Mathematics for Chemistry with Symbolic Computation

J. F. Ogilvie @ 2016 June 30

2016 June, edition 5.0

The software in these fifteen *Maple* worksheets and associated materials comprising this book, Parts I and II, is freely distributed without cost from www.cecm.sfu.ca. Any sale of this software whatsoever is unauthorised and contrary to the terms of ownership of the copyright of this material by J. F. Ogilvie and contributors.

This interactive electronic book is organised into several separate computer files, each called a *Maple* worksheet; each worksheet contains executable commands and assignments in a sequence, with interspersed text to explain their purposes and effects, but particularly to introduce mathematical concepts and principles and to implement the mathematical operations. These assignments and commands are generally intended to be executed in a linear sequence from the top to the bottom of each section or page, and the user is intended to read the interspersed text to understand the progression of these commands. Each file of this book is a document, comprising a single chapter that contains text, and commands executable on a computer operating software *Maple* ^{TM}; this electronic document, a worksheet, is operable entirely interactively: a user reads the text, executes the commands and views the output; according to an experimental approach that is encouraged, a user can freely modify the command to discover the nature of the correspondingly altered output. We recommend that a user retain a separate archival copy of this original document with no output from *Maple*, and that in general a worksheet containing output should not be saved on exit from *Maple* unless such an archival copy of the original document be available; if alterations to any text or command in a worksheet of this electronic book be essential, one should purge the worksheet of output, through use of the command `Remove output` at the bottom of menu `Edit` near the top left corner of this display, and collapse all sections, through use of the bottom command in menu `View` also near the top left of this display.

Within this particular worksheet that includes introductory chapter 0, there are also a preface in a traditional manner, a table of contents and a brief overview of *Maple*. In its separate file, each chapter that treats a traditional area of arithmetic or mathematics comprises in turn groups of sections under a particular title; each section, identifiable on *Maple*'s *classic* interface by a grey square containing a plus sign, +, at the left margin and a heading to the right thereof, is normally collapsed so that content remains hidden until that section is opened according to a mechanism described below. The material is equally accessible with the 'standard' interface in the 'worksheet' mode. To view content of such an indented section with the *classic* interface, set, with a mouse or equivalent device that serves as

pointer, the cursor on a chosen square containing such "+" and depress the left lever of the mouse, or equivalent mechanism of activation; to close an open section for compactness, click on "-" in its square. With *Maple*'s *standard* interface, which tends to operate slowly with large files on old computers because of its implementation in Java, a section is identifiable by a grey triangle at the left margin ">" that points to its right side and its accompanying heading; opening that section by placing the curson on that triangle and depressing the left lever of a computer mouse, or equivalent mechanism, causes the triangle to point downward,and closing that section involves the same mechanism -- 'clicking' left with the cursor on the arrow. With such indented material exposed, one can read the text, in generally black letters as in this paragraph, proceeding down the worksheet in various ways; one can execute an input statement on a line of red characters wherever it appears by ensuring that the cursor is located before the end of that line, even at the first colon or semicolon in that line. Reversing the mechanism to open that section causes the section to become closed and collapsed so that the content is concealed from view. Interspersed among these sections are many examples and exercises, also present normally in collapsed or hidden form. Deemed to be an integral component of the learning or discovery, by means of symbolic computation, these examples and exercises of the presented mathematics should be examined and undertaken in a linear temporal order down any worksheet, and each worksheet in increasing order of chapter number; each example bears an identification at the left margin with initial character *x*, and exercises and their solutions are named explicitly. The reader should attempt the solution of the exercise in a separate 'Window'. The solutions of exercises are available in a separate worksheet.

Using the left lever of a mouse, or equivalent device, to *click* on any word or set of characters displayed both in green letters and underlined activates a hyperlink, which opens a new screen that contains information pertinent to that hyperlink; to return to the former screen after reading that information, one should click on the *lower* x in the upper right corner of the classic display -- that black x with a white or grey background, *not the upper* x in white on a red background or black on a grey background depending on computer, which initiates termination of the *Maple* session. Test this mechanism of hyperlink by clicking on the next word here -- worksheet; so activating a hyperlink is also a general method of quickly altering material displayed on this monitor from one point in a worksheet to either

- a remote point in the same worksheet, or

- another worksheet that is located appropriately or for which an external linking mechanism is arranged, or

- a specific page that provides help on a pertinent topic in relation to usage of *Maple* comprising descriptions of syntax, data types and functions, or

- a pertinent page in a dictionary of mathematical terms within *Maple*, or

- a remote site through an internet browser.

The pages of either `help` or `dictionary` generally contain further hyperlinks to other pages of the same or other type. In worksheets of this book, almost all such hyperlinks in this electronic document invoke pages of `Help` or the dictionary, thus not requiring a connexion to internet.

For these files we commend use of *Maple* in both release 11 or subsequent and its **classic** interface;

or the standard interface and its worksheet mode; use of its **standard** interface might be slow if a computer lack sufficient hardware properties such as speed of processor or amount of memory. Within a *Maple* session thus with its classic interface, a square at the left margin that contains a plus sign, "+" , like that just below this paragraph beside a heading **P preface**, indicates an indented section that contains material about a particular topic to which the heading pertains. To view content of such an indented section, set the cursor, with a mouse or equivalent device to serve as pointer, on a chosen square containing "+" and depress the left lever of the mouse or other mechanism of activation; to close an open section for compactness, click on " – " in its square. With such indented material exposed, one can read the text, and proceed down this worksheet in various ways; one can execute an input statement on a line of red characters wherever it appears by ensuring that the cursor is located before the end of that line, even to the left of a colon or semicolon in that line. If, for a particular operating system, a classic interface be not available, use of the standard interface in mode *worksheet* can be set through selection of Tools -> Options -> Interface and setting the default format to be Worksheet, rather than Document, which might then be applied globally -- to all future sessions -- rather than the session in which this option is set.

−   **T table of contents**

     A section of a chapter, according to indented headings below, of which the heading bears an asterisk * might be omitted at first reading.

P   Preface
T   Table of contents
I Mathematical Operations -- Mathematics for Chemistry
        overview of *Maple*
 0   Exemplary illustrations of use of *Maple*
        overview
   0.1 examples of calculations
      11 solution of equations
      12 a difficult derivative
      13 plot of unit cell of crystalline caesium iodide CsI
      14 chemical and physical data in *Maple*
      15 spreadsheet
   0.2 cursory tour
      21 essential commands and operations
      summary
 1   Numbers, symbols and elementary functions
        overview and principles
        seeking help
   1.1 working with numbers real and complex
      1 classification of real numbers
      2 types of real numbers in *Maple*
      3 basic arithmetical operations
      4 assignment
      5 unassignment

overview and principles

6.1 matrix and determinant
    1 formation, properties and plotting of matrix, and determinant
    2 addition and subtraction of matrices
    3 multiplication of matrix by scalar quantity
    4 multiplication of matrix by matrix
    5 raising a square matrix to a non-negative power
    6 transpose of a matrix
    7 adjoint matrix
    8 trace of a square matrix
    9 minor and cofactor
    10 inverse of a matrix
    11 pseudo-inverse of a matrix
    12 raising a matrix to a negative integer power
    13 decomposition of a matrix
    14 relation of matrix to linear equations
    15 solution of linear algebraic equations
    16 numerical precision in calculations with matrices

6.2 vector
    1 forming and plotting a vector
    2 product of a vector with a scalar quantity
    3 addition and subtraction of vectors
    4 basis vectors
    5 specification of a vector
    6 scalar product of two polar vectors
    7 vectorial space
    8 vectorial product of two vectors
    9 unit cell and dual or reciprocal space
    10 dyadic product of two polar vectors
    11 operation of matrix on vector
    12 linear equations in terms of matrix and vector

6.3 eigenvector and eigenvalue
    1 significance of eigenvector and eigenvalue
    2 diagonalization of a matrix
    3 other relations of eigenvalues
    4 matrix with real or complex eigenvalues and eigenvectors
    5 real symmetric and hermitian matrices
    6 degenerate eigenvalues
    7 matrix with complex elements
    8 graphical interpretation of eigenvector

6.4 calculus with vectors
    1 free vector and vectorial field
    2 gradient of a scalar function

3 divergence and curl of a vector
4 gradient of a function or procedure
5 jacobian, hessian, laplacian and wronskian
6.5 tensor
1 quadratic forms
2 tensor in right-angled systems
3 inertial tensor
6.6 spreadsheet
1 description of spreadsheet and elementary operations
2 manipulation of data and transfer of data into a spreadsheet
3 transfer of data from a spreadsheet
summary

7   **Differential and integral equations**
overview and principles
7.1 nature of ordinary-differential equation
1 notation and properties of ordinary-differential equation
2 classification of ordinary-differential equation
3 generation of ordinary-differential equation
7.2 solution of ordinary-differential equation
1 commands to treat an ordinary-differential equation
2 solution of ordinary-differential equation in numeric form
3 ordinary-differential equations in a system
4 phase portrait
5 numerical solution of ordinary-differential equations in a system
6 Laplace transform and application to ordinary-differential equation
7 application of Fourier transform
8 solution in series
9 differential equations and special functions
7.3 ordinary-differential equation of first order and application to chemical kinetics
1 order of ordinary-differential equation versus kinetic order
2 general reaction of first kinetic order
3 general reaction of second kinetic order
4 reaction of second kinetic order, with both reactants at equal concentration
5 reaction of pseudo-first kinetic order
6 general reaction of third kinetic order
7 reaction of third kinetic order, with two reactants at equal concentration
8 reaction of third kinetic order, with three reactants at equal concentration
9 general reaction of kinetic order n
10 opposing reactions, both of first kinetic order
11 opposing reactions with forward reaction of first kinetic order, reverse of second kinetic order
12 opposing reactions with both forward and reverse reactions of second kinetic order
13 consecutive irreversible reactions of first kinetic order

I index of chemical applications in part I

In this textbook designed to teach, and for the learning of, mathematics for chemistry, the collected material naturally includes sections or groups thereof, examples and exercises with a chemical theme or application; this section provides an index to those components, classified as thermodynamics and general, chemical kinetics and structure in a broad sense. Items in the succeeding lists with initial letter e pertain to exercises and letter x to examples; items lacking either such letter pertain to sections, or a set thereof.

thermodynamics and general

0.11 solution of equations

kinetics

>

I Mathematical Operations -- Mathematics for Chemistry

with contributions from G. Doggett, G. J. Fee, M. B. Monagan and others

O overview of *Maple*

Maple is primarily a computer programme, or software, with which one undertakes interactive calculations involving mathematical objects, although in its mature state its capabilities far transcend that quintessential aspect. One can work with *Maple* in more than one way, even on a particular computer and under a particular operating system on that machine. For most purposes involving exploration of mathematical topics and general interactive calculations, a *Maple* worksheet provides a convenient display that includes

- input statements and commands and their output in numeric and symbolic form,

- plotting instructions and their output as figures embedded within a worksheet,

- spreadsheets also embedded within a worksheet, and

- associated explanatory text.

For intensive programmed calculations a worksheet is less efficient, or executes less rapidly, than *Maple* used in another form that allows input only as command lines, not by clicking on icons in menu and context bars et cetera.

