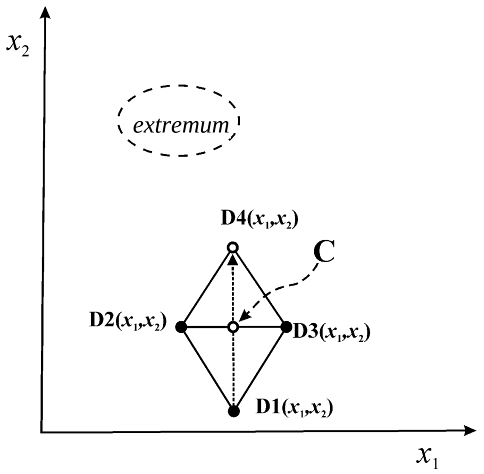


Fig. 10.5. Determining parameters for a new experiment in the simplex method.

In order to determine parameter values at the next simplex point (D4), we can use the general formula:

$$\mathrm{D4} = \mathrm{C} + (\mathrm{C} - \mathrm{D1}) = 2\mathrm{C} - \mathrm{D1} \tag{10.13}$$

which for individual parameters (variables x_1 and x_2) can be put as follows:

$$x_1(\mathrm{D4}) = x_1(\mathrm{C}) + (x_1(\mathrm{C}) - x_1(\mathrm{D1})) = 2x_1(\mathrm{C}) - x_1(\mathrm{D1}) \tag{10.14}$$

$$x_2(\mathrm{D4}) = x_2(\mathrm{C}) + (x_2(\mathrm{C}) - x_2(\mathrm{D1})) = 2x_2(\mathrm{C}) - x_2(\mathrm{D1}) \tag{10.15}$$

In these equations C is the point lying in the middle of the segment joining the remaining points, the coordinates of which can be calculated from the formulas:

$$\mathrm{C} = \frac{\mathrm{D2} + \mathrm{D3}}{2} \tag{10.16}$$

$$x_1(\mathrm{C}) = \frac{x_1(\mathrm{D2}) + x_1(\mathrm{D3})}{2} \tag{10.17}$$

$$x_2(\mathrm{C}) = \frac{x_2(\mathrm{D2}) + x_2(\mathrm{D3})}{2} \tag{10.18}$$

The new vertex of the simplex is therefore at the same distance from point C as D1 from C.

The object response for the new parameter values (point D4) should be better than at points D2 and D3. In this case, the point rejection procedure is repeated for the new simplex D2–D3–D4.

If the object response (result) is no better, it is recommended to reduce the distance between the vertices (reduce the size of the simplex) so as to determine more accurate optimal parameter values. If the modifications do not give better results, this means that the simplex circles around the optimal value (or it covers it). The procedure stops then and the mean values of the last simplex are taken as optimum.

If the simplex method is applied to the optimization of many parameters, it is often necessary to repeat the whole process with a new initial simplex due to the fact that the response surface can show local extrema.

If rule 2 fails, then rule 3 is applied.

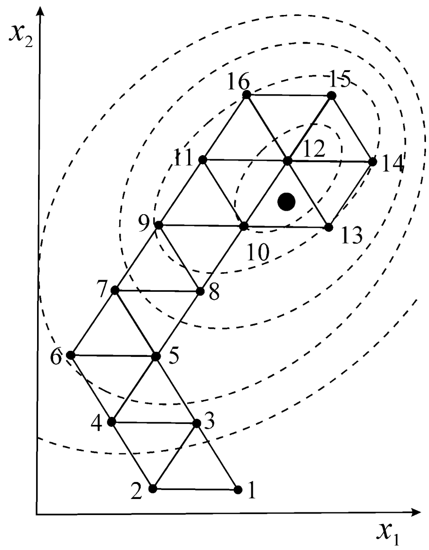


Fig. 10.6. An example of rule 3 with respect to simplex 10–12–13.

Rule 3

If the results obtained for the new parameter values are as bad or worse, the experiment should be repeated again to check if there is a random error. If the result is the same, it means that rule 2 cannot be applied. In this case, the next and worst point is discarded. An example of this rule is simplex 10–12–13 (Fig. 10.6) were a better result is obtained by discarding point 10 rather than 13

Rule 4

If after several simplex modifications one of the vertices remains among the new simplexes (e.g. point 12, Fig. 10.6), this means that the result of the experiment may be overestimated due to an experimental error. If the repeated result is true, then the procedure stops and the parameter values for that point are taken as optimal. Otherwise, the whole optimization procedure should start again, taking a new, improved initial simplex.

Rule 5

If the new vertex parameters go beyond acceptable limits in their physical meaning, or cannot be generated in the course of the conducted experiment, then:

a) a parameter value is established on a physically justifiable level and the procedure continues with a reduced simplex, or
b) we return to the previous simplex and repeat the procedure so as not to exceed the imposed limit for the optimized parameters.

The simplex method has also several limitations which include:

a) a possibility of determining the optimum area only in so far as permitted by the desired size of the simplex (simplex design scale)
b) problems with the way of verifying the optimum area achieved,
c) problems with choosing the size of the initial simplex, and
d) achieving a local maximum in the place of a global maximum.

In the case of the simplex method we should also remember not to draw conclusions regarding the conduct of a given process based on a simplex movement. In order to determine the relationships between a result and parameters, we can apply regression models, generated from the corresponding full factorial designs.

10.4.1. Variable-size simplex

A modified simplex method developed by Nelder and Mead [J.A. Nelder, R.Mead, A simplex method for function minimization, *Comput. J.*, 7 (1965) 308–313] introduces a simplex expansion in the direction of a growth of the response function or contraction in the direction opposite to a clear decrease of this quantity.

10.4.1.1. Expansion

Simplex expansion (Fig. 10.7) can be carried out if the value of system response (the result of experiment) for point DR is better than for the discarded (D1) and remaining points. The distance between points C and DE is generally 2 times greater than that between points C and DR. In this case, to determine the parameter values at the expanded point, we can apply the general formula:

$$\mathrm{DE} = \frac{\mathrm{D2}+\mathrm{D3}}{2} + 2\cdot\left[\frac{\mathrm{D2}+\mathrm{D3}}{2} - \mathrm{D1}\right] = 1.5\cdot(\mathrm{D2}+\mathrm{D3}) - 2\mathrm{D1} \qquad (10.19)$$

This equation for individual parameters can be put as:

$$x_1(\mathrm{DE}) = 1.5\cdot[x_1(\mathrm{D2}) + x_1(\mathrm{D3})] - 2\cdot x_1(\mathrm{D1}) \qquad (10.20)$$

$$x_2(\mathrm{DE}) = 1.5\cdot[x_2(\mathrm{D2}) + x_2(\mathrm{D3})] - 2\cdot x_2(\mathrm{D1}) \qquad (10.21)$$

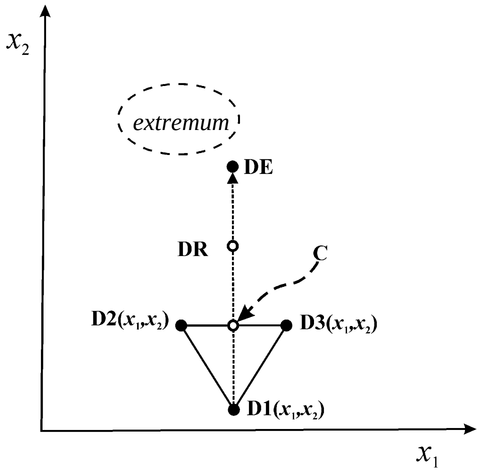


Fig. 10.7. Expansion of the D1–D2–D3 simplex.

If the result of the experiment that used the parameter values of point DE is better than that at the rejected point, the procedure is continued using a new, expanded simplex (D2–D3–DE). Otherwise, we go back to the unexpanded simplex and continue the procedure according to the simple simplex metod.

10.4.1.2. Contraction

If the value of the system response at point DR is worse than the results obtained for the remaining points, and at the same time it is better than the value obtained at the remaining point, simplex can be contracted. Taking into account the value of the result obtained at point DR, contractions can be carried out in two ways:

Positive contraction

If the value of the response function (f) of an object decreases in a range f(D3) > f(DR) > f(D1) (Fig.10.8), then the parameters of a new point (DK$^+$) can be calculated from the general formula:

$$\mathrm{DK}^{+}=\frac{\mathrm{D2}+\mathrm{D3}}{2}+0.5\cdot\left[\frac{\mathrm{D2}+\mathrm{D3}}{2}-\mathrm{D1}\right]=0.75\cdot(\mathrm{D2}+\mathrm{D3})-0.5\cdot\mathrm{D1} \qquad (10.22)$$

which for each variable can be given as follows:

$$x_1(\mathrm{DK}^{+})=0.75\cdot[x_1(\mathrm{D2})+x_1(\mathrm{D3})]-0.5\cdot x_1(\mathrm{D1}) \qquad (10.23)$$

$$x_2(\mathrm{DK}^{+})=0.75\cdot[x_2(\mathrm{D2})+x_2(\mathrm{D3})]-0.5\cdot x_2(\mathrm{D1}) \qquad (10.24)$$

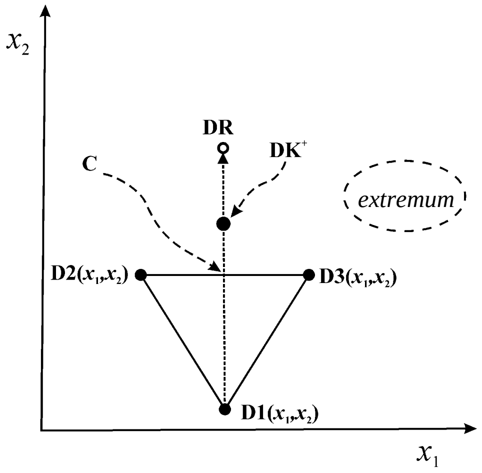


Fig. 10.8. Positive contraction of the D2–D3–DR simplex.