In a form such as this worksheet, our content is designed to take advantage of a graphic interface for a user: three components --

- an essential kernel that executes mathematical commands,

- libraries of procedures for more or less special purposes, and

- an interface between machine and user

-- combine to constitute an operational version of *Maple*. The libraries, some of which are automatically invoked on input of particular commands whereas others require explicit invocation, and the kernel of *Maple* in a particular release are common to computers of almost all types, but a graphic interface is peculiar to each type to the extent that features and operations might vary among those types. A worksheet is a document that can contain descriptive text, like this paragraph printed in black, execution groups with input in red and output in blue, such as

`> (3*x + 5)*(2*y  + 3);`

$$6\,x\,y + 9\,x + 10\,y + 15$$

and embedded graphics and spreadsheets; such a worksheet has a common appearance across various computers and operating systems, and is entirely transferable between computers, provided that they run *Maple* in the same release and, for *Maple* 9 or afterward, that the computer operates the same interface -- either classic or standard. As a result of the graphic interface, operations are undertaken in various ways, such as with explicitly typewritten commands that make no profound use of such an interface, or with control or command keys depressed concurrently with various other keys, or with moving a cursor onto a pictograph near the top or bottom of a display on a monitor and clicking a mouse -- or equivalent mechanism.   To initiate a desired action, we generally employ the former mechanism, namely with commands and statements typed explicitly, because an explanation of such a procedure from this printed text is easier that way than otherwise. We assume generally a computer on which this worksheet executes to possess a mouse with two or more levers, and express actions accordingly to effect some result; if a particular computer have an alternative peripheral device, one must undertake the corresponding action.

When one has opened a worksheet within a *Maple* session to read this text, one can generally observe at the top of the display on the monitor four horizontal rows of words or pictographs, each of which is called a *bar*. Uppermost, a *title bar* likely indicates at left a version of *Maple* and a name of a file that is open in the present window; clicking at extreme left or right ends of this file is likely to terminate, after confirmation, this *Maple* session. Below this title bar is a menu bar: clicking, with a computer 'mouse' or equivalent device, on any word therein, from `File`, at left, to `Help`, near the centre or extreme right, invokes a menu corresponding to that heading. Three items at top of menu `View` relate to three further bars -- for tool, context and status, which one can thereby make to appear or to disappear according to their status altered on clicking with a mouse.   Menu `File` operates in a more or less standard manner according to the underlying operating system of the computer on which *Maple* is running, whereas other menus relate more specifically to operation of *Maple*. That menu `File` contains an item `Preferences` according to which a user may set conditions of operation of *Maple*, for instance to save automatically the file on which a user is working at the end of a particular specified interval, or to activate or to deactivate Balloon Help. If that `Balloon Help` be activated, on clicking left on that name, subsequent motion, by means of a mouse, of the cursor, in the form of an arrow, onto a particular pictograph, or item of a menu, causes display of a short description to advise about a function of that item. Below the menu bar, a tool bar contains pictographs relating to common operations such as copying, printing, reading an existing worksheet or restarting to clear the internal memory. The fourth row down from the top of the display is a context bar, the content of which depends on the nature of a location of the cursor within the visible portion of a worksheet: if that cursor be

located within text such as these words, the context menu pertains to properties of text such as font, size of letters, justification at right or left or not at all, et cetera, whereas, if the cursor lie on a command executable with *Maple,* or a graph, or a spreadsheet, a separate context bar that contains pertinent pictographs appears automatically. On activating an icon, containing *!!!,* in the context bar that appears when the cursor is on a *Maple* input or output item, one can even execute automatically all *Maple* commands in an entire worksheet. At the bottom right of the display appears a status bar that indicates the cumulative duration of execution involving the computer processor, size of workspace and available memory.

   To effect a particular action, mechanisms alternative to typing commands in an input line might exist, such as use of key *Ctrl* or *Alt* in combination with a key for an alphabetical or numeric character; through consultation of `Help` invoked in a menu at the top of the display, or of printed manuals, or of experienced users of *Maple,* one can become acquainted with these mechanisms. For instance, clicking on `Help` above, or, equivalently, depressing concurrently keys *Alt* and *H* is an alternative to typing ? in an input line to invoke help; depressing function key F1 whilst the cursor is on a *Maple* command, or even a key word within this text -- for *Maple* 11, invokes the Help page for that command or term. In this text we habitually neglect such vital ancillary activities as beginning, suspending and ending a *Maple* session, saving and recalling files, use of a mouse with one, two or three levers or buttons et cetera. Experience with other software on a computer of a particular type is helpful in relation to operation of *Maple* on the same machine, as *Maple* is designed to operate under conventions fairly standard for a machine of that type, but even without such experience one can learn quickly how to undertake mathematical operations by executing commands and statements, as we cursorily sample in section 0.21 of this worksheet. With *Maple* 9.5 or subsequent release, a dictionary of mathematical terms is accessible through menu `Help` or through clicking on a particular hyperlink with a lever on a mouse; explanations and examples of applications of many mathematical terms appear there.

   Once this worksheet is open, so that one can read this text, one proceeds by placing the cursor in a line of input to *Maple* displayed in red lettering -- the best location at which to place the cursor is just to the right of a black symbol **>** displayed at the left margin -- and by then depressing key "Return" or a key marked <--' or possibly "Enter", depending on a computer of a particular type; after execution of that input line, the cursor moves automatically to the left edge of the next line of input, which is similarly executed. We expect that, on encountering a particular portion of this content for the first time, one reads the text in black displayed between lines of input in red; therein we recall mathematical principles and explain how to implement them in practice with this computer programme. According to a convention under which *Maple* is designed, a line of input intended to be executed begins in the classic interface beside a black character **>** and is printed in red letters; if the corresponding output expression, printed in blue, be short, it is centred across the screen or page, otherwise it is printed beginning from the left margin. Unexecutable text such as these words, intended for human not mechanical attention, is printed in black letters. At those locations within text at which we mention *Maple* commands rather than just terms or algebraic quantities, these commands might be printed in **red** letters and a distinctive font, whereas output quoted within the text might be printed in blue letters, but such commands are not there executable.

   Just as for any traditional spoken or written language, vocabulary, punctuation and rules to

construct a command -- syntax -- are associated with a programming language. Unlike a spoken or written sentence, in which lack of grammatical precision or imprecise spelling or pronunciation might not preclude understanding, with computer programmes almost no deviation from rules is allowed in construction of a statement, corresponding to a succession of operators with precise names and punctuation. There is not just one way to achieve a desired outcome: just as in any spoken language, a command can be posed with words in disparate sequences to achieve a given objective. In programming environments such as *Maple*, economical use of commands is a preferred style. In early chapters of this text we generally endeavour to avoid using too succinct collections of operators, so as to preclude a reader puzzling over programming syntax rather than an underlying logical basis to solve a particular problem.

In all languages punctuation is important: in both computer and written languages, commas separate items in a sequence, list or set; all commands to a computer have a particular terminating character analogous to a full stop, period or point at the end of a printed sentence, whether it be an explicit mark or a generally invisible character to signal the end of a line. In *Maple*, a mathematical sentence or statement in only one form -- a *command* or *instruction* -- invokes action of *Maple*, whereas a *comment* is ignored by this processor and appears solely for information of a reader; anything following **#** on an input line is treated as a comment, such as in this example,

> **# This is a comment.**

whereas anything else on an input line or anything before **#** is treated as input and must accordingly obey rules for an error message to be averted.

A semicolon **;** is important!

When *Maple* is invoked ready for use according to a classic interface that presents **>** at the left margin as a prompt, to terminate the specification of any input to *Maple* within that line, one must type either a *semicolon* **;** or a *colon* **:** and depress key "Enter" or (carriage) "Return" before execution can begin; merely depressing a key for "Enter" without presence of a semicolon or colon is insufficient to initiate execution, but generally elicits a Warning about premature end of input. If a semicolon be used, any output appropriate to an input is displayed, in blue; if a colon be used, a command is executed but no output is displayed. Messages to advise of an error in input or a warning about altered meanings of names of operators appear in magenta or blue. In the worksheet mode with the standard interface, such use of a semicolon is superfluous. To have displayed a result of a calculation we generally terminate an input line with **;** and initiate execution of a statement or instruction by depressing key "<---'" or "Enter"; a few commands produce no output even when terminated with a semicolon, whereas invoking help with **?** or **?topic**, in which topic denotes a name of a command of interest, requires no terminating punctuation.

An alternative mode -- mathematical input -- exists for which a question mark **?** serves as a prompt; under these conditions no colon or semicolon is required to terminate a particular command or statement to initiate execution -- a depression of key "Enter" suffices. To enter this mode involves clicking on the icon x at the left of the context bar when the context is a line of input.

Details about use of reserved names and arithmetical operators, distinction between parentheses ( ), brackets [ ] and braces { }, and related matters we introduce summarily in section 0.21 below or as required in section 1.118 and elsewhere in chapter 1 and subsequently; on

proceeding through exemplary illustrations in chapter 0 here below, one achieves a glimpse of both the nature of contemporary symbolic computation, and the use of *Maple* in particular, concurrently with discovering the immense mathematical capabilities of this processor.

>

chapter 0   Exemplary illustrations of use of *Maple*

− **0.0 overview**

 Maple is a mathematical programme for electronic digital computers that contains mathematical knowledge accumulated during 5000 years. Here follow five diverse illustrations to demonstrate how powerful and flexible is *Maple* software applied to mathematical aspects of chemical and physical phenomena. First we solve equations of two kinds: in a direct chemical context we treat six linear equations in a set with seven unknowns to balance a chemical equation, deriving thereby a numerical answer; we then solve an equation involving a cubic formula that arises in a problem of physico-chemical interest, seeking a symbolic answer.   We exhibit *Maple*'s symbolic capability also with an algebraic operation according to differential calculus on a simple expression, which might be a challenging manual task. To demonstrate a graphic capability we display approximately a unit cell of a crystal in mock three-dimensional form. *Maple* is not merely a symbolic and numeric calculator with graphical capabilities but also a repository of much mathematical and scientific knowledge; on 'clicking' to activate an underscored word or phrase, such as *linear equations* above, one views material from a dictionary of definitions of mathematical and statistical terms, whereas *Maple* is a repository also of much information about chemical elements and fundamental physical constants, among other topics. With a spreadsheet in *Maple* one can implement both traditional numeric operations and a novel symbolic capability.

  To proceed throught this chapter, or any succeeding chapter, one simply opens a section of which a number and title appear at the left side of this display, reads the text displayed in black letters and executes the commands displayed in red letters, by depressing key "Enter" or equivalent depending on computer type; the latter generally causes display of output dictated by that command, in blue type for mathematical expressions or as a graph for a plotting command. Sample these sections to acquire a flavour of this powerful software for mathematical applications; explore the menus at the top of this display to discover how to use *Maple* as a text editor and diverse other capabilities.

  Following these five sections designed to whet a user's appetite for symbolic computation and associated capabilities in their advanced form in *Maple*, a further section presents a brief introduction to some essential commands and operations that a reader might find directly useful and applicable on the basis of already learned mathematical knowledge; these commands are likely pertinent beyond the mathematical context of these worksheets concurrent with this study of mathematics, and every topic or command in this summary is explained at an appropriate length in succeeding chapters in part I of this book.