Negative contraction

When: f(DR)<f(D1), then a negative contraction (Fig. 10.9) is applied according to the formulas:

$$\mathrm{DK}^{-}=\frac{\mathrm{D2}+\mathrm{D3}}{2}-0.5\cdot\left[\frac{\mathrm{D2}+\mathrm{D3}}{2}-\mathrm{D1}\right]=0.25\cdot(\mathrm{D2}+\mathrm{D3})+0.5\cdot\mathrm{D1} \qquad (10.25)$$

$$x_1(\mathrm{DK}^{-})=0.25\cdot[x_1(\mathrm{D2})+x_1(\mathrm{D3})]+0.5\cdot x_1(\mathrm{D1}) \qquad (10.26)$$

$$x_2(\mathrm{DK}^{-})=0.25\cdot[x_2(\mathrm{D2})+x_2(\mathrm{D3})]+0.5\cdot x_2(\mathrm{D1}) \qquad (10.27)$$

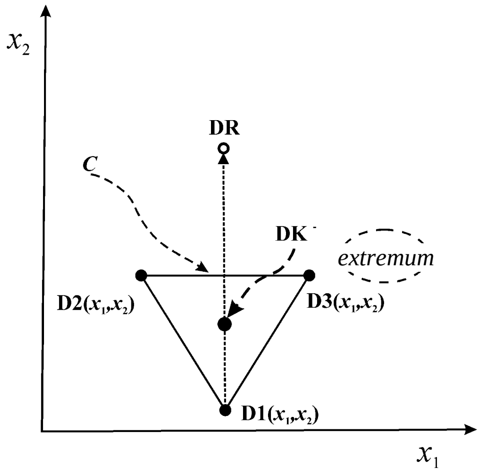


Fig. 10.9. Negative contraction of the D2–D3–DR simplex.

When a modified simplex method is used, a simplex ceases to be a regular figure and in the course of its movement it changes its size (Fig. 10.10).

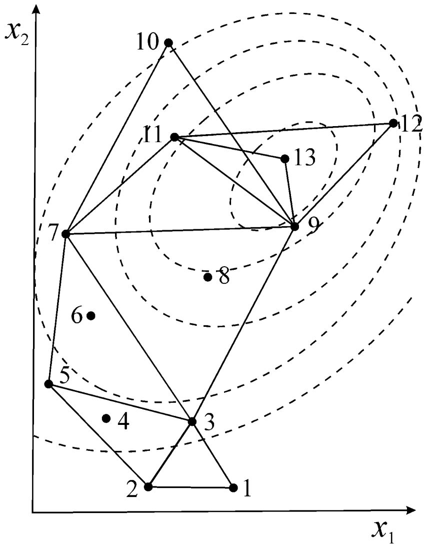


Fig. 10.10. Movements of a variable-size simplex.

10.4.2. Optimization criteria

In order to complete the optimization in a proper moment, it is necessary to predetermine an adequately defined criterion. At the same time it is not possible to formulate a general criterion, adequate for all the analysed cases due to various, often conflicting criteria, such as yield and purity of a product. Therefore, in accordance with the general rules of conduct, a properly formulated criterion should include:

1) The analysis of the obtained results to estimate whether the changes observed are below the threshold value, which indicates the achievement of a response surface area with the extremum of the desired size.
2) The analysis of parameter values by testing for their changes during optimization, allowing completion of the procedure on the basis of the size of the simplex.
3) The analysis of gradients to evaluate the effectiveness of the procedure which unfortunately fails when the response surface includes saddle points.
4) The analysis of the results to generate a mathematical model describing the response surface to obtain an image or cross-sections of the surface.

In experimental studies it is necessary to use many of these criteria simultaneously and to evaluate a result during optimization, which is rarely carried out automatically.

Literature presents many examples of practical use of the simplex method, among others, to determine the optimum operating conditions of an atomic absorption spectrophotometer, to optimize the magnetic field homogeneity in NMR, to optimize separation of an isomeric mixture in chromatography and to optimize various reaction yields, etc.

11. Monte Carlo methods - integration and simulation

11.1. Pseudorandom number generators

Random numbers are used in many iterative computations. The Monte Carlo method is one of the methods requiring generation of random numbers with uniform distribution.

The Neuman–Metropolis method is among the basic methods of generating random numbers. In order to obtain pseudorandom numbers with uniform distribution in an interval ⟨0,1⟩, this method uses any *m*–digit number from this interval which can be put in *m* binary positions:

$$x_1 = \alpha_1 2^{-1} + \alpha_2 2^{-2} + \cdots \alpha_m 2^{-m} \tag{11.1}$$

In this formula, α_i is a figure at the i–th position in the m–digit number. By squaring the number, we get:

$$x_1^2 = \beta_1 2^{-1} + \beta_2 2^{-2} + \cdots \beta_m 2^{-m} \tag{11.2}$$

The middle of this number, i.e.:

$$x_2 = \beta_{\frac{m}{2}+1} 2^{-1} + \beta_{\frac{m}{2}+2} 2^{-2} + \cdots \beta_{\frac{3m}{2}} 2^{-m} \tag{11.3}$$

is another pseudorandom number.
Continuing the calculations, i.e. squaring and setting the next middle number, we obtain a sequence of numbers x_1, x_2 ... x_n uniformly distributed in the interval ⟨0,1⟩.

EXAMPLE

Calculate two pseudorandom numbers (x_2 and x_3) from the interval from 0 to 1, using the Neuman–Metropolis method. The initial value to be assumed in the calculations is: $x_1 = 0.1107$

SOLUTION

By taking it as the initial value $x_1 = 0.1107$ and squaring, we get:

$$x_1^2 = (0.1107)^2 = 0.01225449$$

The middle four digits of the number are another pseudorandom number:

$$x_2 = 0.2254$$

Proceeding analogously with the value of $x_{2,}$ we obtain:

$$x_2^2 = (0.2254)^2 = 0.05080516$$

and:

$$x_3 = 0.0805$$

There are many computer programs, called pseudorandom number generators for creating subsequent numbers by operations on a preceding number. As the computer is 100% deterministic, a sequence of pseudorandom numbers initiated with the same value is always the same.

The most commonly used pseudorandom number generation algorithm is the recursive method developed by Lehmer, called a Linear Congruential Generator (LCG):

$$x_i = \left(a \cdot x_{i-1} + c\right) \operatorname{mod} m \tag{11.4}$$

where a is the multiplier, and c – gain, mod – integer remainder.
Because, according to the equation (11.4), another pseudorandom number is obtained from a previous one, it is necessary to specify the initial value x_0 from which an algorithm starts. Depending on the value c, we can distinguish two basic types of generators LCG. Where $c \neq 0$, the generator is called additive (mixed), when $c=0$ – multiplicative.

The formula (11.4) has been applied in many pseudorandom number generators, however, different values of coefficients are used depending on a program (or programming language). Some typical examples are shown in Table 11.1.

Tab. 11.1. Some examples of pseudorandom number generators, and coefficients

Name	*m*	*a*	*c*
RANDU	2^{31}	65539	0
DERIVE	2^{32}	3141592653	1
RNB	2^{31}	$2^2 \cdot 23^7+1$	0
RAND	2^{32} or 2^{31}	1103515245	12356
MINSTD	$2^{31}-1$	16807	0

Correct operation of a pseudorandom number generator (uniform distribution and no correlation between the values) can be determined by calculating appropriate statistical tests. A chi–square test (χ^2) is among the most popular distribution compliance tests

In order to check whether 10 000 selected numbers (cardinality in each interval: 971, 1015, 974, 1014, 1012, 1015, 981, 1005 1006 and 1007) is characterized by a uniform distribution, we can calculate statistics χ^2 from the formula:

$$\chi^2_{l-1} = \sum_i^l \frac{(n_i - n)^2}{n} \tag{11.5}$$

where n_i is cardinality in the *i*–th interval, *n* is the expected value n_i equal to 1000, and *l* is the number of subintervals. The calculated value of $\chi^2 = 2.78$ is compared with critical values, for example, from a distribution table χ^2 which for the number of degrees of freedom $l - 1 = 9$ are: 2.088, 2.532 and 3.325 for respectively 99%, 98% and 95% confidence levels. Based on these values, it can be stated that with 95–98% probability, the sequence of 10 000 pseudorandom numbers is characterized by a uniform distribution.

The uniform distribution obtained in most pseudorandom number generators can be transformed into the Gaussian distribution, in accordance with the statement:
"*If a population of variables has finite variance σ^2, and mean x, then, with increasing n, the sample mean distribution aims at the normal distribution with mean and variance σ^2/n.*"
called the central limit Theorem.
An adequate formula can be applied for the transformation:

$$x_{k,G} = 2\sum_{i=1}^{N} x_k - N \tag{11.6}$$

where x_k is numbers with uniform distribution, while $x_{k,G}$ – numbers with normal distribution. A correct normal distribution can be obtained for values *N* ranging from 10 to 12.

The Box–Müller transform is a much more common method for generating normally distributed numbers. In this method, independent random variables x_1 and x_2 with uniform distributions in an interval $\langle 0,1 \rangle$ are transformed into random variables y_1 and y_2 with the Gaussian distribution, according to the following formulas:

$$y_1 = \sqrt{-2 \cdot \ln(x_1)} \cdot \cos(2 \cdot \pi \cdot x_2) \tag{11.7}$$

$$y_1 = \sqrt{-2 \cdot \ln(x_1)} \cdot \sin(2 \cdot \pi \cdot x_2) \tag{11.8}$$

Such transformation results in normally distributed variables with mean equal to zero and standard deviation equal to one.

11.2. Monte Carlo integration

Numerical integration methods are used to estimate integrals of the general form:

$$I = \int_a^b f(x)dx \tag{11.9}$$

For this purpose, we can use the rectangle, trapezoid, Simpson, and Gaussian methods presented in Chapter 5, or apply pseudorandom numbers and calculate the integral defined with the Monte Carlo method.

The simplest variant of the MC method is a variant called "success–failure" , "Hit-or-Miss" or "Acceptance — Rejection " where an integrand ($f(x)$) in an interval $\langle a, b\rangle$ is limited to a rectangle $P = \langle a, b\rangle \times \langle 0, M\rangle$ where M is the maximum value of the integrand ($M = \max f(x)$ for $x \in \langle a,b\rangle$). Then we randomize N points of the rectangle P, each of which can be placed above or below the graph of the $f(x)$ function. The approximate value of the integral can be calculated from the equation:

$$\int_a^b f(x)dx \approx \frac{k}{N}(b-a)\cdot M \tag{11.10}$$

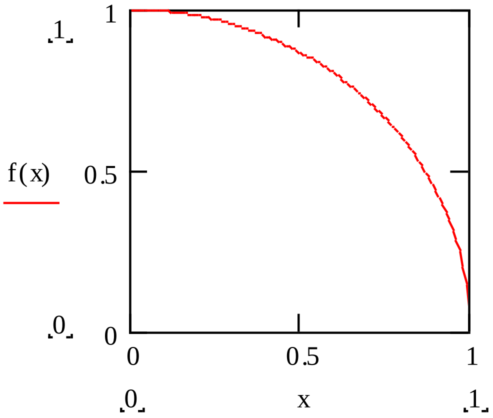
where k is the number of points located below the graph of the $f(x)$ function.

EXAMPLE

Calculate the approximate value of number π using the Monte Carlo method.

SOLUTION

The surface area of the circle described by the function $y^2 + x^2 = 1$ is equal to π. The graph of the function $y = \sqrt{1-x^2}$ is shown in the figure:



In order to calculate the value π using the Monte Carlo method, we need to:

a) randomize N points of a square with coordinates [0,0], [0,1], [1,1] i [1,0]

b) check after each randomization whether the coordinates satisfy inequality:

$$y^2 + x^2 \leq 1 \qquad \text{(belong to ¼ of the circle)}$$

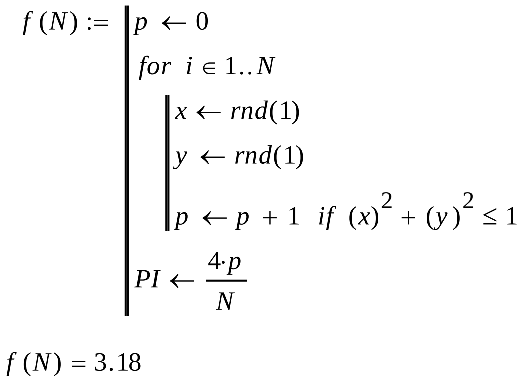
c) calculate the area of the circle which is:

$$PI = 4{\cdot}p/N$$

p – number of samples satisfying the inequality, N – number of randomized points.

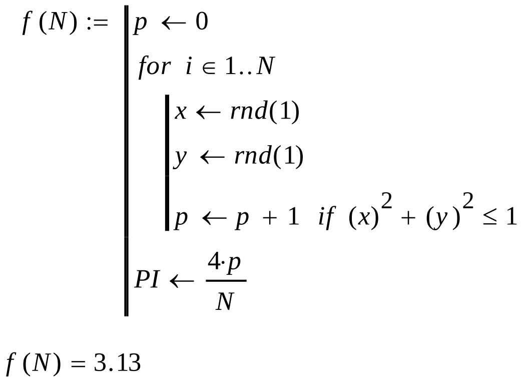
Below is a sample algorithm for solving this problem in Mathcad for $N = 800$ and $N = 8000$:

N:=800



```
f(N) := | p ← 0
        | for i ∈ 1..N
        |     | x ← rnd(1)
        |     | y ← rnd(1)
        |     | p ← p + 1  if (x)^2 + (y)^2 ≤ 1
        | PI ← 4·p / N

f(N) = 3.18
```

N:=8000



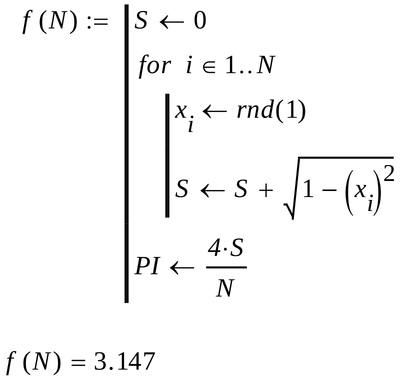
```
f(N) := | p ← 0
        | for i ∈ 1..N
        |     | x ← rnd(1)
        |     | y ← rnd(1)
        |     | p ← p + 1  if (x)^2 + (y)^2 ≤ 1
        | PI ← 4·p / N

f(N) = 3.13
```

Computations show that with increasing N, the error between the calculated and theoretical values (3.14159…) is reduced.

The approximate value of an integral can be obtained by applying the basic Monte Carlo (*Crude Monte Carlo*) method which counts only the points lying below the graph of the $f(x)$ function. For the example presented above, the surface of the circle's part lying in the first quadrant can be calculated from the formula

$$I = \int_0^1 \sqrt{1 - x^2}\,dx$$

The corresponding solution algorithm in Mathcad for N= 8000 is as follows:



```
f(N) := | S ← 0
        | for i ∈ 1..N
        |     | x_i ← rnd(1)
        |     | S ← S + √(1 − (x_i)^2)
        | PI ← 4·S / N

f(N) = 3.147
```

The basic Monte Carlo method has a much higher accuracy compared to the "success–failure" metod.

11.3. Monte Carlo simulation

Computers with ever-increasing computational power and software designed for chemists, enable the study of various processes without their physical conduct by using an appropriate mathematical model. At the same time more and more realistic models lead to results confirmed by results of experiments. One example of simulation methods used for a modelling of chemical and physico-chemical processes is the MC method applied, among others, for simulation:

1) molecular dynamics (e.g. modeling of liquid water),

2) reaction dynamics (collision theory),
3) chromatographic processes,
4) quadrupole mass analyzer,
5) adsorption processes.

The Monte Carlo method is considered to be a stochastic method which does not model the movement of molecules but only transitions from one configuration to another. Applications of the Monte Carlo method are widely described in literature. One interesting example is modelling activated carbon adsorption [P. A. Gauden, A. P. Terzyk, S. Furmaniak, Modele budowy węgla aktywnego wczoraj, dzisiaj, jutro*, Wiadomości Chemiczne* 62 (5-6) (2008) 403-447]. For this purpose, an appropriate simulation cells are used where a number of adsorbate configurations are generated as components of a specific statistical unit. In studies on adsorption in heterogeneous systems with the Monte Carlo method, they use a grand canonical ensemble which is an open system (can exchange mass and energy with the surrounding), while chemical potential, volume and temperature are constant.

Simulation methods enable the evaluation of a model validity by comparing experimental results with simulation results, as well as verifying a theory by comparing theoretical results with simulations on the same system. They are also much cheaper than experimental studies, and allow for conducting simulations in conditions inaccessible to experiments (high pressure, temperature). They also provide information not only about the macroscopic properties of a system, but also a structure at the molecular level

II. LABORATORY

EXCERSISE No. 1
STATISTICAL ANALYSIS OF EXPERIMENTAL DATA
1. The mean, standard deviation, dispersion measures.

EXCERSISE No. 2
STATISTICAL ANALYSIS OF EXPERIMENTAL DATA
2. Dependence of the mean and measures of statistical dispersion on the number of samles.

EXCERSISE No. 3
REGRESSION ANALYSIS
Application of the linear regression to calculate the first-order reaction rate constant

EXCERSISE No. 4
CALCULATION OF THE pH OF THE TWO ACIDS MIXTURE

EXCERSISE No. 5
MULTIPLE LINEAR REGRESSION

EXCERSISE No. 6
LINEAR REGRRESION –LINEARIZING TRANSFORMATION

EXCERSISE No. 7
NUMERICAL INTEGRATION
THE RECTANGULAR, TRAPEZOIDAL AND SIMPSON’S RULE METHOD

EXCERSISE No. 8
NUMERICAL SOLVING OF DIFFERENTIAL EQUATIONS
EULER, RUNGE – KUTTA, MILNE METHODS

EXCERSISE No. 9
SIMPLEX OPTIMIZATION

INTRODUCTION

Laboratory exercises are closely related to the topics presented in the lectures, but they require a basic knowledge about the use of a spreadsheet (Microsoft Office Excel). A brief introduction to a spreadsheet, the basic features of the spreadsheet and sample functions are described below.

The spreadsheet is a computer program used to calculate arrays. In the spreadsheet, it is possible to present figures and other data in arrays consisting of rows and columns. Columns are denoted with letters, rows – with numbers. At the intersection of each column and row is a cell, uniquely defined by its address. The address of a cell consists of a letter (or letters) defining a column, and a number indicating a row in which it is located (e.g. B2).

We can enter three kinds of data to each cell: label, number, or formula (equation). Labels are adequate names such as Data, Sum, Product, etc., used to identify (describe) calculations made in the spreadsheet. The number is a combination of figures from 0–9, while the formula is a specific relationship between cells. Formulas used for arithmetic calculations, for example the formula =B2*B3 multiplies the content of the cell with the address B2 by the value of the cell with the address B3. A sign = is a mandatory operator when performing arithmetic calculations. The program also features standard formulas, available by selecting the formula creator (icon f_x – insert function) or, if we know the name of the function – by typing it in a spreadsheet cell. Below are examples of standard spreadsheet functions which can be used in solving specific laboratory tasks. Due to differences between the names of individual features in the latest version of MS Excel 2010 and earlier versions, names are provided according to the version.

Selected spreadsheet functions related to the descriptive statistics:

- =sum(number1,number2,…)
- =sqrt(number)
- =average(number1,number2,…)
- =median(zakres_komórek)
- =var(number1,number2,…)
- =stdev(number1,number2,…)
- =tinv(probability;degrees_freedom)
- =chiinv(probability;degrees_freedom)
- =frequency(data_array,bins_array)

FREQUENCY is an example of an array function which is introduced in a strictly defined manner. After selecting a function and selecting data (array_data) and intervals (array_intervals), we select a range of cells where relevant results shall appear (the same size as array_intervals). Then we press the F2 key on the function keyboard and end the computations by pressing Ctrl+Shift+Enter.

TINV and CHIINV functions calculate the value t (from the student's t-distribution) and χ^2 (from the chi-square distribution), necessary to define a confidence interval for a mean value and standard deviation (or variance), respectively

Statistical calculations can be carried out using the Excel add-in – Data analysis. As we enable this option (toolbar Quick Access→Excel Options→Extras→Go and choose the option *Analysis ToolPak*) the *Data analysis* button is available in the *Data* tab.

When we choose the *Descriptive statistics* tool from the available list, and select appropriate data (*Input range*) and options (*Summary statistics* and *Confidence level for the mean*), we get an analysis summary in the form of an appropriate array.

Selected spreadsheet functions related to the regerssion analysis:

- =slope(known_ys; known_xs)
- =intercept(known_ys; known_xs)
- =r.kwadrat(known_ys; known_xs)
- =steyx(known_ys; known_xs)
- =rsq(array1;array2)
- =minverse(array)

The last two functions, just as the FREQUENCY function, are table functions and need to be entered in the manner described above.

A complete regression analysis can be obtained by choosing the *Regression* tool from a list of available tools (*Data analy*sis). After selecting the input data (*Input Y range, Input X range*) and the options (*Confidence level* and *Residuals*) in the same, or a new spreadsheet (*Output options*) a summary of the calculations is generated.

The spreadsheet also provides a graphical representation of figures in the form of charts. In order to generate a chart, we select a block of data, we can use a chart creator (in older versions) or the corresponding menu (*Insert→Charts...*).