>

− **summary of chapter 0**

  After these few samples of what software for symbolic computation and associated operations

can accomplish, we proceed, in eight succeeding chapters within part I, to introduce systematically both the important mathematical concepts and principles and the corresponding commands and operations in *Maple* that we require to undertake significant chemical applications, such as those for which we deploy mathematical methods in part II. The mere sample of some common commands and operations in section 0.21 that concludes this chapter is intended not to guide a user to undertake confidently those operations but only to demonstrate the possibilities; a user should avoid implementing those, or any other, commands without either consulting menu `Help` on the properties of each particular item and option or progressing through the ensuing text to the appropriate explanation provided therein.

After six decades of intensive development, software for symbolic computation has achieved an astonishing performance and power, matching the enormous development of the hardware on which it operates; though not lacking in minor deficiencies, *Maple* in particular offers incomparable resources to solve problems of a mathematical, or mathematically expressible, nature to enhance the capability of a chemist, or of a scientist and engineer in general, to accomplish practically all technical aspects of his or her professional tasks. Even the content of the next eight chapters is far from a comprehensive survey of the total extent of commands and operators that *Maple* offers, but anybody who has achieved a reasonably firm knowledge of these chapters is likely to find that he or she possesses a capability to develop effective solutions to almost all mathematically expressible problems for which formal, algebraic, numeric and graphical methods might be applicable separately or in combination. A prime objective of the use of software for symbolic computation is to enable a user to concentrate on formulating the mathematical problem, rather than to be concerned with tedious details of its solution according to a well defined formulation: a substantial knowledge of both the mathematical principles and their implementation with software is a requisite for this purpose.

chapter 1   Numbers, symbols and elementary functions

1.0   overview and principles

This chapter serves to introduce many commands and operators that prove invaluable in undertaking numeric or symbolic operations, first in essentially arithmetic, next in algebra, and then working with elementary mathematical functions exponential and logarithmic. When a reader encounters this book in a university context, he or she has undoubtedly encountered already many of these arithmetical and mathematical topics during preceding years at a primary or secondary school; the presence of this material serves here a purpose to enable a reader to understand how symbolic computation functions generally without a hindrance of mostly fresh mathematical content to obscure a distinction between mathematics and its implementation on a computer. In conducting a discussion of even arithmetic at this stage of progress in learning mathematics, we employ, however, a formalism and approach more abstract than would be practicable in a context of an elementary school. Arithmetic is a branch of mathematics concerned with numbers, relations among numbers and observations on numbers and their use to solve problems. After distinguishing between a function and a formula, we introduce methods to define our own functions with an arrow notation, and proceed to explore the elementary functions. Analysis, which originally meant solution backwards, includes all mathematical methods in which

the existence of the quantity sought is first assumed as an <u>unknown</u> <u>variable</u> <u>quantity</u>, and its <u>value</u> derived by means of a mathematical process is regarded as analytic; analysis evolved to encompass symbolic methods that yielded equations, in contrast with a geometric mode of solution. The solution of equations and inequalities is an important aspect of any application of mathematics in science and technology. In what follows we formally define pertinent mathematical concepts.

One distinguishes mathematically between a <u>continuous</u> aspect, such as <u>real numbers</u> and <u>limits</u>, and a <u>discrete</u> aspect, such as <u>natural numbers</u> and in <u>number theory</u>; methods of number theory are applied to yield subtle <u>coding</u> of data and <u>digital</u> <u>information</u>.

set and ring

An important concept for which a precise definition is difficult is a <u>set</u>, which implies a collection, possibly uncountable, of distinct numbers or objects; each set is an entity in its own right, and the identity of a set depends only on its members, not on their order. A <u>subset</u> is a set of which each <u>member</u> is also a member of a larger set. In these definitions we assume a property <u>equality</u> implied with an <u>'equal' sign</u> (Robert Recorde, 1557), =, that is fundamental to arithmetic and mathematics; for logical clarity one must distinguish between <u>equal</u> that implies an <u>identity</u>, or <u>equivalence</u>, of quantities on either side of that sign, and an <u>assignment</u> according to which a <u>quantity</u>, generally to the right side of that sign, becomes known to the author and to the computer processor by a name, generally to the left of that sign. That distinction between equality and assignment is crucial in the operation of this software *Maple*.

A <u>set</u>, commonly displayed as a list between braces $\{...\}$, is completely defined according to a rule that determines whether a particular object is a member; an <u>empty</u> or null set exists, denoted $\{\}$ or Φ. A <u>finite</u> set can not be put into <u>correspondence</u> <u>one to one</u> with a <u>proper</u> <u>subset</u> of itself. A <u>linearly ordered</u> set S has a <u>relation</u> R such that the <u>domain</u> and <u>range</u> of R are contained in S, and S satisfies the conditions that

- for any x, y in S, either $x < y$ or $x = y$ or $y < x$, in which $<$ implies 'precedes',

- for $x < y$ and $y < z$, then $x < z$,

- there exists a first member α such that $\alpha < x$ for any other member x of the set, and

- there exists a last member β such that $x < \beta$ for any other member x of the set.

A set is finite if, no matter which linear ordering is applied, each non-empty subset has both a first member and a last member. Any set of positive integers has a first member.

A set **R** becomes a <u>ring</u> if, for three members a, b and c therein,

- an associated <u>sum</u> $a + b$ and product $a\, b$ are also within that ring,

- $a + b = b + a$,   <u>commutative</u> property,

- $(a + b) + c = a + (b + c)$,   <u>distributive</u> property,

- for a member 0 of that ring, $x + 0 = 0 + x = x$ for all x in **R**, so there is a unique real number 0 or <u>zero</u> or <u>nought</u> that is the <u>additive identity</u>,

- for each member a in **R**, there exists another member $-a$ such that $a + (-a) = 0$, so for each real number there is a real number $-a$ called the <u>additive inverse</u>,

- for a member 1 of that ring, product $1\, x$ = product $x\, 1 = x$ for all x in **R**, so there is a unique real number 1 or unity that is the multiplicative identity,

- $(a\, b)\, c\; =\; a\, (b\, c),$ associative property,

- $a\, (b + c) = a\, b + a\, c,$ distributive property.

  The following two conditions are not necessarily properties of a ring:

- $a\, b = b\, a,$ commutative property,

- for each member a in **R** except $a = 0$, there exists another member $\dfrac{1}{a}$ such $a\,\dfrac{1}{a} = 1$, so $\dfrac{1}{a}$ serves as the multiplicative inverse or reciprocal of a.

Here a quantity $a\, b$ implies a product as a result of an operation multiplication of two multiplicands or multipliers for which no operator such as * between the two factors appears explicitly, merely a space between the two quantities to distinguish from a compond symbol ab; such usage constitutes implicit multiplication; in text we might use x or * to indicate explicit multiplication where necessary, but in *Maple* input we must use * (or **.** in special circumstances) for this purpose.

numbers

One classifies numbers according to the following sets. A natural number, denoted **N**, includes zero and each counting number 1, 2, 3, 4 ... in a set; it is a cardinal number that need take no account of its position in any sequence, for contrast with an ordinal number -- first, second, third ... -- that defines a position in such a sequence: for instance, the third item in a sequence 5, 10, 15, 20 ... is cardinal number 15. An integer, denoted **Z**, extends natural numbers to include negative numbers: counting numbers and natural numbers each constitute a subset of integers. Integers are both discrete and countable, but denumerable. A natural number greater than 9 is represented using arabic numerals that constitute natural numbers 0, 1, 2 ... up to 9; such a number can be written as a sum of numerals each multiplied by ten as a base to an integer power as an exponent, such as

$$1234 = 1*10^3 + 2*10^2 + 3*10^{\,1} + 4*10^{\,0}.$$

In that displayed relation appears = , known as an equality operator or equals sign, which implies that the quantities on either side evaluate to identical values; operator * between a number and 10 to some power constitutes explicit multiplication. If two quantities x and y are not identical or equivalent, so that $x \neq y$, one quantity x might be less than, $x < y$, or greater than, $x > y$, the other quantity y; we thus establish the meaning of four operators, namely =, $\neq$, < and >.

Thus zero

- is a cardinal number,

- is an essential member of any system of numbers beyond counting numbers,

- is a place holder, in place-value notation, to distinguish magnitudes with the same significant digits, as in 37, 370, 3700,

- is an identity element for operation addition, such that $a + 0 = 0 + a = a,$

- has a [multiplicative](#) property such that $a * 0 = 0 * a = 0$,

- is an invalid divisor, such that $\dfrac{x}{0}$ is undefined for arbitrary x, and

- operates as an [exponent](#) to yield [unity](#), such that $x^0 = 1$ for arbitrary x, as implemented in *Maple* even though there are exceptional conditions in which this result be invalid.

For comparison, [unity](#)

- is also a cardinal number,

- is the digit other than zero in a binary system of numbers,

- is an identity element for operation multiplication, such that $a * 1 = 1 * a = a$,

- is a valid divisor such that $\dfrac{x}{1} = x$ for arbitrary x, which is a trivial result, and

- operates as an exponent likewise to yield the base as a trivial result, $x^{'1'} = x$ for arbitrary x.

natural numbers and prime numbers

For a property $c = a\, b$ with a, b, c integer, a and b become [divisors](#) of c, with these properties:

- if an integer be divisible by 2, it is called [even](#), otherwise [odd](#);

- a natural number n is [divisible](#) by 2 if it be even;

- a natural number n is divisible by 3 if the sum of its digits be divisible by 3;

- a natural number n is divisible by 4 if its rightmost two digits be divisible by 4;

- a natural number n is divisible by 5 if its rightmost digit be 0 or 5;

- a natural number n is divisible by 6 if it be even and if the sum of its digits be divisible by 3;

- a natural number n is divisible by 9 if the sum of its digits be divisible by 9;

- a natural number n is divisible by 10 if its rightmost digit be 0;

- a natural number n with $n \geq 2$ is [prime](#) if it be divisible by only 1 and itself.

According to [Euclid's](#) theorem, the prime numbers are [uncountable](#), or [infinitely](#) many, and, according to the [fundamental theorem of arithmetic](#), each natural number n with $n \geq 2$ is a product of prime numbers; this decomposition is unique if the prime numbers be ordered by magnitude. The [least common multiple](#) of two natural numbers a and b is obtained on multiplying all distinct prime numbers decomposed from a and b. The [greatest common divisor](#) of two natural numbers a and b is obtained on forming a product of all prime numbers that occur in both [decompositions](#) of a and b into prime numbers. Two natural numbers are [relatively prime](#) if their greatest common divisor be unity; for instance, a number 7 is relatively prime to 8, 9, 10, 11, 12 and 13, but not to 14. For a real number $x \geq 2$, the prime number function $\pi(x)$ is the number of prime numbers $\leq x$. The fundamental theorem of prime numbers is that, for a large number x, an [asymptotic](#) [equality](#) $\pi($

$x) \sim \dfrac{x}{\ln(x)}$ as $x \to \infty$. Riemann's ζ function of argument s is

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

with s a possibly complex number for which $\Re(s) > 1$.