Solver is another Excel add-in used during classes. Solver can be used for calculations where it is necessary to change values in certain cells (*changing cells*) in order to obtain a result which is defined by a user as an adequate formula in a target cell (*target cell*).

After calling Solver, a window is displayed in which we must enter: *target cell* (that contains a formula) which can take a specified, maximum or minimum value. The *target cell* is directly or indirectly related to the *changing cells*. The program will change numeric values in these cells as long as the formula in the cell shown in the *Target cell* takes a certain value. Additionally, we can enter appropriate restrictions (*Constraints*) affecting the numeric values changed. The *Options* button loads or saves models, or changes standard calculation parameters. The *Solve* button starts computations.

Literature:

M. Pilch, *Ćwiczenia z Excel dla chemików*, Mikom, 2001
K. Mądry, W. Ufnalski, *Excel dla chemików i nie tylko*, W. N.-T., 2000
E. Joseph Billo, *Excel for Chemists: A Comprehensive Guide.* John Wiley & Sons, Inc., 2001
R. de Levie, *How to use Excel in analytical chemistry and in general scientific data analysis*, Cambridge University Press, 2004
Z. Smogur, Excel w zastosowaniach inżynieryjnych, Wydawnictwo Helion, 2008
V. Gupta, Statistical analysis with Excel, VJ Books Inc, 2002

EXERCISE No. 1

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

1. The mean, standard deviation, dispersion measures.

The water content in the samples (10 g) of fertilizer was investigated according to the rules described in Polish Norm PN/C-04500. The results are as follows:

No.	1	2	3	4	5	6	7	8	9	10
mass	0.182	0.088	0.095	0.170	0.176	0.075	0.159	0.150	0.155	0.141
No.	11	12	13	14	15	16	178	18	19	20
mass	0.101	0.121	0.111	0.140	0.118	0.132	0.108	0.127	0.115	0.138
No.	21	22	23	24						
mass	0.125	0.129	0.126	0.131						

Calculate the statistical parameters using the formulas given below, and compare the results with values calculated using standard spreadsheet functions. Calculate the following quantities using the standard spreadsheet functions. A detailed instruction on the calculation and presentation of the results is presented below (see COMMENTS).

I. STATISTICAL CALCULATIONS

Calculate:

(a) The arithmetic mean of the water content in the samples:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

(b) The median:

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{for even values of } n \\ \dfrac{x_{n/2} + x_{(n/2)+1}}{2} & \text{for odd values of } n \end{cases} \tag{2}$$

(c) The variance:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \tag{3}$$

where *n*-1 denote the number of degree of freedom (d.f.), i.e. the number of independent observations, which are used in calculating of *s*.

(d) Standard deviation, s_x:

$$s_x = \sqrt{s_x^2} \tag{4}$$

(e) Relative standard deviation:

$$v_x = 100\frac{s_x}{\overline{x}} \tag{5}$$

(f) Standard uncertainty (standard error of the mean):

$$u(x) = s_{\overline{x}} = \frac{s_x}{\sqrt{n}} \tag{6}$$

(g) Expanded uncertainty:

$$U = k \cdot u(x) = k \cdot s_{\overline{x}} = k \cdot \frac{s_x}{\sqrt{n}} \tag{7}$$

k – coverage factor (*k* = 2 or 3)

(h) Confidence interval:

$$\text{c.i.} = t_{\alpha,n-1} \cdot s_{\overline{x}} = t_{\alpha,n-1} \cdot \frac{s_x}{\sqrt{n}} \tag{8}$$

t – value from Student's distribution (function **TINV()**).

II. ESTIMATION OF VARIANCE AND STANDARD DEVIATION

Based on the calculations determine the confidence interval for the variance s_x^2 containing the "true" value σ_x^2 with 95% probability. Assume that the sample comes from a normally distributed population, and the random variable:

$$\frac{rs_x^2}{\sigma_x^2} \tag{6}$$

has a normal distribution χ^2 of *r* degrees of freedom , i.e.:

$$P\left\{\frac{r\,s_x^2}{\chi_{r,\alpha/2}^2} \le \sigma_x^2 \le \frac{r\,s_x^2}{\chi_{r,1-\alpha/2}^2}\right\} = 1-\alpha \tag{7}$$

assume a confidence coefficient of 1-α =0.95, *and find the values of* χ^2 *in the relevant statistical tables* (*e.g. Metody statystyczne dla chemików, J.B. Czermiński, A. Iwasiewicz, Z. Paszek, A. Sikorski*).

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE1\Exe01.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames.

3. **Calculate the statistical parameters using the equations and compare the results with values calculated using standard spreadsheet functions and data analysis (descriptive statistics).**

4. Arrange the measurement results by increasing the water content in the samples and determine the number of results in the intervals (function FREQUENCY):

$$
\begin{gathered}
0.06 \le x_i \le 0.08\\
0.08 < x_i \le 0.10\\
0.10 < x_i \le 0.12\\
0.12 < x_i \le 0.14\\
0.14 < x_i \le 0.16\\
0.16 < x_i \le 0.18\\
0.18 < x_i \le 0.20
\end{gathered}
$$

5. Calculate the relative multiplicity of the water content in each interval, i.e.:

$$l = \frac{n_{x_i}}{\sum n_{x_i}}$$

where n_{x_i} - "number of results in a given interval

$\sum n_{x_i}$ - total number of results

and make a histogram of the water content denoting the corresponding interval as: I, II, III, IV, V, VI, VII.

6. Make a distribution curve of the water content in the samples by plotting *l* in the function $x_i^{przedzid}$, where $x_i^{przedzid}$ corresponds to the mean value of x_i in the above intervals, i.e. 0.07, 0.09, 0.11 etc. Does the plot corresponds to the normal distribution curve?

7. **Using a recursive method calculate the mean and the standard deviation.**

8. **The result quote according to the *i*) standard uncertainty, *ii*) expanded uncertainty (*k*=2), and *iii*) confidence interval for the mean.**

9. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the performer should be placed in the footer or header (optional).

APPENDIX I
RECURSIVE METHOD

The mean and the standard deviation can be calculated using the recursive method. In this method the first trial value of the mean (m_1) is equal to the first measured value (x_1), i.e.:

$$m_1 = x_1 \tag{11}$$

and the first trial value of the sum-squared deviation (q_1) is equal to zero:

$$q_1 = 0 \tag{12}$$

Subsequent values of the mean (m_i) and the sum-squared deviations (q_i) can be evaluated from the following equations:

$$m_i = \frac{(i-1)m_{i-1} + x_i}{i} \tag{13}$$

$$q_i = q_{i-1} + \frac{(i-1)(x_i - m_{i-1})^2}{i} \tag{14}$$

After completing the calculations for all values of *i* (*i* = 1, 2, ..*n*), the final m_i value is the mean of the entire data set (m_n) whereas the standard deviation (*s*) is computed using the equation:

$$s = \sqrt{\frac{q_n}{n-1}} \tag{15}$$

where q_n denote the final value of sum-squared deviations (q_i).

EXERCISE No. 2

STATISTICAL ANALYSIS OF EXPERIMENTAL DATA

2. Dependence of the mean and measures of statistical dispersion on the number of samles.

The water content in the samples (10 g) of fertilizer was investigated according to the rules described in Polish Norm PN/C-04500. The results are as follows:

No.	1	2	3	4	5	6	7	8	9	10
mass	0.182	0.088	0.095	0.170	0.176	0.075	0.159	0.150	0.155	0.141
No..	11	12	13	14	15	16	178	18	19	20
mass	0.101	0.121	0.111	0.140	0.118	0.132	0.108	0.127	0.115	0.138
No..	21	22	23	24						
mass	0.125	0.129	0.126	0.131						

It was found that a 24–hour production cycle always gives results analogous to those given in the table above. In order to reduce the cost of laboratory tests, it was decided to limit the number of analyses and samples were taken every 2 hours.

The purpose of this task is to examine the relationship between the average water content and other statistical quantities, and the frequency of sampling for the analysis.

Calculate the following quantities using the standard spreadsheet functions. A detailed instruction on the calculation and presentation of the results is presented below (see COMMENTS).

I. STATISTICAL CALCULATIONS

Calculate:

(a) The arithmetic mean of the water content in the samples:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (1)$$

(b) The median:

$$\tilde{x} = \begin{cases} x_{(n+1)/2} & \text{for even values of } n \\ \dfrac{x_{n/2} + x_{(n/2)+1}}{2} & \text{for odd values of } n \end{cases} \qquad (2)$$

(c) The variance:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad (3)$$

where *n*-1 denote the number of degree of freedom (d.f.), i.e. the number of independent observations, which are used in calculating of *s*.

(d) Standard deviation, s_x:

$$s_x = \sqrt{s_x^2} \qquad (4)$$

(e) Relative standard deviation:

$$v_x = 100\frac{s_x}{\overline{x}} \tag{5}$$

(f) Standard uncertainty (standard error of the mean):

$$u(x)=s_{\overline{x}} = \frac{s_x}{\sqrt{n}} \tag{6}$$

(g) Expanded uncertainty:

$$U=k \cdot u(x)=k \cdot s_{\overline{x}} = k \cdot \frac{s_x}{\sqrt{n}} \tag{7}$$

k – coverage factor (*k* = 2 or 3)

(h) Confidence interval:

$$\text{c.i.}= t_{\alpha,n-1} \cdot s_{\overline{x}} = t_{\alpha,n-1} \cdot \frac{s_x}{\sqrt{n}} \tag{8}$$

t – value from Student's distribution (function **TINV()**).

II. ESTIMATION OF VARIANCE AND STANDARD DEVIATION

Based on the calculations determine the confidence interval for the variance s_x^2 containing the "true" value σ_x^2 with 95% probability. Assume that the sample comes from a normally distributed population, and the random variable:

$$\frac{rs_x^2}{\sigma_x^2} \tag{6}$$

has a normal distribution χ^2 of *r* degrees of freedom , i.e.:

$$P\left\{\frac{r\,s_x^2}{\chi^2_{r,\alpha/2}} \le \sigma_x^2 \le \frac{r\,s_x^2}{\chi^2_{r,1-\alpha/2}}\right\} = 1-\alpha \tag{7}$$

assume a confidence coefficient of 1-α =0.95, *and find the values of* χ^2 *in the relevant statistical tables* (*e.g. Metody statystyczne dla chemików, J.B. Czermiński, A. Iwasiewicz, Z. Paszek, A. Sikorski*).