   Numbers in other than a decimal or denary system, based on 10, are analogously composed as a sum of natural numbers less than a base or radix multiplied by that base to various powers; such systems include binary -- with base 2 and symbols 0 and 1 as intrinsic numerals; octal -- with base 8 and symbols 0, 1, 2, 3, 4, 5, 6, 7; duodecimal -- with base 12 and symbols 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, T, E; and hexadecimal -- with base 16 and symbols 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, A, B, C, D, E, F, or the corresponding minuscules *a, b, c, d, e, f.*

   A rational number or vulgar fraction, denoted **Q**, is a ratio of integers represented with a numerator and a denominator; for a common or proper or simple fraction, such as $\dfrac{3}{4}$, the magnitude of a ratio of numerator and denominator is less than unity or 1, whereas an improper fraction, such as $\dfrac{5}{3}$, can be expressed as a mixed fraction having both an integer and a fractional part, correspondingly $1\dfrac{2}{3}$. As an integer can be formally expressed as a ratio of the same integer and unity, integers can be practically considered a subset of rational numbers. Rational numbers are dense and denumerable.

   A real number, denoted **R**, might contain an embedded decimal point in a form such as 0.1234 or 1.2345 for a decimal fraction; a real number is generally associated with such a decimal fraction but can in practice be considered to include a rational number or an integer as a subset. Real numbers are uncountable and non-denumerable. A real number is rational only if its decimal expansion terminates. For n even, $(x^n)^{\left(\frac{1}{n}\right)} = |x|$, whereas for n odd, $(x^n)^{\left(\frac{1}{n}\right)} = x$, and

$$\left(\left(x^{\left(\frac{1}{n}\right)} \right)^{\left(\frac{1}{m}\right)} \right) = x^{\left(\frac{1}{mn}\right)}.$$

   According to Peano's axioms that number five,

- we define a set **N** of natural numbers to be non-empty and to have a distinguished element unity.

- For each member of that set there exists one and only one successor.

- The member unity is the successor of no other member.

- For each successor there is at most one member of that set.

- In a subset M of set **N** that contains unity, if n is in M, $n + 1$ is also within M: this axiom involves mathematical induction.

According to this foundation of Peano's axioms, one introduces first zero and negative integers to define **Z**, and then ratios of integers to define **Q**.

For two numbers a, b, if $a > b$, $a - b$ is a positive number, and vice versa. For three real numbers a, b, c, with $a > b$, $a \pm c > b \pm c$. If a, b, c are real numbers with $a > b$ and $c > 0$, $c\,a > c\,b$. If $a > b$, a boolean test whether $a - b > 0$ yields a result true, otherwise false or possibly undecidable.

A geometric interpretation of real numbers **R**, which hence contain rational numbers **Q** and integers **Z**, associates these quantities with points along an euclidean line (straight line in euclidean space, hence **R**^{1} implying one dimension). A distance between two points a and b is the absolute value of their difference, $|b - a|$, that is also expressible as a (positive) square root of the square of that difference, $\sqrt{(b - a)^2}$. A map or mapping procedure that converts one quantity into another and that preserves distance is called an isometry; two isometries of **R**^{1} are a translation, such that $x \to k + x$, and a reflexion, such that $x \to k - x$. The general orthogonal group of one dimension, designated GO_1, comprises two isometries $x \to \pm x$ that fix the origin, or position of zero, along the euclidean line.

The properties of real numbers include, beside those -- associative, commutative, indentity, inverse and distributive -- of a ring listed above, also closure whereby sum $a + b$ is unique and product $a\,b$ is unique.

We define an interval as a set **I** of real numbers with two properties:

- the interval contains at least two numbers;

- if two numbers belong to **I**, any number between those two numbers belongs also to **I**.

The set **R** of all real numbers is an interval. For each other interval **I** there exists a number a or there exist two numbers a and b, with $a < b$, such that **I** comprises one among the following eight sets for some number x:

$$
\begin{array}{ll}
x < a, & x \le a, \\
x > a, & x \ge a, \\
a < x < b, & a \le x < b, \\
a < x \le b, & a \le x \le b
\end{array}
$$

in which '<' denotes *less than*, '≤' denotes *less than or equal to*, '>' denotes *greater than*, and '≥' denotes *greater than or equal to*. An interval **I** can hence be defined as a set of real numbers of one type among these nine: all real numbers **R** or the eight displayed sets. Each interval **I** except **R** has at least one endpoint. A closed interval includes the endpoints, if any. An open interval excludes the endpoints, if any. The real numbers **R** thus constitute both an open and a closed interval. Any interval that involves only ≤ is closed, of which there are three among the displayed eight sets. Any interval that involves only < is open, of which there are three among the displayed eight sets. The other displayed intervals are either half closed or half open. A closed interval, which thus includes endpoints, might be denoted $[a, b]$, whereas an open interval might be denoted (a, b); one half-open interval might be analogously denoted $[a, b)$. An interval is bounded if it possesses two endpoints; the latter four intervals among the displayed eight are hence bounded. A condition $x \ge a$ implies an unbounded interval $[a, \infty)$, which is regarded as closed,

whereas a condition $x > a$ implies an unbounded interval (a, ∞), which is regarded as open; ∞ denotes infinity.

arithmetic

   Arithmetic involves application of the theory of numbers in elementary aspects required for mensuration and numerical calculations -- addition, subtraction, multiplication, division, raising to a power and extraction of a root. In a collection (or set) of 1, 2, 3 or, in general, n objects (or elements), the act of determining the number of objects present is called counting. For an empty set, no object is present, and the count yields the number 0. Numbers n thus obtained are called natural numbers, represented as **N**; whether natural numbers include zero is debatable. Natural numbers constitute a subset of integers, represented as **Z**, that thus include positive numbers and negative numbers; a negative number is formed on subtraction of a positive number from zero. The scope of numbers becomes extended to rational numbers, represented as **Q**, which are considered to be formed on division of two integers of which the quotient might not evaluate or simplify to an integer. A rational number thus has the form of a numerator as dividend divided by a denominator as divisor. The numerator and denominator are together called the terms of the rational number or fraction. When the numerator but not the denominator of a rational number is unity, the rational number is a fractional unit; multiplication of such a positive fractional unit by a natural number generates a common fraction if the magnitude of the numerator be less than that of the denominator, otherwise an improper fraction. An integer is invariably expressible as a fraction with unity in the denominator; the fractions and integers constitute together the *rational numbers*. A decimal fraction has a magnitude less than unity and is written in a notation of a decimal place value following a decimal point. A percentage signifies a number of parts per hundred, or per cent ; a permillage signifies a number of parts per thousand, or per mil. A ratio signifies a quotient or proportion of two numbers or magnitudes or expressions; for instance, if a mixture contain by mass one quarter of one compound and three quarters of another compound, the two compounds are present in a ratio 1:3 by mass.

   A ratio of two numbers a and b, expressed as $a : b$, is a fraction $\dfrac{a}{b}$ provided that $b \neq 0$. A proportion expresses an equality of two ratios, in a form $a : b :: c : d$; here a and d are called the extremes, b and c are called the means, and d is called the fourth proportional.   In a proportion $a : b :: b : c$, b is a mean proportional between a and c and c is a third proportional to a and b. In these proportions the middle pair of colons might be replaced by an equals sign, as in $a : b = c : d$, and the laws of proportions follow from its expression $\dfrac{a}{b} = \dfrac{c}{d}$ as equal ratios: $a\,d = b\,c, \dfrac{b}{a} = \dfrac{d}{c}$, $\dfrac{a}{c} = \dfrac{b}{c}, \dfrac{a+b}{b} = \dfrac{c+d}{c}, \dfrac{a-b}{b} = \dfrac{c-d}{c}$ and $\dfrac{a+b}{a-b} = \dfrac{c+d}{c-d}$, providing that no denominator equal zero. A proportion or ratio equation has a form of an equality between two ratios, expressed for instance as $\dfrac{a}{b} = \dfrac{c}{d}$ or $a : b :: c : d$, in which a and d are the outer terms; b and c are the inner terms; a and c are the antecedents, whereas b and d are the consequents. The product $a\,d$ of the outer terms equals the product $b\,c$ of the inner terms. Interchanging the terms of a proposition yields $a :$

$c :: b : d$, $b : d :: a : c$ and $d : c :: b : a$. The derived proportions include $(a + b) : a :: (c + d) : d$ and $(a + b) : (a - b) :: (c + d) : (c - d)$. A continued proportion is expressed $a : b :: b : c$, which is a proportion with equal inner terms in which $c = \dfrac{b^2}{a}$; for the mean proportion in which $a : b :: b : c$, $b = \sqrt{a c}$. For a continuous proportion, $a : b : c : d$ **...** $a_1 : b_1 : c_1 : d_1$ **...**, which implies rearrangement into individual proportiions such as $a : b :: a_1 : b_1$, $a : c :: a_1 : c_1$ et cetera.

If a be proportional to b or a vary directly with b, expressed as $a \; \alpha \; b$, an equality requires an inserted parameter k as factor of proportionality as in $a = k\, b$; conversely, if a be inversely proportional to b or if a vary inversely with b, expressed as $a \; \alpha \; \dfrac{1}{b}$, an equality requires an inserted parameter k of proportionality as in $a = \dfrac{k}{b}$. If a vary jointly with b and c, the equality becomes $a = k\, b\, c$.

In chemistry or physics, a number is typically accompanied by <u>units</u>, such as a length of extent one metre or 1 m. According to the <u>International System of Symbols, Units and Notation</u>, such a unit might have a prefix to generate a unit of a convenient size, such as kilometre, abbreviated km, equivalent to 1000 m, or millimetre, mm, equivalent to $\dfrac{1}{1000}$ m, but only m -- neither km nor mm -- is a SI unit. According to engineering or scientific notation, a number might be expressed as a product of a decimal fraction and ten to some positive or negative power, so 1456.789 as 1.456789 10^3.

complex numbers

A <u>complex number</u> resolves the failure to find a real number x that satisfies a simple quadrative equation $x^2 + 1 = 0$. Between real and complex numbers there are similarities and differences:

- real numbers might be ordered whereas complex numbers can not, in general, be ordered;

- the notion of <u>infinity</u> for a complex number differs from that for a real number;

- the set of all real numbers is a proper subset of the set of complex numbers.