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE2\Exe02.xls, where AA denote the number of class, BB – user number.Prepare a table containing the results of the water content determinations in the samples corresponding to each of the 6 series.

2. **Make calculations and put the results in a separate array.**
3. Pay attention to the careful planning of tables, descriptions, and frames.

4. Present the calculation results graphically in the form of curves:

$$\bar{x} = f(n)$$
$$\tilde{x} = f(n)$$
$$s_x^2 = f(n)$$
$$v = f(n)$$

   Pay attention to the careful description of curves by placing relevant comments in the figure.

5. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

EXERCISE No. 3

REGRESSION ANALYSIS

Application of the linear regression to calculate the first-order reaction rate constant

Ester hydrolysis reactions were conducted in the presence of hydrochloric acid as a catalyst. Samples of the reaction mixture were taken during the reaction and the concentration of the resulting carboxylic acid $[C]_t$ was determined:

REACTION									
time (min)	15	30	47	67	80	95	115	127	142
$[C]_{A,t}$ (mol/dm^3)	0.035	0.059	0.072	0.096	0.123	0.139	0.168	0.171	0.174

In order to complete the task, use the spreadsheet and:

1. Calculate the hydrolysis reaction rate constant as a first-order reaction,
2. Verify the assumption of the first-order reaction based on linear correlation analysis,
3. Calculate statistical values to assess regression coefficients (S_{a_1} and S_{a_0}) and confidence interval for the constant $k \equiv a_1$ at a significance level α= 0.05.

I. CALCULATION OF THE REGRESSION COEFFICIENTS

The integral form of the kinetic equation of the first-order reaction is given by:

$$\ln \frac{[C]_{t\to\infty} - [C]_t}{[C]_{t\to\infty} - [C]_{t=0}} = -kt \qquad (1)$$

where k - reaction rate constant (s^{-1}), t – reaction time in s ($[C]_{t\to\infty}$ = 0.5)

Using the linear regression equation in the form:

$$Y = a_0 + a_1 X \qquad (2)$$

where $\qquad Y \equiv \ln f([C]_{R,t}) \qquad X \equiv t \qquad a_1 \equiv k \qquad (3)$

calculate regression coefficients (and thus k) using the method of least squares.

Make the calculations using the linear regression subroutine of the spreadsheet and independently using the formulas given below.

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} \qquad (4)$$

$$a_0 = \overline{y} - a_1 \overline{x} \qquad (5)$$

where $\overline{y}$ and $\overline{x}$ denotes the arithmetic mean of y_i and x_i:

$$\overline{y}_{i} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \overline{x}_{i} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{6}$$

II. EVALUATION OF THE LINEAR MODEL

To assess an error made while trying to describe the phenomenon of ester hydrolysis using the linear first-order reaction model, calculate:

(a) residual variance:

$$s_y^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n}(y_i - \overline{y})^2 - a_1^2\sum_{i=1}^{n}(x_i - \overline{x})^2\right] \tag{7}$$

n–2 corresponds to the number of degrees of freedom.

(b) residual standard deviation (mean deviation from the regression):

$$s_y = \sqrt{s_y^2} \tag{8}$$

(c) coefficient of linear correlation:

$$r = \frac{\sum_{i=1}(x_i - \overline{x})(y_i - \overline{y})}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\sum_{i=1}^{n}(y_i - \overline{y})^2\right]^{1/2}} \tag{9}$$

(d) coefficient of determination (squared correlation coefficient)

$$\text{det.coeff.} = r^2 \tag{10}$$

III. STANDARD DEVIATION AND CONFIDENCE INTERVALS FOR THE REGRESSION COEFFICIENTS.

The kinetic constant *k* should be reported according to the standard deviation and confidence interval at a significance level $\alpha = 0.05$

The standard deviation of the a_1 regression coefficient (equivalent to a reaction rate constant *k*) calculate from the equation:

$$s_{a_1} = s_y\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \overline{x})^2}} \tag{11}$$

In order to calculate the confidence interval, we assume that the error:

$$\varepsilon_i = y_i - B - Ax_i \tag{12}$$

is normally distributed (*A* and *B* are the regression coefficients in the general population). In this case, the variable t_{a_1} :

$$t_{a_1} = \frac{a_1 - A}{s_{a_1}} \qquad (13)$$

has a Student's distribution with n-2 degrees of freedom. This means that the confidence interval for k at the given confidence level β=1-α is as follows:

$$P(a_1 - t_{\alpha,n-2} \cdot s_{a_1} < A \equiv k < a_1 + t_{\alpha,n-2} \cdot s_{a_1}) = 1 - \alpha \qquad (14)$$

Accordingly, calculate the confidence interval for k (α=0.05) expressed by the formula:

$$\text{c.i.} = \pm\, t_{\alpha,n\text{-}2} \cdot s_{a_1} \qquad (15)$$

where $t_{\alpha,n\text{-}2}$ is the tabulated value of the Student's distribution (e.g. *Metody statystyczne dla chemików, J.B. Czermiński, A. Iwasiewicz, Z. Paszek, A. Sikorski*).

The k±c.i. means that the constant k is in the given interval with probability 100× (1-α), i.e. for α = 0.05 equal to 95%.

The standard deviation of the regression coefficient calculated using the equation:

$$s_{a_0} = s_y \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i - \bar{x})^2}} \qquad (16)$$

The confidence interval for a_0 expressed by the formula can be calculated from:

$$\text{c.i.} = \pm\, t_{\alpha,n\text{-}2} \cdot s_{a_0} \qquad (17)$$

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE3\Exe03.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames

3. **Calculate the statistical parameters using the formulas described and compare the results with values calculated using standard spreadsheet functions and data analysis (regression). Present in a separate table the results of the constant value calculations according to the confidence interval and place a conclusion drawn from the size of the correlation coefficient and the determination coefficient.**

4. Make a graph illustrating the relationship between the experimentally measured concentrations and time i.e. C]$_t$=f(t), in the form of points and a trend line calculated from the equation (18) (transformed equation (1)):

$$[C]_t = [C]_{t\to\infty} - [C]_{t\to\infty} \cdot e^{-kt} \qquad (18)$$

where $[C]_{t\to\infty} = e^{a_0}$ and $k = a_1$.

5. Create a graph of the logarithmic dependence:

$$Y = \ln\left([C]_{t\to\infty} - [C]_t\right) = f(t) \tag{19}$$

calculated from the experimental data. In the same figure present the relationship calculated from the linear regression analysis $\hat{Y} = f(t)$ as a continuous line without exposing the estimated values, in the form of points.

6. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

EXERCISE No. 4

CALCULATION OF THE pH OF THE TWO ACIDS MIXTURE

A mixture of two acids HM and HP has been prepared with total concentration [C] = [HM] + [HP]. The mole fraction of HM acid in the subsequent mixtures was as follows:

$X_1=0.9$ $X_2=0.8$ $X_3=0.7$ $X_4=0.6$ $X_5=0.5$ $X_6=0.4$ $X_7=0.3$ $X_8=0.2$ $X_9=0.1$

pK for acids – see tab. 1, a teacher gives the concentration value [C].

In order to complete the task, derive a polynomial binding total concentration of hydrogen ions [H] with acid dissociation constants, their molar fraction (X) and the total concentration[C]. Use the SOLVER add-in to calculate [H] satisfying the derived equation for each value of X_{HM}. Then, using the estimated values of [H], calculate:

$$pH= -\log[H] \qquad (1)$$

1. Anion M^- and P^- concentration.

$$[M]=K_{HM}[C]X/([H]+K_{HM}) \qquad (2)$$
$$[P]=K_{HP}[C](1-X)/([H]+K_{HP}) \qquad (3)$$

2. Concentration of undissociated acid HM and HP

$$[HM]=[C]X-[M] \qquad (4)$$
$$[HP]=[C](1-X)-[P] \qquad (5)$$

3. Dissociation degree

$$\alpha_{HM}=[M]/[C]X$$
$$\alpha_{HP}=[P]/([C](1-X)) \qquad (6)$$

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\EXERCISE4\Exe04.xls, where AA denote the number of class, BB – user number.
2. Prepare a reference calculation block for X=0.1.
3. Perform calculations by making successive corrections of the result to get the maximum compatibility between the left and right sides of the equation.
4. Perform calculations for other values of X after you copy and modify the reference block.
5. Make dependency graphs:

   $pH=f(X)$ $[M]=f(X)$ $[P]=f(X)$ $\alpha_{HM}=f(X)$ $\alpha_{HP}=f(X)$

6. For **X=0.1 find the root using:**
   a) bisection method between $[H]_1=0$ and $[H]_2=1$,
   b) secant method between$[H]_1= -2$ and $[H]_2=1$,
   c) tangent method (Newton-Raphson method) between $[H]_1=0$ and $[H]_2=1$.
7. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

APPENDIX I

Derive the equation

$$HP \leftrightarrow H^+ + P^- \qquad (1)$$

$$HM \leftrightarrow H^+ + M^- \qquad (2)$$

$$K_{HM} = \frac{[H][M]}{[HM]} \qquad (3)$$

$$K_{HP} = \frac{[H][P]}{[HP]} \tag{4}$$

$[HM] = [C]X – [M]$ (5)

$[HP] = [C](1 –X) – [P]$ (6)

$$[H] = \frac{K_{HM}([C]X - [M])}{[M]} \tag{7}$$

$[H] = [M] + [P]$ (8)

$[M] = [H] – [P]$ (9)

$$[H] = \frac{K_{HM}([C]X - [H] + [P])}{[H] - [P]} \tag{10}$$

$[H]^2 – [H][P] = K_{HM}[C]X – K_{HM}[H] + K_{HM}[P]$ (11)

$K_{HM}[P] +[H][P] = [H]^2 + K_{HM}[H] – K_{HM}[C]X$ (12)

$$[H] = \frac{K_{HP}([C](1 - X) - [P])}{[P]} \tag{13}$$

$$[P] = \frac{[H]^2 + K_{HM}[H] - K_{HM}[C]X}{K_{HM} + [H]} \tag{14}$$

$$[H] = \frac{K_{HP}\left\{[C](1 - X) - \dfrac{[H]^2 + K_{HM}[H] - K_{HM}[C]X}{K_{HM} + [H]}\right\}}{\dfrac{[H]^2 + K_{HM}[H] - K_{HM}[C]X}{K_{HM} + [H]}} \tag{15}$$

$$[H] = \frac{K_{HP}\{[C](1 - X)(K_{HM} + [H]) - [H]^2 - K_{HM}[H] + K_{HM}[C]X\}}{[H]^2 + K_{HM}[H] - K_{HM}[C]X} \tag{16}$$