A complex number z is defined as an ordered pair $z = (a, b)$ in which both a and b are real numbers. We express z in terms of a and b as $z = a + i\, b$. An <u>imaginary number</u>, which has no specific symbol, has a form conventionally expressed as $b\, i$, which is a <u>product</u> of real number b with $i = \sqrt{-1}$ as the square root of <u>minus</u> <u>unity</u>. A <u>sum</u> of a <u>real part</u> a and an <u>imaginary part</u> $b\, i$, such as $z = a + b\, i$, in which a and b are real numbers, forms a <u>complex number</u> z, denoted $\underline{\mathbf{C}}$; the real part is denoted $\Re(z) = a$ and the imaginary part is denoted $\Im(z) = b$. If $\Im(z) = 0$, $z = (a, 0)$ reduces to real number a; if $\Re(z) = 0$ and $b = 1$, $(0, 1) = i$, a special number that is called the imaginary unit. As complex numbers are defined as ordered pairs, two such complex numbers $z_1 = (a_1, b_1)$ and $z_2 = (a_2, b_2)$ are equal only if parts real $a_1 = a_2$ and imaginary $b_1 = b_2$ are separately equal, as written. As <u>arithmetical</u> properties of complex numbers, such as $z_1 = a_1 + i\, b_1$ and $z_2 = a_2 + i\, b_2$ obey

- addition according to $z_1 + z_2 = (a_1 + a_2, b_1 + b_2) = (a_1 + a_2) + i\,(b_1 + b_2)$, and

- multiplication according to $z_1\,z_2 = (a_1\,a_2 - b_1\,b_2,\ a_1\,b_2 + a_2\,b_1) = (a_1\,a_2 - b_1\,b_2) + i\,($ $a_1\,b_2 + a_2\,b_1)$.

As <u>algebraic</u> properties of complex numbers, complex numbers conform to these <u>axioms</u>:

- $z_1 + z_2, z_1\,z_2$ are within the set of complex numbers $\underline{\mathbf{C}}$;

- addition is <u>commutative</u>, $z_1 + z_2 = z_2 + z_1$;

- addition is <u>associative</u>, $z_1 + (z_2 + z_3) = (z_1 + z_2) + z_3$;

- multiplication is commutative, $z_1\,z_2 = z_2\,z_1$

- multiplication is associative, $z_1\,(z_2\,z_3) = (z_1\,z_2)\,z_3$, and

- multiplication is <u>distributive</u> over addition, $z_1\,(z_2 + z_3) = z_1\,z_2 + z_1\,z_3$.

For these reasons $z + 0 = z$ and $z\,1 = z$, so that numbers zero and unity retain their identity properties in the <u>field</u> of complex numbers. Hence $z + (-z) = 0$, and $z\,\dfrac{1}{z} = 1$ for $z \ne 0$; $z = a + i\,b$ has an <u>additive inverse</u> $-z = -a - i\,b$. For subtraction of two complex numbers we apply the additive inverse, $z_1 - z_2 = z_1 + (-z_2) = (a_1 - a_2) + i\,(b_1 - b_2)$. For real numbers a, b in any couple, ordering implies $a < b$ or $a > b$ or $a = b$, whereas for complex numbers ordering is practicable only when imaginary parts are all zero.

A complex number in a <u>cartesian</u> form defined as above by replacing a and b to obtain $z = x + i\,y$ with real numbers x and y implies a correspondence one to one between that number and a point in <u>plane</u> xy, also called a <u>complex plane</u> or plane z. In that plane, axis x represents a real number and the pertinent axis becomes the <u>real axis</u>; analogously a point along axis y represents an imaginary number, and axis y become the imaginary axis. A complex number may be regarded also a a <u>vector</u> in the complex plane, cf section group 6.2; a complex number is polar coordinates is explained in section 2.311.

A complex number $z = a + b\,i$ has a <u>conjugate</u>, represented $\bar{z}$, and of form $\bar{z} = a - b\,i$, with these properties:

- $\overline{(\bar{z})} = z$, or a complex conjugate of a complex conjugate regenerates the original quantity,

- $\bar{z} = z$ only if z be a real number,

- for two complex numbers z and w, the complex conjugate of their sum is $\overline{(z + w)} = \bar{z} + \bar{w}$ or of their difference is $\overline{(z - w)} = \bar{z} - \bar{w}$,

- the complex conjugate of their product is $\overline{(z\,w)} = \bar{z}\,\bar{w}$ and of their quotient is $\overline{\left(\dfrac{z}{w}\right)} = \dfrac{\bar{z}}{\bar{w}}$ for $w \ne 0$, and

- for each natural number n the power law is $\overline{z^n} = (\bar{z}^n)$.

In the complex plane, complex conjugate number $\bar{z} = (x, -y) = x - iy$ located as a reflexion of $z = (x, y) = x + iy$ across the real axis. The <u>absolute value</u> or <u>magnitude</u> or <u>modulus</u> of a real number is the value of that number disregarding the sign; the absolute value of a complex number $z = a + bi$ is this non-negative square root of the squares of real and imaginary parts, $|z| = \sqrt{a^2 + b^2}$, which is the length of the vector from its base at the origin of the complex plane.

For a <u>multiplicative inverse</u>, for $z \neq 0$ we apply $\dfrac{1}{z} = \dfrac{\bar{z}}{z\,\bar{z}} = \dfrac{a - ib}{a^2 + b^2}$. For division, we apply the multiplicative inverse of the divisor $z_2 \neq 0$ as $\dfrac{z_1}{z_2} = z_1(\dfrac{1}{z_2}) = \dfrac{a_1 a_2 + b_1 b_2 + i(a_2 b_1 - a_1 b_2)}{a_2^{\,2} + b_2^{\,2}}$. In practice, addition and subtraction are conveniently performed with these cartesian forms, but multiplication and division might be more conveniently performed with polar forms.

A <u>gaussian integer</u> is a complex number of which each real and imaginary part is separately an integer, such as $3 + 4i$; such gaussian integers form an <u>euclidean domain</u>. A gaussian integer z is <u>composite</u> if it be factorizable into a form $z = uv$ in which u and v are both gaussian integers excluding ± 1 and $\pm i$, and <u>prime</u> otherwise; hence 2 is composite because $2 = (1 + i)(1 - i)$, but 3 is prime because no analogous relation holds.

For two complex numbers a and b, the absolute value of a product equals the product of the absolute values of the factors:
$$|ab| = |a||b|,$$
and analogously for a quotient providing that the denominator be not zero,
$$\left|\frac{a}{b}\right| = \frac{|a|}{|b|}$$
For a sum we find
$$|a + b|^2 = (a + b)(\bar{a} + \bar{b}) = a\bar{a} + b\bar{b} + (a\bar{b} + b\bar{a})$$
which we rewrite as
$$|a + b|^2 = |a|^2 + |b|^2 + 2\,\Re(a\bar{b})$$
The difference is accordingly
$$|a - b|^2 = |a|^2 + |b|^2 - 2\,\Re(a\bar{b})$$
from which we obtain a relation
$$|a + b|^2 + |a - b|^2 = 2(|a|^2 + |b|^2).$$
Regarding inequalities, from a definition of the absolute value or modulus, we deduce that
$$-|a| \leq \Re(a) \leq |a|$$
$$-|a| \leq \Im(a) \leq |a|$$
Applying these conditions to a formula for an absolute value of a sum above we obtain
$$|a + b|^2 \leq (|a| + |b|)^2$$
or
$$|a + b| \leq |a| + |b|$$
This relation is called the <u>triangle inequality</u> because of its implication that the length of one side of a triangle is less than or equal to the sum of the lengths of the other two sides. As a special

case, for complex number $z = a + b\,i$ we find
$$\left|a + b\,I\right| \le \left|a\right| + \left|b\right|$$
This relation is extensible to an arbitrary sum as
$$\left|a + b + c + ...\right| \le \left|a\right| + \left|b\right| + \left|c\right| + ...$$
which implies that the maximum value of the absolute value of a sum is the sum of the absolute values of its addends.

Because complex numbers conform to commutative, associative and distributive rules and because additive and multiplicative inverses exist, complex numbers in set $\mathbf{C}$ constitutes a field, of which real numbers in set $\mathbf{R}$ form a subset.

In relation to complex ∞, the following rules apply for all z in $\mathbf{C}$:

- $z + \infty = \infty,$

- $z\,\infty = \infty,$

- $\infty + \infty = \infty,$

- $\infty\,\infty = \infty,$

- $\dfrac{z}{\infty} = 0,$ and

- $\dfrac{z}{0} = \infty.$

Hence product $-1\,\infty = \infty$, but product $0\,\infty$ and quotient $\dfrac{\infty}{\infty}$ are undefined.

numbers of other types

There exist also irrational numbers, such as the archimedean number π or pi that is the ratio of the circumference of a circle to its diameter, the Euler number e that is the base of natural logarithms, and $\sqrt{2}$ that implies the square root of 2. An irrational number, which is represented approximately as a decimal number that lacks a pattern of repeating groups of digits, is not expressible as a ratio of integers of finite magnitude. A number such as π or $\mathbf{e}$ that is not a root of a polynomial equation with rational coefficients is called a transcendental number. A random number is a member of a sequence having a property that no member is predictable from preceding elements or items within that sequence; members of such a sequence can form no progression nor follow a regular or repetitive pattern. Among real numbers of other types are an algebraic irrational number that is inexpressible as a ratio of two integers, such as a square root or cube root of a rational number, and a transcendental number that is not a root of a polynomial equation with rational coefficients. We encounter such numbers in subsequent sections.

Concerning operations with numbers, addition or multiplication of one counting number by another generates a further counting number, whereas division of one counting number by another might generate a rational number; subtraction of one counting number by another might generate a negative integer. Addition or subtraction of a number with zero generates no other number, whereas multiplication or division of a number by unity generates no other number.

A factorial function of integer n is a product of the first n counting numbers and hence yields

only an integer. The numbers $2! + 2, 3! + 3, 4! + 4, ...$ for $n = 2, 3, 4, ...$ are not prime numbers, and for increasing n the sets of non-primes in gaps between primes become increasingly long. Gamma function $\Gamma(n)$ for integer n equals an integer that is a factorial of $n - 1$ and integer $n > 0$, such that $\Gamma(n + 1) = n!$ Restricting the argument of a gamma function $\Gamma(n)$ to a positive integer n thus generates a factorial function, but for a general complex argument this function produces complex real numbers. An important special function, beta function $\beta(p, q)$, is related to the gamma function in that $\beta(p, q) = \dfrac{\Gamma(p)\,\Gamma(q)}{\Gamma(p + q)}$.

sequence, set, list, table and array

   Of quantities available to pertain to prospectively multiple items within a collection, a sequence and a set have mathematical significance, but *Maple* provides also a list for which many applications exist, including a package of specific commands. Consider a sequence $(x_1, x_2, x_3, ... , x_n)$ of real numbers; the set of all such sequences constitutes an n-space, and is denoted $\underline{\mathbf{R}}^n$. The first member of $(x_1, x_2, x_3, ... , x_n)$ is x_1, the second member is x_2, et cetera. $\underline{\mathbf{R}}^2$ denotes a space having two dimensions according to which one can plot a graph in a plane, and $\underline{\mathbf{R}}^3$ denotes analogously a space in three dimensions. Many results and techniques that one might develop for $\underline{\mathbf{R}}^n$ with $n > 3$ become useful mathematical tools, but lack a direct geometrical significance or graphic depiction.

   A sequence of integers that arises in diverse cases in both biological and physical sciences is that attributed to Fibonacci, who is credited with introducing arabic numerals into Europe to replace roman numerals; in such a sequence, each number is a sum of the preceding two: with 0 and 1 as initial values, further members are hence 1, 2, 3, 5, 8, 13 ...

   As chemical instances of a sequence, here are symbols of chemical elements

$$n \;=\; 1, \;\; 2, \;\; 3, \;\; 4, \;\; 5, 6, \;\; 7, \;\; 8, \;\; 9, \;\; 10 \; \text{...}$$
$$\text{elements} = \text{H, \; He, \; Li, \; Be, \; B, \; C, \; N, \; O, \; F, Ne ...}$$

in which the latter line is hence a sequence, and the number of alkyl derivatives of benzene, $C_n H_{2n-6}$, with carbon atoms numbering $n = 6, 7, 8, ...$ [N. J. A. Sloane and S. Plouffe, *Encyclopaedia of Integer Sequences*, Academic Press, San Diego USA, 1995]:

$$n \;=\; 6, \;\; 7, \;\; 8, \;\; 9, \;\; 10, \;\; 11, \;\; 12, \;\; 13, \;\; 14, \; \text{...}$$
$$\text{No.} = 1, \;\; 1, \;\; 4, \;\; 8, \;\; 22, \;\; 51, \;\; 136, \;\; 335, \;\; 871, ...$$

A sequence implies not only the ordering of events, which might be numerical or non-numerical, in a set with respect to time but also the use of an attribute possessed by members of the set for that ordering, such as atomic number for chemical elements. A sequence differs from a series in that a sequence is an ordered set but a series involves a sum of quantities in a set. A sequence is finite or infinite according to the number of elements or terms therein. Term n of arithmetic sequence $a, a + d, a + 2\,d, ...$, with common fixed difference d is $a_n = a + (n - 1)\,d$; the sum of an arithmetic series having n terms in that arithmetic sequence is $n\,\dfrac{a + a_n}{2}$. Term n of a geometric sequence $a, a\,r, a\,r^2, ...$ is $a\,r^{(n-1)}$; the sum of a geometric series having n terms in that sequence

is $\dfrac{a\,(1-r^{n})}{1-r}$ with $r \neq 1$. For an infinite geometric series of the same form and with $|r| < 1$, the sum is $\dfrac{a}{1-r}$.

A union of two sets A and B denotes the set C described with a rule that x be a member of C if x belong to either A or B or both. An intersection of two sets A and B denotes a set C described with a rule that x be a member of C if x belong to both sets A and B. An intersection of two sets is hence the common part of the two sets whereas a union of two sets is formed from a consolidation of two sets into one set. A difference of set A and set B contains the members of set A that are not members of set B, whereas a symmetric difference of two sets A and B contains members of sets A and B that are not members of both sets.

As another term for a compilation of quantities that is not a specifically mathematical entity, a table has elements to which we can refer with an index other than a positive integer. As a specialization of a table for which indices to define an entry or element must be integer or a symbol that evaluates to an integer, an array represents such an extended structure in *Maple* to contain data that can have 0 -- 63 dimensions or indices; its name or symbol can correspondingly bear up to 64 integers as indices or subscripts. As an extension of a concept of an array in *Maple*, a table enables one to work with natural notation, and serves as a basis of not only an array but also a matrix and a vector that have symbolic elements by default; the latter features are implemented in the superseded package linalg. In contrast, a rectangular table, or rtable, is a distinct structure used internally in *Maple* and that serves in turn as a basis of an array, a matrix, and a column or row vector of which each element or component is zero by default, all implemented within package **LinearAlgebra**, introduced in chapter 6. As a list is an inefficient mechanism to treat numerous items as a collection, an array or table is preferable.

algebra

Algebra, from an arabic word meaning reunion, is a study of four basic arithmetical operations -- addition, subtraction, multiplication and division -- typically involving symbolic quantities, and the solution of equations that arise thereby; such an exercise is practicable because the objects upon which these operations act might all be left indefinite. An algebraic expression might comprise numbers of any kind, parameters or constants that have a fixed value in a particular context, and variables or unknown quantities that might assume one value in a set within that context and within a particular domain, separated with arithmetical operators. In algebra the symbols used instead of numbers were originally viewed as numbers not determined, or in a sense that a quantity that a symbol represented was left indefinite, but in modern abstract algebra even the quality of the symbols might be left indeterminate, yielding a genuine theory of operations. For symbols to represent mathematical variables in an abstract algebraic context, one generally employs letters near the end of the alphabet, typically x for independent variable and y for dependent variable, whereas, for parameters that can represent constant or invariant quantities within those formulae, letters near the beginning of the alphabet, such as a, b, c, ... are in common use, following Descartes, but greek or other letters might be alternatively applied for particular purposes. Symbols to denote variables for chemical or physical quantities involve typically the first letter of the name, such as T for temperature, V for volume, ...; such names and their symbols

are subject to conventions adopted by International Unions of Pure and Applied Chemistry, and Physics, or International Organization for Standardization.

Algebra is a language comprising not words but symbols: algebra is a branch of elementary mathematics that generalizes arithmetic in using variables to range over numbers; a symbol can denote an unknown quantity within a mechanism to determine its value through elementary operations of arithmetic. Such an operation involves an operator and its operand or argument: to effect an operation, an operator operates on an operand to produce a result whereby one quantity is converted, or mapped, into another. Such an operator might be simply a plus sign that converts two quantities into their sum, according to arithmetic, or a squaring operator that yields a product of a quantity with itself, or a differential operator with respect to a specified variable that yields a derivative, as explained in chapter 3, or an integration operator with respect to a specified variable that yields an antiderivative, as explained in chapter 4, or even a matrix that acts on a vector so as to yield a rotated result, as explained in chapter 6.

An associative operation for at least addition is one for which parentheses are superfluous, such as

$$(a + b) + c \; = \; a + (b + c) \; = \; a + b + c \,.$$

A binary operation involves two elements or members of a set, or applies to two elements in its domain. More abstract algebras serve for the study of systems such as rings, groups and fields with operations not involving implicitly or explicitly infinite sets: a ring is a set of numbers on which operations addition, subtraction and multiplication can be performed without restriction; if a ring contain two or more, equal or unequal, numbers p and q, $p + q$, $p - q$ and $p\,q$ are also members of that ring. A group is a set that is closed under an associative binary operation, generally called multiplication; a field is a set of numbers subject to two binary operations, such as multiplication and division (except by zero). Real numbers and rational numbers are thus fields, whereas integers constitute a ring. A complex number that is not zero, such as $3 + 2\,i$, comprises two linearly independent parts -- 3 and 2 i -- in a field of real numbers, but not in a complex field. The theory of groups has important applications in chemistry, in classification of symmetries of molecular structures and their deformations from their equilibrium conformations, treated in chapter 10 and elsewhere; we allude to such groups in a mathematical sense at various pertinent points.

A field is a set of elements -- numbers -- having two operations, addition denoted $+$ and multiplication denoted $.$, and an equality operator $=$ to satisfy the following seven postulates and a further qualification about that equality operator;

- closure : for each couple x, y of elements in the set, a sum $x + y$ and a product $x \,.\, y$ are in the set;

- commutation: for each couple x, y of elements in the set, $x + y = y + x$ and $x \,.\, y = y \,.\, x$;

- association: for each triple x, y, z of elements in the set, $x + (y + z) = (x + y) + z$ and $x \,.\, (y \,.\, z) = (x \,.\, y) \,.\, z$;

- additive identity -- zero: there exists an element, 0, in the set such that for each x in the set $x + 0 = 0 + x = x$;

- multiplicative identity -- unity: there exists an element, 1, in the set such that for each x in the set $1 \cdot x = x \cdot 1 = x$;

- distribution: for each triple x, y, z of elements in the set, $(x + y) \cdot z = x \cdot z + y \cdot z$ and $x \cdot (y + z) = x \cdot y + x \cdot z$;

- inverse: for each element x in the set, there exists in the set another element $-x$, such that $x + (-x) = 0$; if x be not zero, there exists in the set an element $\frac{1}{x}$ such that $x \cdot \frac{1}{x} = 1$; the element $-x$ is called an additive inverse or negative of x, and an element $\frac{1}{x}$ is called a multiplicative inverse or reciprocal of x; the elements $-x$ and $\frac{1}{x}$ represent single elements, not the results of a subtraction or division.

- equality operator: $x = y$ implies that $x + z = y + z$ and $x \cdot z = y \cdot z$; for $z \neq 0$, $x \cdot z = y \cdot z$ implies that $x = y$, which constitutes the law of cancellation.

Typical examples of fields include the field of complex numbers, the field of rational numbers and the field of real numbers. Further properties -- reflexive, symmetric and transitive -- of a field and its elements arise in connection with modular arithmetic. Applications of a field are implicit in the solution of an equation.

  Although numbers - real and complex - and scalar symbolic or common algebraic quantities obey the commutative law for multiplication, other quantities that are important in chemical calculations, such as matrices in a context of linear algebra, lack this property; W. R. Hamilton discovered such non-commuting quantities in 1843, and there have been devised other quantities, such as *octonions*, that fail to commute even for addition, but these are unimportant in chemistry.

  The set of all integers fails to constitute a field because 1 is the only non-zero integer that has a multiplicative inverse that is also an integer. These postulates omit either subtraction or division because the existence of an additive inverse implies that adding that additive inverse effects subtraction, and the existence of a multiplicative inverse implies that multiplying that multiplicative inverse effects division.

  The set of all even integers is closed under both addition and multiplication, contains zero, and for each even integer x contains the number $-x$ that is also an even integer. Two properties of a field that are not satisfied are the lack of a multiplicative inverse and the lack of a multiplicative identity, unity; the set of all even integers hence fails to constitute a field.

  Besides numbers, mathematics is concerned with variables each of which might be represented with an appropriate symbol and take any value among those in a given set; the set forms the domain of that variable. A real variable has as domain either all real numbers or a subset thereof. A variable might be continuous in a particular interval or take only discrete values in a particular domain. If a set of a particular domain contain only one value, the corresponding variable is a constant. Common algebraic operations involve expansion and factoring of expressions containing variables and numbers, and simplifying the results, but for such simplification there is no absolute criterion.