$[H]^3 + (K_{HM} + K_{HP})[H]^2 + \{K_{HP}K_{HM} – K_{HM}[C]X – K_{HP}[C](1\text{-}X)\}[H] = K_{HM}K_{HP}[C]$ (17)

EQUATION FOR CALCULATION:

$$a[H]^3 + b[H]^2 + c[H] = 1$$

where

$$a = \frac{1}{K_{HP}K_{HM}[C]} \qquad b = \frac{K_{HM} + K_{HP}}{K_{HM}K_{HP}[C]} \qquad c = \frac{1}{[C]} - \frac{X}{K_{HP}} - \frac{1 - X}{K_{HM}}$$

APPENDIX II
TAB.1

Acid dissociation constant values:

ACID	pK	K
formic	3.75	1.78×10^{-4}
lactic	3.86	1.38×10^{-4}
acetic	4.75	1.75×10^{-5}
propionic	4.87	1.33×10^{-5}

APPENDIX III (REMINDER)

Bisection method

In the bisection method, to determine the approximate zero of a function, the interval $\langle x_1,x_2\rangle$ gradually decreases so as to contain the element sought. The starting point in this method is two argument values for which the function $f(x)$ changes its sign.
In the first step we calculate $f(x_3)$ at the midpoint of the interval:

$$x_3 = ½\cdot(x_1+x_2)$$

If $f(x_3)>0$, then the solution is between points x_1 and x_3:

$$x_4 = ½\cdot(x_1+x_3)$$

The calculations are repeated several times until a sufficiently good estimate of zero is obtained. In practice, the iterative calculations end after fulfilling any of the following conditions:

$$\left|x_{n+1}-x_n\right|<\varepsilon$$

which means that the difference between successive approximations is small enough, or:

$$\left|f(x_n)\right|<\varepsilon$$

i.e. the value of the function at the designated point is close to 0 (lower than ε). In these equations, ε is the assumed accuracy of calculations (criterion specified by a user). These equations are also used in the secant and tangent methods.

Secant method (regula falsi)

In this method, also called the false position method, a chord is drawn through points x_1 and x_2, for which the function $f(x)$ changes its sign, with the following equation:

$$y-f(x_1)=\frac{f(x_2)-f(x_1)}{x_2-x_1}(x-x_1)$$

The abscissa x_3 of the point at which the fixed chord AB intersects the axis OX (Fig. 7.4), is assumed as the first approximation of the desired zero location.

$$x_3=x_2-f(x_2)\frac{x_2-x_1}{f(x_2)-f(x_1)}$$

etc.
The general recursive equation can be written as:

$$x_{(k+2)}=x_{(k+1)}-f(x_{(k+1)})\frac{x_{(k+1)}-x_k}{f(x_{(k+1)})-f(x_k)}$$

where k = 1,2, ...

Tangent method (Newton-Raphson)

In this method it is necessary to know the function *f*(*x*) and its derivative *f'*(*x*). The slope of the tangent to the plot at point x_2 can be evaluated from:

$$f'(x_2) = \frac{f(x_2)}{x_2 - x_3}$$

The first approximation of the root (x_3) can be calculated from the equation:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)}$$

The general recursive formula is as follows:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

EXERCISE No. 5

MULTIPLE LINEAR REGRESSION

The general linear multiple regression equation for *p* independent variables takes the form:

$$y = a_0 + a_1x_1 + a_2x_2 + ... + a_px_p \tag{1}$$

Linear dependence coefficients can be determined in a simple way using the method of least squares from which the following dependence is obtained for a vector of regression coefficients (***a***):

$$\boldsymbol{a = (x^{T}x)^{-1}x^{T}y} \tag{2}$$

where ***x*** is the matrix of value *x*, ***y*** – the matrix of value *y*:

$$\boldsymbol{x} = \begin{bmatrix} 1 & x_{1(1)} & x_{2(1)} & \cdots & x_{p(1)} \\ 1 & x_{1(2)} & x_{2(2)} & \cdots & x_{p(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1(n)} & x_{2(n)} & \cdots & x_{p(n)} \end{bmatrix}, \ \boldsymbol{y} = \begin{bmatrix} y_{(1)} \\ y_{(2)} \\ \vdots \\ y_{(n)} \end{bmatrix} \tag{3}$$

$\boldsymbol{x}^{\mathbf{T}}$ transposition of the matrix ***x*****,** and $\boldsymbol{(x^{T}x)^{-1}}$ inverse of the matrix product.

The next stage of regression analysis is the assessment of a model's goodness of fit. Adequate sums of squared deviations resulting from the regression function (Q_2), experimental errors (Q_3) and the total variability (Q_1) can be calculated from the following formulas:

$$Q_2 = \sum(\hat{y}_i - \bar{y})^2 = \boldsymbol{a^{T}x^{T}y} - n\cdot\bar{y}^2 \tag{4}$$

$$Q_3 = \sum(y_i - \hat{y}_i)^2 = \boldsymbol{y^{T}y - a^{T}x^{T}y} \tag{5}$$

$$Q_1 = \sum(y_i - \bar{y})^2 = \boldsymbol{y^{T}y} - n\cdot\bar{y}^2 \tag{6}$$

where *n* de notes the number of of data points, $\bar{y}$ – the mean value the dependent variable.
The coefficient of determination (r^2) can be calculated from:

$$r^2 = Q_2/\,Q_1 \tag{7}$$

In analytical chemistry, linear models are widely used in calibration. At the same time, an explained (response) variable relatively rarely depends only on one explanatory variable.

In the case of atomic absorption spectrometry, the analytical signal value, measured with a solution of fixed concentration of a determined element, is affected by many factors. These factors may be spectral (frequency of emitted or absorbed radiation, atomic energy transition probability, statistical weights of energy states, and others), or related to the transport of a solution to the flame (determined by the so–called nebulization efficiency), conditions in the flame (composition, shape and temperature of the flame) and reactions occurring in it (e.g. ionization of atoms of the determined element, dissociation of its salt particles, formation of chemical compounds with particles of flame gases). The presence of other substances in the test solution (in addition to the determined metal cation) can be a source of spectral interferences (involving mainly the coincidence of the spectral lines of these components), or changes in physical properties of the solution (viscosity, surface tension) and, consequently, changes in nebulization efficiency.

Components accompanying the determined element may affect the analytical signal in various ways. Using the modified data presented below [P.C. Jurs, *Computer Software Applications in Chemistry,* J. Wiley, New York 1996], determine the coefficients of linear dependence between the analytical signal *R* (dependent variable) and the concentrations c_1, c_2, c_3 of accompanying components (independent variables).

c_1 [mol dm^{-3}]	c_2 [mol dm^{-3}]	c_3 [mol dm^{-3}]	*R*
0.071	0.288	0.107	0.425
0.107	0.265	0.102	0.779
0.150	0.264	0.107	0.937
0.217	0.268	0.101	0.646
0.295	0.268	0.113	1.010
0.338	0.290	0.113	0.485
0.361	0.264	0.107	0.853
0.488	0.266	0.117	1.144
0.538	0.271	0.102	0.410
0.597	0.259	0.111	1.015
0.636	0.267	0.106	0.637
0.718	0.284	0.110	0.349
0.746	0.288	0.102	0.073
0.823	0.269	0.114	0.769
0.838	0.275	0.108	0.415
0.852	0.264	0.110	0.744
0.972	0.267	0.111	0.656
1.052	0.265	0.107	0.518
1.044	0.277	0.116	0.595
1.133	0.277	0.102	0.012

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE5\Exe05.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames.

3. **Calculations:**

 a) **Calculate the coefficients of the equation and the linear correlation for each pair of variables separately (using any method).**

$$R=a_{0(1)}+a_{1(1)}\cdot c_1$$
$$R=a_{0(2)}+a_{1(2)}\cdot c_2$$
$$R=a_{0(3)}+a_{1(3)}\cdot c_3$$

 b) **Calculate the coefficients of the equation and the linear correlation for the following relations (using any method):**

$$R=a_{0(12)}+a_{1(12)}\cdot c_1+a_{2(12)}\cdot c_2$$
$$R=a_{0(13)}+a_{1(13)}\cdot c_1+a_{2(13)}\cdot c_3$$
$$R=a_{0(23)}+a_{1(23)}\cdot c_2+a_{2(23)}\cdot c_3$$

 c) **Perform statistical calculations using the equations((2)-(7)) given in the description. Compare the results (Q_1, Q_2, Q_3, r^2) with the values calculated using the standard procedure of the spreadsheet (Regression) and the SOLVER add-in (only regression coefficients). If differences occur, put an explaining comment. Present the results of the calculations (regression coefficients) with the corresponding confidence intervals in a separate table.**

4. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

EXERCISE No. 6

LINEAR REGRRESION –LINEARIZING TRANSFORMATION

Equations used to describe experimental data in chemistry are often nonlinear. At the same time, in many cases a non-linear model, through a simple transformation (substitution of variables) can be reduced to a linear dependence. Typical non-linear functions and appropriate linearizing substitutions are shown below:

Nonlinear function	Linearizing substitutions

$y = a + \frac{b}{x}$	Y = y
	X = $1/x$

$\frac{1}{y} = a + b \cdot x$	Y = $1/y$
	X = x

$y = a \cdot b^x$	Y = log(y)
	X = x

$y = a \cdot x^b$	Y = log(y)
	X = log(x)

$y = a \cdot e^x$	Y = ln(y)
	X = x

$y = a + b \cdot x^n$	Y = y
	X = x^n

$y = \frac{a \cdot x}{b + x}$	Y = x/y or Y = $1/y$
	X = x X = $1/x$

Exercise *a)*

According to the Arrhenius-Guzman equation, viscosity dependence on temperature takes the form:

$$\eta = A \cdot e^{\frac{E_\eta}{RT}} \qquad (1)$$

where E_η denote the activation energy [J·mol^{-1}], T – temperature [K], R – is the gas constant [J·K^{-1}·mol^{-1}].

Based on the experimental results (Tab.1.) [J. Demichowicz-Pigoniowa, Obliczenia fizykochemiczne, PWN, Warszawa, 1984] determine the values of constants A and E_η.