A polynomial in one or more variables, which is a common algebraic expression, is a mathematical expression comprising a sum of terms each of which is a product of a constant and one or more variables or indeterminates raised to a non-negative integer power. A polynomial in one independent variable has this form,

$$P(x) = c_0 + c_1 x + c_2 x^2 + \dots + c_{n-1} x^{(n-1)} + c_n x^n$$

and contains an independent variable x to various powers and coefficients c_j with index or subscript j taking integer values from 0 to n; these coefficients c_j are symbols for quantities that take finite numerical values in a particular case; such a polynomial might be considered both a function of that variable, $P(x)$, and an algebraic entity comprising that variable and the set $\{x, c_j\}$ of coefficients separated with appropriate arithmetical operators. The degree of a polynomial is the greatest power of the variable therein, n in the displayed formula above; for small degrees, particular descriptors of polynomials are degree one -- linear, two -- quadratic, three -- cubic, four -- quartic, five -- pentic, et cetera. The domain of this polynomial is a set of all real and finite numbers, so that $P(x)$ yields a finite result for any real and finite value of independent variable x. If evaluating a polynomial $f(x)$ for $x = a$ yields zero as a result, $x - a$ is a factor of that polynomial. A polynomial such as $x^2 - 3$ for which substitution of no integer for x yields a zero result is irreducible over integers. A polynomial of degree greater than first has a regular graph or smooth curve, without discontinuity or cusp; a polynomial of degree zero or unity plots in cartesian coordinates as a straight line. The graph of a polynomial of degree n has at most $n - 1$ turning points. In the immediate vicinity of the abscissal axis, the graph of a polynomial for which $(x - a)^n$ is a factor closely resembles a graph of $\alpha (x - a)^n$. When a polynomial $f(x)$ is divided by $x - r$ for r not a root, the remainder is $f(r)$; if $f(r) = 0$, $x - r$ is a factor of $f(x)$ and r is a root of that polynomial, and conversely. According to the fundamental theorem of algebra, every polynomial equation $P(x) = 0$, with $P(x)$ of degree n and of the above form, has n complex roots, of which some might be multiple; for these roots $r_1, r_2, \dots$ that might be real, $P(x)$ is expressible as a product of n linear factors $P(x) = (x - r_1)(x - r_2)\dots$ If rational number $\dfrac{p}{q}$, with p and q having no common factors other than ± 1, be a root of $P(x) = 0$, with $P(x)$ of the form displayed above, p is a factor of c_0 and q is a factor of c_n. For $P(x)$ a polynomial with all coefficients c_j being real numbers, if, for real numbers a and b, $P(x)\big|_{x=a} = P(a)$ and $P(x)\big|_{x=b} = P(b)$ have opposite signs, the equation $P(x) = 0$ has at least one real root between a and b. Likewise, for polynomial $P(x)$ with all coefficients being real numbers, for each complex root $a + b\,i$ there exists another root its complex conjugate $a - b\,i$ with $i = \sqrt{-1}$; analogously, for $P(x)$ with rational coefficients and p, q, r being rational numbers but $\sqrt{r}$ being irrational, for each root of form $p + q\sqrt{r}$ there exists another root $p - q\sqrt{r}$.

A formula $y = f(x)$ is algebraic if, for all x in its domain, it satisfies an equation of form

$$p_0(x)\,y^n + p_1(x)\,y^{(n-1)} + \dots + p_{n-1}\,y + p_n(x) = 0$$

in which $p_0(x), p_1(x), \dots, p_n(x)$ are polynomials in x and n is a positive integer. A formula that is not algebraic is transcendental, of which exponential, logarithmic and trigonometric formulae are

instances.

According to the remainder theorem, for a constant r and a polynomial $P(x)$, the remainder of the quotient $\dfrac{P(x)}{x-r}$ is $P(r)$. If r be a root of polynomial $P(x)$, $x - r$ is a factor of $P(x)$. According to the rule of signs discovered by Descartes, for a polynomial $P(x) = 0$ with real coefficients and written with descending powers of x in which a sign reversal occurs between consecutive terms, the number of positive roots is either equal to the number of those reversals or is less than that number by an even integer; the number of negative roots is either equal to the number of sign reversals in $P(-x)$ or is less than that number by an even integer. For a polynomial

$$P(x) = x^n + p_1 x^{(n-1)} + p_2 x^{(n-2)} + \ldots + p_{n-1} x + p_n = 0$$ of order n in which the coefficient of the leading term is unity, so that that term is x^n,

- the sum of roots is $-p_1$;

- the sum of all binary products of roots is p_2;

- the sum of all ternary products of roots is $-p_3$; ...

- the product of all roots is $(-1)^n p_n$.

If $x = a$ be substituted into a polynomial $f(x)$, the value $f(a)$ so obtained is the remainder that would result from the quotient of $f(x)$ and $x - a$. Likewise, if a polynomial $f(x)$ of degree n be divided by $(x - a)(x - b)$ with $a \neq b$, the quotient becomes a polynomial of degree $n - 2$ with a remainder of form $c x + d$.

An expression in a single variable and containing an embedded equality operator is either an identity, true for any value of that operator, or a conditional equation, true for only particular values of that variable. The preceding displayed equation for the associative law involving a, b and c is an identity that is true for arbitrary values of those quantities, but a conditional equation,

$$3 x + 7 = 5 x + 1$$

is true for only a particular value $x = 3$; that value hence satisfies that equation that is a linear equation in a single unknown quantity identifiable as the symbol x. Finding such a particular value, or solving an equation for a particular variable, generally signifies isolation of that variable on the left side of an equality and having all other numbers and variables on the right side; a solution of an equation, or of equations in a set, implies values in a unique set that yield a true statement when substituted for unknown quantities in that equation, or equations in that set. Procedures that yield equivalent equations are

- adding or subtracting the same quantity on both sides of an equality operator,

- multiplying or dividing both sides of the equality operator by the same non-zero quantity, and

- simplifying an expression on either side of an equality operator, such as by factoring or expanding.

A symmetric equation has coefficients arranged symmetrically, such as in $a x^3 + b x^2 + b x + a = 0$; such equations might have simpler solutions than for non-symmetric

equations.

Fundamental results of mathematics are expressible not only as equalities, involving operator =, but also inequalities, involving operators $<, >, \leq$ and $\geq$. A value of a variable satisfies an inequality in the same way that it might satisfy an equality, but a solution set for an inequality might be large or define a domain rather than comprising a single discrete value or multiple discrete values. An inequality has these properties:

$$\text{if } a < c, a + b < c + b ;$$

$$\text{for } b \text{ positive, if } a < c, a\,b < c\,b \text{ and } \frac{a}{b} < \frac{c}{b} ;$$

$$\text{if } a < b \textbf{ and } b < c , a < c , \text{ according to the transitive property;}$$

$$\text{if } a > 0 , |b| < a \text{ only if } -a < b < a ;$$

$$\text{if } a > 0, |b| > a \text{ only if } b < -a \textbf{ or } b > a .$$

solution of equations

An equation is a statement of equality between two expressions called members. An equation that is true for only a certain value or values of symbolic quantities therein, such as $3\,x + 5 = 11$ that is true for only $x = 2$, is called a conditional equation; an equation that is true for all permissible values of symbolic quantities therein, such as $2\,(x + 3) = 2\,x + 6$ that is true for any value of x, is an identity. To evaluate numerically or symbolically the value of an unknown quantity that appears in a conditional equation is to solve that equation; that solution satisfies that equation. Among operations that one might apply to solve an equation are adding, subtracting, multiplying or dividing equals to equals to obtain results that are equal, provided that in the latter case there is no division by zero; for instance, to solve $x - 2 = 3$, we add 2 to each side of the equality to produce $x - 2 + 2 = 3 + 2$, so that $x = 5$. Equivalent equations have the same solutions, such as $x - 2 = 3$ and $2\,x = 10$, for which in either case $x = 5$. Operations addition and subtraction of equals to equals invariably yield equivalent equations, but multiplication and taking powers of equals might introduce extraneous or redundant solutions beyond those of the original equation; if an operation decrease the number of solutions, the derived equation is described as defective. Operations division or taking of reciprocals of equals might yield defective equations.

A processor for symbolic computation, such as *Maple*, provides powerful means to solve equations and even inequalities of many kinds and extents of complicatiion, but there remains a task of the user of such computer algebra to express the problem to be solved, which is likely stated in more or less formal language, in algebraic terms that are susceptible to formal mathematical solution; such a task might not be trivial, requiring clear thinking and converting words into meaningful symbols and their coefficients. For multiple equations of linear type in their systems, a powerful formalism has been developed in a form called linear algebra involving matrix, vector and other quantities, as we describe in chapter 6; for non-linear systems of one or multiple equations, a processor for symbolic computation might still enable an attack upon a problem by both algebraic and numeric means, although, as the degree, or extent of complication, of the system increases, the prospect of finding an exact algebraic solution fades rapidly. Under those conditions, numeric methods, generally approximate, remain, but they are persistently sensitive to numerical error. It is incumbent upon a student of mathematics to develop a talent for converting problems expressed in ordinary words into algebraic conditions that are susceptible to

solution by whatever methods and facilities be available.

A rational integral equation is a statement of equality between two rational integral expressions or polynomials, each of which contains rational integral terms of form $a\,x^\alpha\,y^\beta\,z^\gamma$ in which a might denote a number of any kind but exponents α, β, γ of unknown quantities x, y, z are non-negative integers. A linear equation or an equation of first degree is expressible in a form $a\,x + b = 0$, in which x is a variable and a and b are parameters that denote real numbers. For two linear simultaneous equations in two variables, the equations have graphs either of two lines that intersect at exactly one point, yielding one consistent solution to the system, or of two lines that are parallel -- and have the same slope -- and fail to intersect, in which inconsistent case there is no solution, or of two lines that coincide, for which consistent case the solutions number uncountably. Alternative to that graphical method to solve two simultaneous linear equations in two unknown quantities are solution by addition or subtraction and solution by substitution. To solve three simultaneous linear equations as a system in three unknown quantities or variables, one unknown might be eliminated between two equations, and then the same unknown from any other couple of equations.

An equation expressible in a form $a\,x^2 + b\,x + c = 0$ in which appear variable x and parameters a, b, c that denote real numbers is quadratic or of second degree. For a pure quadratic equation for which $b = 0$ in the preceding formula so yielding a form $a\,x^2 + c = 0$, the solution is simply one of two roots, $x = +\sqrt{-\dfrac{c}{a}}$ or $-\sqrt{-\dfrac{c}{a}}$. For a general quadratic equation with $a \neq 0$ and $b \neq 0$,

$$y = a\,x^2 + b\,x + c$$

if the roots are not conveniently found by factoring, completing the square yields an expression that one factors to generate the roots; completing the square for that general formula and subsequent algebraic operations yield two general equations that express the roots for $y = 0$ to be

$$x_1 = \frac{-b + \sqrt{b^2 - 4\,a\,c}}{2\,a} \quad \text{and} \quad x_2 = \frac{-b - \sqrt{b^2 - 4\,a\,c}}{2\,a},$$

which *Maple* finds directly; if a, b, c be real numbers and if discriminant $b^2 - 4\,a\,c > 0$, the two roots are real and disparate, whereas if a, b, c be real numbers and if $b^2 - 4\,a\,c < 0$, the two roots are disparate and complex, one being the complex conjugate of the other; for $b^2 - 4\,a\,c = 0$, the roots are real and equal, so amounting to one root repeated. The sum of the two roots is evidently $-\dfrac{b}{a}$ and their product is $\dfrac{c}{a}$. A quadratic formula, such as that named y above, plots as a parabola, of which the real roots occur at intersection of the abscissal axis of that curve. For a radical equation, such as $\sqrt{x - 5} = \sqrt{x}$ or other containing cube roots or other and smaller exponents, in which one or more unknown occurs within a radical, isolating one radical on one side of an equality and raising both sides to clear that radical, and continuing likewise until all radicals are cleared, is a method of solution, but extraneous roots must be rejected by testing of the solutions. A graphical method is effective to find the real roots of two simultaneous quadratic equations in two unknowns. For equations that are symmetric in unknown quantities, such as $3\,x^2 + 3\,y^2 + 2\,x\,y + 5\,x + 5\,y = 7$, a possible solution might be obtained on substituting $x = u + v$

and $y = u - v$. Invoked with a particular command, *Maple* attempts to solve all equations or systems thereof, but a solution might not be practicable in a particular case.