Tab.1. The measured values of viscosity as a function of temperature

T [K]	$\eta \cdot 10^3$ [N s m^{-2}]
288.16	2.1858
291.16	2.0211
298.16	1.7017
308.16	1.3428
318.16	1.0960
328.16	0.9095

- Make a chart (Fig.1.) showing the dependence of the experimentally measured values of viscosity as a function of temperature ($\eta = f(T)$),
- Present on a chart (Fig.2.) the linear dependence obtained from the transformation with the corresponding trend line, the equation and the value of r^2,

- Determine the regression coefficients in the linear equation from the corresponding formulas, data analysis, and using the SOLVER add-in,
- In a separate table (Tab.2.) present the results of the calculations (coefficients of the model), confidence intervals and the corresponding dimension of the determined coefficients.

Exercise *b)*

The Arrhenius equation describes the dependence of reaction rate on temperature:

$$k = A \cdot e^{-\frac{E_a}{RT}} \tag{2}$$

where k – denote the reaction rate constant [s^{-1}], E_a – the activation energy [J·mol^{-1}], R – is the gas constant [J·K^{-1}·mol^{-1}], T – temperature[K].
Based on the experimental results (Tab.3.) [J. Demichowicz-Pigoniowa, Obliczenia fizyko-chemiczne, PWN, Warszawa, 1984] determine the values of activation energy and A.

Tab.3. The measured values of reaction rate constant as a function of temperature

T [K]	k [s^{-1}]
273	$7.8 \cdot 10^{-7}$
298	$3.3 \cdot 10^{-5}$
318	$5.0 \cdot 10^{-4}$
338	$5.0 \cdot 10^{-3}$

- Make a chart (Fig.1.) showing the dependence of the experimentally measured values of reaction rate constant as a function of temperature ($k = f(T)$),
- Present on a chart (Fig.2.) the linear dependence obtained from the transformation with the corresponding trend line, the equation and the value of r^2,
- Determine the regression coefficients in the linear equation from the corresponding formulas, data analysis, and using the SOLVER add-in,
- In a separate table (Tab.2.) present the results of the calculations (coefficients of the model), confidence intervals and the corresponding dimension of the determined coefficients.

Exercise *c)*

Carboxylic acid adsorption isotherm on activated carbon can be described by equations:

$$\frac{x}{m} = k \cdot c^{\frac{1}{n}} \tag{3}$$

$$\frac{x}{m} = \frac{a \cdot b \cdot c}{1 + b \cdot c} \tag{4}$$

where x/m denote the weight of the acid adsorbed per unit weight of adsorbent [g/g], c – equilibrium concentration of the acid [mol dm^{-3}], k, n, a, b – isotherm equations constants.
Based on the experimental results (Tab.5.) determine the values of isotherm equations constants.

Tab.5. The measured values of the weight of the acid adsorbed per unit weight of adsorbent as a function of concentration

x/m [g/g]	c [mol dm^{-3}]
0.1043	0.2103
0.07638	0.09373
0.05835	0.04038
0.04761	0.01847
0.02814	0.007074

- Make a chart (Fig.1.) showing the dependence of the experimentally measured values of the weight of the acid adsorbed per unit weight of adsorbent as a function of concentration (*x/m* = *f*(*c*)),
- Present on a chart (Fig.2.) the linear dependence obtained from the transformation with the corresponding trend line, the equation and the value of r^2,
- Determine the regression coefficients in the linear equation from the corresponding formulas, data analysis, and using the SOLVER add-in,
- In a separate table (Tab.2.) present the results of the calculations (coefficients of the model), confidence intervals and the corresponding dimension of the determined coefficients.

Exercise *d)*

The rate of enzymatic reactions, can be described by the Michaelis-Menten equation:

$$r = \frac{r_{\max} \cdot [S]}{[S] + K_{MM}} \tag{5}$$

where K_{MM} denote the Michaelis-Menten constant [mol·dm^{-3}], [*S*] – concentration of a substrate [mol·dm^{-3}], $r_{\max}$ – the maximum reaction rate [mol·dm^{-3}·s^{-1}].
Based on the experimental results (Tab.7.) [J. Demichowicz-Pigoniowa, Obliczenia fizyko-chemiczne, PWN, Warszawa, 1984] determine the values of K_{MM} and $r_{\max}$

Tab.7. The measured values of the reaction rate as a function of concentration [*S*]

[*S*] [mol·dm^{-3}]	r·10^3 [mol·dm^{-3}·s^{-1}]
0.0052	0.256
0.0104	0.403
0.0208	0.616
0.0416	0.823
0.0833	0.985
0.1670	1.082
0.3330	1.087

- Make a chart (Fig.1.) showing the dependence of the experimentally measured values of the of the reaction rate as a function of concentration (*r* = *f*([*S*])),
- Present on a chart (Fig.2.) the linear dependence obtained from the transformation with the corresponding trend line, the equation and the value of r^2,
- Determine the regression coefficients in the linear equation from the corresponding formulas, data analysis, and using the SOLVER add-in,
- In a separate table (Tab.2.) present the results of the calculations (coefficients of the model), confidence intervals and the corresponding dimension of the determined coefficients.

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE6\Exe06.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames.

3. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

EXERCISE No. 7

NUMERICAL INTEGRATION
THE RECTANGULAR, TRAPEZOIDAL AND SIMPSON’S RULE METHOD.

I. INTRODUCTION

To approximate the definite integral $\int_a^b f(x)dx$ using a numerical methods,), the interval $\langle a, b\rangle$ is divided into n equal subintervals $\Delta x = \frac{b-a}{n}$.

For designated points from the interval $x_1, x_2, \ldots, x_{n-1}$ the values of the integrand $y = f(x)$ ($y_0 = f(a)$, $y_1 = f(x_1)$, …, $y_{n-1} = f(x_{n-1})$, $y_n = f(b)$) were calculated.
In the final calculation, the following equations can be used:

1. Rectangle method

$$\int_a^b f(x)dx \approx \Delta x(y_0 + y_1 + \ldots + y_{n-1})$$

2. Trapezoidal method

$$\int_a^b f(x)dx \approx \Delta x\left(\frac{y_0 + y_n}{2} + y_1 + \ldots + y_{n-1}\right)$$

3. Simpson method (n must be even)

$$\int_a^b f(x)dx \approx \frac{\Delta x}{3}\left[y_0 + y_n + 4(y_1 + y_3 + \ldots y_{n-1}) + 2(y_2 + y_4 + \ldots y_{n-2})\right]$$

II. CALCULATIONS

1. **Estimate the definite integral:**

a) $\mathrm{B} = \int_1^7 \frac{tdt}{\sqrt{1+3t}}$ b) $\mathrm{V} = \int_1^{2.6} \frac{dc}{c}$

using the rectangular, trapezoidal and Simpson’s rule method with n = 6, 8, 10 steps.
The results compare in the table:

n	Rectangular rule	Trapezoidal rule	Simpson’s rule
6			
8			
10			

2. **Determine the definite integral:**

$$D = \int_3^{15} f(t)dt$$

$c = f(t)$

for a given results, using the rectangle, trapezoid and Simpson methods.

t	c
3	5.531
4	6.302
5	6.625
6	6.578
7	6.239
8	5.686
9	4.997
10	4.25
11	3.523
12	2.894
13	2.441
14	2.242
15	2.375

Determine a regression equation (3-degree polynomial) describing the presented dependence and calculate the analytical value of the integral. Knowing the analytical value of the integral *D*, determine a relative error for different methods of integration.

3. **Determine the respective definite integrals using the rectangle, trapezoid and Simpson methods.**

The standard heat of iodine hydrogen from iodine and hydrogen formation at 1000 K can be calculated from the following equation [J. Demichowicz-Pigoniowa, Obliczenia fizykochemiczne, PWN, Warszawa, 1984]:

$$\Delta H^o_{r,1000} = \Delta H^o_{r,298} + \int_{298}^{438} \sum \nu_i \mathscr{C}^o_{p,i} dT - \tfrac{1}{2}\Delta H^o_{p.f.k} + \int_{438}^{1000} \sum \nu_i \mathscr{C}^o_{p,i} dT$$

where $\Delta H^o_{r,298}$ is the standard enthalpy of formation of hydrogen iodide in temperature 298 K (25.94 kJ·mol^{-1}), $\Delta H^o_{p.f.k}$ the heat of sublimation of iodine (59.8·kJ mol^{-1}).

The sum of the molar heat capacity for the temperature range ⟨298,438⟩ is::

$$\sum \nu_i \mathscr{C}^o_{p,i} = -6.12 - 22.94 \cdot 10^{-3}\, T + 1.00 \cdot 10^{-6}\, T^2 \quad [\text{J} \cdot \text{K}^{-1}]$$

For the temperature range ⟨438,1000⟩:

$$\sum \nu_i \mathscr{C}^o_{p,i} = -4.76 - 1.66 \cdot 10^{-3}\, T + 1.00 \cdot 10^{-6}\, T^2 + 0.36 \cdot 10^{5}\, T^{-2} \quad [\text{J} \cdot \text{K}^{-1}]$$

The respective definite integrals calculate using the rectangle, trapezoid and Simpson methods (*n*=10).

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE7\Exe07.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames

3. **Calculate the definite integrals using the appropriate equations.**

4. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

EXERCISE No. 8

NUMERICAL SOLVING OF DIFFERENTIAL EQUATIONS
EULER, RUNGE – KUTTA, MILNE METHODS

An ordinary first order differential equation can be presented as follows:

$$y^{'}(x)=\frac{dy}{dx}=f(x) \tag{1}$$

The solution is function $y(x)$ satisfying this equation and one of the initial conditions $y(x_0) = y_0$.

A typical example of a differential equation application is a description of changes in substrate concentration during a reaction. For example, the kinetic equation for an irreversible first-order reaction is as follows:

$$\frac{dc}{dt}=-k\cdot c \tag{2}$$

where k denote reaction rate constant [s^{-1}], c – concentration of substrate [mol/dm^3], t –time [s].

This equation, after solving, leads to dependence of substrate concentration on time:

$$c=c_0\cdot e^{-k\cdot t} \tag{3}$$

where c_0 is the initial concentration of substrate [mol/dm^3]

The concentration-time dependence can also be calculated using an appropriate numerical method for solving differential equations. For this purpose, we can use e.g. the Euler method, Runge-Kutta and predictor-corrector methods. In the case of numerical methods for solving differential equations, it is necessary to define the starting point (x_0, y_0) and the slope of the function which is the solution of the equation at that point (y’).

1. Euler method

In the Euler method, the value of the function at point x_0+Δx (y_1) is calculated from the equation:

$$y_1=y_0+\Delta y=y_0+(\Delta x)\times f(x_0,y_0) \tag{4}$$

where $f(x_0, y_0)$ is equal to the slope of the function being the solution at that point. The general equation can be given as follows:

$$y_{n+1}=y_n+\Delta y=y_n+(\Delta x)\cdot f(x_n,y_n) \tag{5}$$

2. Runge-Kutta method

In the fourth-order Runge-Kutta method, adequate calculations can be performed using the following formulas:

$$y_{i+1}=y_i+\frac{1}{6}\Delta x\left(c_1+2c_2+2c_3+c_4\right) \tag{6}$$

$$c_1 = f(x_i y_i) \tag{7}$$

$$c_2 = f(x_i + \frac{1}{2}\Delta x,\ y_i + \frac{1}{2}\Delta x c_1) \tag{8}$$

$$c_3 = f(x_i + \frac{1}{2}\Delta x,\ y_i + \frac{1}{2}\Delta x c_2) \tag{9}$$

$$c_4 = f(x_i + \Delta x,\ y_i + \Delta x\, c_3) \tag{10}$$

where c_1 denote the value of the slope of the solution function at the starting point of the interval (x=x_0) c_2 and c_3 – the values of the slope at the midpoint, c_4 the value of the slope at the end of the interval.

3. Milne method (predictor-corrector)

An alternative method for solving differential equations is the multi-step Milne method (predictor-corrector). In this method, we need to have the values:

$$\begin{matrix} y_0, & y_{-1} & y_{-2} & y_{-3} \\ y'_0, & y'_{-1} & y'_{-2} & y'_{-3} \end{matrix} \tag{11}$$

and the function:

$$\frac{dy}{dx} = f(x, y) \tag{12}$$

Calculations are performed according to the following equations:

$$y_{1,p} = y_{-3} + \frac{4\Delta x}{3}\left(2y'_{-2} - y'_{-1} + 2y'_0\right) \tag{13}$$

$$y'_{1p} = f(x_1, y_{1p}) \tag{14}$$

$$y_{1c} = y_{-1} + \frac{\Delta x}{3}\left(y'_{-1} + 4y'_0 + y'_1\right) \tag{15}$$

$$y'_{1c} = f(x_1, y_{1c}) \tag{16}$$

where $y_{1,p}$ denote the predicted value of y_1, y'_{1p} the estimated value of a derivative at point x_1, y_{1c} *a* corrected value of y_1, y'_{1c} a corrected derivative of the estimated value at point x_1.

I. CALCULATIONS

Using the equation (3), and assuming that k = 0.18 s^{-1}, c_0 = 0.1 mol/dm^3 and Δt = 1 s, calculate the substrate concentration changes over time ($t_{\max}$=18 s).

Calculate the dependence of the substrate concentration on time, applying the Euler (Equation (5), Milne (Equations (13)-(16)) and Runge-Kutta (Equations (6)-(10)) methods.
Knowing the actual values of the concentration (Equation (2)) and the results obtained for each method, calculate the relative error [%]. In the Milne method, use the initial values calculated with the Runge-Kutta method as the starting points.

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE8\Exe08.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames

3. Make graph 1 illustrating the concentration-time dependence $c = f(t)$ calculated from the equation (3) in the form of points. Present the dependencies calculated with the Euler and Runge-Kutta methods on the same graph as a continuous line without exposing the estimated values c in the form of points

4. Make a graph 2 illustrating the dependence $\ln(c/c_0) = f(t)$ calculated from equation (3) in the form of points. Present the dependencies calculated with the Euler and Runge-Kutta methods ($\ln(c/c_0)$) on the same graph as a continuous line without exposing the estimated values c in the form of points

5. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

EXERCISE No. 9

SIMPLEX OPTIMIZATION

A simplex is a geometric shape with regular edges and $n + 1$ vertices (n is the number of optimized parameters). Below examples of simplexes in the one-, two-and three-dimensional space [1] are presented.

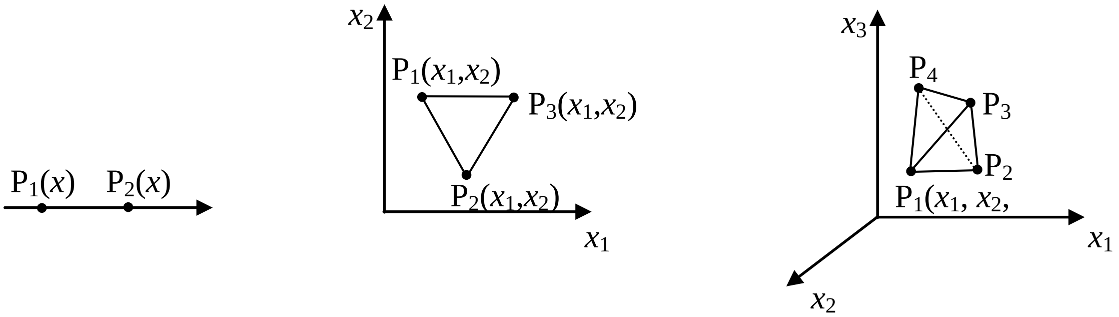


Fig.1. The simplexes in the one-, two-and three-dimensional space.

The simplex method is based on a systematic analysis of the response surface to locate the optimum response function. Optimization begins with generating an initial simplex (n+1 experiments). Gorsky and Brodsky [2] proposed a method with the coordinate origin in the simplex centre. A corresponding initial matrix (3) can be generated from the following formulas:

$$k_i = \left(\frac{1}{2i \cdot (i+1)} \right)^{\frac{1}{2}} \tag{1}$$

$$R_i = \left(\frac{i}{2 \cdot (i+1)} \right)^{\frac{1}{2}} \tag{2}$$

The general form of the matrix is as follows:

$$A = \begin{bmatrix} k_1 & k_2 & \cdots & k_{n-1} & k_n \\ -R_1 & k_2 & \cdots & k_{n-1} & k_n \\ 0 & -R_2 & \cdots & k_{n-1} & k_n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -R_{n-1} & k_n \\ 0 & 0 & \cdots & 0 & -R_n \end{bmatrix} \tag{3}$$

For n = 3, the matrix A can be written in the form:

$$A = \begin{bmatrix} 0.500 & 0.289 & 0.204 \\ -0.500 & 0.289 & 0.204 \\ 0 & -0.578 & 0.204 \\ 0 & 0 & -0.612 \end{bmatrix} \tag{4}$$

The matrix A (initial simplex) is expressed in dimensionless units and shows n+1 parameter values of individual experiments. Designated coordinates can be obtained from a simple calculation using the following formula:

$$x_{mi} = x_{0,i} + z_i A_i \tag{5}$$

x_{mi} is a designated value of the *i*-th parameter, $x_{0,i}$ – designated output value of the *i*-th parameter, z_i – designated value of a unit on the axis of the variable, *A* – dimensionless value of the *i*-th parameter corresponding to the value from the matrix *A*

After a series of experiments (initial simplex), the results are assessed in terms of a feature that best characterizes the outcome (quality criterion). Then we select an experiment whose quality criterion is the lowest (point C, Fig.2). This point is replaced by a new one (point D), symmetrical to the point with the lowest quality criterion, formed by symmetrical reflection with respect to the opposite edge of the simplex.

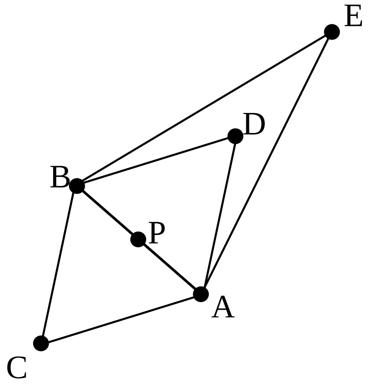


Fig. 2. Reflection (D) and expansion (E) in the simplex method.

The coordinates of the new point, symmetrical to the rejected point (for each parameter separately) can be calculated from the formula:

$$D = P + (P - C) \tag{6}$$

where P is the mean of all parameter values without the dismissed result, C – values of the rejected point parameters.

In the case of significant growth of the response functions at the new point, it is possible to expand the simplex in that direction (point E). The coordinates of the expanded point can be calculated from the formula

$$E = D + (P - C) \tag{7}$$

If the criterion of quality at point D is not worse than the result at the rejected point and is not better than the remaining ones, simplex contraction can be applied. Positive (point K^+, Fig. 3.) or negative contractions (point K^-, Fig. 3.) can be used.

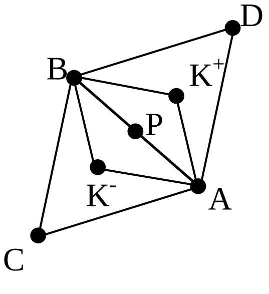


Fig. 3. Positive contraction (K^+) and negative contraction (K^-) in the simplex method

The coordinates of the corresponding points can be calculated with the formulas:

$$K^+ = P + (P - C)/2 \tag{8}$$

$$K^- = P - (P - C)/2 \tag{9}$$

The response surface analysis ends as soon as it reaches the optimum area of a selected optimization criterion.

EXCERCISE:

A chemical yield (*WR*) depends on concentration (*c*) and temperature (*T*) and is described by the following equation:

$$WR = (725 - (10 - c)^2 - (20 - T)^2)/7.25$$

Locate the maximum yield with the simplex method. Generate the initial simplex for the following values the of parameters (*x*) and step (*z*):

$$x_{0,c} = 3.5 \text{ mol/dm}^3,\ z_c = 2 \text{ mol/dm}^3$$
$$x_{0,T} = 10\ ^{o}\text{C},\ z_T = 2\ ^{o}\text{C}$$

COMMENTS:

1. Create a new spreadsheet and save it to a network drive in the directory S:\PinfAABB\EXERCISE\ EXERCISE9\Exe09.xls, where AA denote the number of class, BB – user number.

2. Create a table with data set used for the calculations along with any scheduled frames. Pay attention to the careful planning of tables, descriptions, and frames.

3. **Calculations:**
   After performing the calculations (finding the maximum yield), generate a simplex at a point near the maximum, where the step is: z_c = 0.5 mol/dm^3, z_T = 0.5 ^{o}C. Determine the maximum yield for the new simplex.
   Present the calculated simplex points (*c*, *T*) on a graph.

4. Prepare a worksheet for printing with margins: left 3 cm and upper 2 cm. Information concerning the user should be placed in a footer or header (optional).

[1] R. Wódzki, J. Ceynowa, Sympleksowa metoda planowania doświadczeń ekstremalnych, *Wiadomości Chemiczne*, 1976, 30, 327

[2] W. G. Gorskij, W.Z. Brodskij, *Zawod. Łab.*, 1968, 34, 7, 838