Equations involving polynomials up to quartic have <u>roots</u> expressible in exact algebraic form, but not in general for pentic polynomials or beyond. Even for a general cubic equation, the formulae for the roots are excessively complicated for most purposes, but with unit coefficients *Maple* provides solutions for polynomials of arbitrarily great order. For a general cubic equation $a x^3 + b x^2 + c x + d = 0$, solutions occur in one of three sets depending on the value of coefficients a, b, c, d: three real and distinct roots, three real roots of which two are equal, and one real and two complex roots, of which one of the latter is the complex conjugate of the latter. A cubic equation in reduced form $y^3 + p y + q = 0$ with $a > 0$ has a real root

$$y = \left(-\frac{b}{2} + \sqrt{\frac{a^3}{27} + \frac{b^2}{4}} \right)^{\left(\frac{1}{3}\right)} + \left(-\frac{b}{2} - \sqrt{\frac{a^3}{27} + \frac{b^2}{4}} \right)^{\left(\frac{1}{3}\right)}.$$

For that equation in reduced form, generated on substitution $x = y - \frac{a}{3}$, for three real roots $p < 0$ and the <u>discriminant</u> $4 p^3 + 27 q^2 < 0$;

$$y = q + \frac{2 p \sqrt{-\frac{p}{3}}}{3} \quad \text{or} \quad q - \frac{2 p \sqrt{-\frac{p}{3}}}{3}.$$

For three real roots of which two are identical, that discriminant $4 p^3 + 27 q^2 = 0$. In the other case $4 p^3 + 27 q^2 > 0$, there are one real and two complex roots, the latter as a conjugate pair.

For a quartic equation $x^4 + a x^3 + b x^2 + c x + d = 0$, substitution $x = y - \frac{a}{4}$ yields a reduced form. A special biquadratic equation $a x^4 + b x^2 + c = 0$ that is a special case of a quartic equation is directly soluble through a substitution $x^2 = y$, to become a quadratic equation with roots y_1, y_2 to yield the ultimate solutions $\pm y_1$ and $\pm y_2$.

The solution of a general biquadratic equation,

$$x^4 + p x^2 + q x + r = 0,$$

depends on the behaviour of solutions of the <u>cubic resolvent equation</u>,

$$y^3 + 2 p y^2 + (p^2 - 4 r) y - q^2 = 0$$

that has roots a, b, c; for real coefficients p, q, r and for all $a, b, c > 0$, the original equation has four real roots, whereas for $a > 0$ and $b, c < 0$ there are two pairs of complex conjugate roots; for a real and b, c mutually complex conjugate, there are two real and two complex conjugate roots. The roots of the original biquadratic equation are

$$2 x_1 = u + v + w, \qquad 2 x_2 = u - v + w, \quad 2 x_3 = -u + v + w, \quad 2 x_4 = -u - v - w$$

in which u, v, w are solutions of equations $u^2 = a, v^2 = b, w^2 = c$ with a requirement that $u v w = q$.

If $a, b, c, \acute{} .. \acute{}$ be roots of an equation

$$p_n x^n + p_{n-1} x^{(n-1)} + p_{n-2} x^{(n-2)} + ... + p_1 x + p_0 = 0,$$

the sum of the roots is $-\dfrac{p_{n-1}}{p_n}$, the sum of products of roots taken two at a time is $\dfrac{p_{n-2}}{p_n}$, the sum

of products of roots taken three at a time is $-\dfrac{p_{n-3}}{p_0}$, ... and the sum of products of roots taken n at

a time is $\dfrac{(-1)^n p_0}{p_n}$.

A *reciprocal equation* $f(x) = 0$ of polynomial form is unaltered when the variable is replaced by its reciprocal. For instance, for $a x^2 + b x + c = 0$ to be a reciprocal equation, $a = c$. If $x = r$ be

a root of such an equation, $x = \dfrac{1}{r}$ must also be a root of this equation and the roots must occur in

pairs. If the degree of $f(x) = 0$ be odd, one root must be its own reciprocal, i.e. $x = 1$ or $x = -1$.

These equations,

$$\frac{x^2 + 3x}{x - 2} = \frac{x + 3}{x - 2}, \qquad x^2 + 3x = x + 3$$

are not <u>equivalent</u> because they do not possess the same roots; the proof of this statement requires that no cancellation of denominators be undertaken before solution. The latter equation is called an <u>auxiliary equation</u> because it is useful in determining the solutions of the former. To solve this equation,

$$3x - \sqrt{2x - 3} - 5 = 0$$

by hand, we might isolate the surd to one side,

$$3x - 5 = \sqrt{2x - 3}$$

square both sides, and solve the resulting expression to obtain as roots $x_1 = 2$, $x_2 = \dfrac{14}{9}$; if we

substitute both latter roots separately into the original expression we find that only $x_1 = 2$ satisfies

that equation, $3x - \sqrt{2x - 3} - 5 = 0$; the other value $x_2 = \dfrac{14}{9}$ is called <u>extraneous</u>. Such

multiplication or division of both sides of an equality operator by an expression that involves the variable might thereby introduce extraneous solutions that fail to satisfy the original equation; one should therefore verify all prospective solutions of equations after such operations.

One application of solution of an equation is to find the inverse of a formula or expression. If y be the name assigned to a formula or expression involving an independent variable x, which we

represent as $y = f(x)$, the inverse of that relation is expressed as $x = f^{(-1)}(y)$, distinct from a

reciprocal, $\dfrac{1}{f}$. Although, for a particular formula or expression $f(x)$, an inverse formula might not

be expressible in an explicit algebraic form, one can generally solve numerically and tabulate the results for a domain of x of interest or plot the expression or formula and read coordinates from the graph. In some cases, either variable is inseparable to one side of an equality; such formulae or expressions are expressed $f(x, y)$ and are described as <u>implicit</u>.

An <u>inequality</u> expresses a condition that one expression is greater than or less than another

expression. If $a < b$, the difference $b - a$ is a positive number; if $a > b$, the difference $b - a$ is a negative number. An absolute inequality is true for all values of the quantities involved; for instance, $-1 < x^2$ for all real x. A conditional equality is true for a particular domain of a quantity; for instance $x + 3 > 5$ holds for $x > 2$. Inequalities $a < b$ and $c < d$ have the same sense, whereas inequalities $a < b$ and $c > d$ have the reverse sense. An inequality is invariant if each side be increased or decreased by the same quantity. The sense of an inequality is invariant if each side be multiplied or divided by the same positive quantity, but the sense of that inequality is reversed if each side be multiplied or divided by the same negative quantity. For a, b, n positive quantities and $a < b$, $a^n < b^n$ but $a^{(-n)} > b^{(-n)}$. If $a < b$ and $c < d$, $a + c < b + d$; if $0 < b < a$ and $0 < d < c$, $b\, d < a\, c$. For inequalities,

- for $a < b$, $a + c < b + c$ and $a - c < b - c$;

- for $a < b$ and c positive, $a\, c < b\, c$ and $\dfrac{a}{c} < \dfrac{b}{c}$;

- for $a < b$ and c negative, $b\, c < a\, c$ and $\dfrac{b}{c} < \dfrac{a}{c}$;

- for $a < b$ and $b < c$, $a < c$;

- for $a > 0$ and $|x| < a$, $-a < x < a$;

- for $a > 0$ and $|x| > a$, $x < -a$ or $x > a$.

Maple attempts to solve inequalities with the same command to solve equalities.

For reasons of algebraic or mathematical simplicity, linear relations are most tractable; linear algebra constitutes an immense separate branch of mathematics, having important chemical applications, constructed on such relations, which we treat in chapter 6. After linear relations, naturally quadratic relations are next most tractable, and many chemical and physical models are based on such a relation. Other common relations involve exponential functions and, their inverse, logarithmic functions, treated in section groups 1.4 and 1.5 respectively, with polynomial and other functions that we treat in section group 2..4.

A general, but typically inefficient, method of finding a root x or zero of $f(x) = 0$ is called bisection, or a binary search method that requires $f(x)$ to assume values with opposite signs for values of x that define an interval within which a root must occur. The search operates by finding the sign of $f(x)$ at the middle of the interval and then choosing the subinterval for which the sign change persists; this subinterval is then bisected and the process is repeated until the subinterval containing the root is as small as desired to express the root to digits of satisfactory number. Although convergence is slow in that the number of bisections might be large to obtain a sufficiently small subinterval, the method invariably converges to a solution; for this reason this method might serve as an initial operation for a more efficient method, such as Newton's method, discussed in section 3.308, that involves derivatives of the formula or function, or the secant method or the regula falsi as the rule of false position. One or other method among these might be automatically invoked with *Maple* operator fsolve.

mathematical proof

   A conjecture is an idea that is subject to proof. Some mathematical conjectures have been published for centuries without a proof being demonstrated, even though no counterexample be known. For instance, Goldbach formed a conjecture that each even integer greater than four is expressible as a sum of two odd prime numbers; as examples, $8 = 5 + 3$, $10 = 7 + 3$ ... Whereas some even integers might be such a sum in many ways, for example 150 in twelve ways, others occur as few such sums, for example 98 and 128 in only three ways. For this conjecture there is no known counterexample, but no proof has been discovered to convert this idea into a theorem. For an extension by Goldbach that each odd number is a sum of three prime numbers, Vinogradov formed a proof that enables the latter idea to become a theorem.

   Mathematical induction provides a method to prove a general theorem or formula from particular cases. Such a proof has two steps -- first to demonstrate by substitution that the theorem or formula is true for some single positive integral value of n, such as $n = 1$ or $n = 2$, and then according to an assumption of that truth for $n = k$ to prove its truth for $n = k + 1$. For instance, to prove, for all positive integer values n, a sum from unity, $1 + 2 + 3 + ... + n = \dfrac{n\,(n+1)}{2}$, we test that this relation holds for $n = 2$, for which $\dfrac{2\,(2+1)}{2} = 1 + 2$; assuming this formula for $n = k$, the sum up to $n = k + 1$ is $\dfrac{k\,(k+1)}{2} + k + 1$, which equals $\dfrac{(k+1)\,(k+2)}{2}$ that is the value of $\dfrac{n\,(n+1)}{2}$ when $n = k + 1$.

formula and function

   A formula is an equation that expresses a general fact, rule or principle, such as $C = 2\,\pi\,r$ that states the circumference of a circle to equal twice the product of π and the radius. In undertaking calculations, chemists employ many mathematical formulae and functions, some of an elementary and general nature and others more closely related to chemical applications. We review in this chapter mostly the former and demonstrate common operations involving them. As a formal definition based on two non-empty sets I and O that might represent input and output, a function from I to O is a rule of correspondence that assigns to one element of O exactly one element of I; that correspondence might imply the use of a formula to relate an element of I to a corresponding element of O, so that an element of I acts as input for the output of O. The domain of a function or formula is the set of all inputs I, and the range is the set of all outputs O. The symbol to represent an element in the domain of a function is the independent variable, and the symbol to represent an element in the range is the dependent variable. A graph of a formula or function in plane xy comprises those points (x, y) such that x is in the domain of that formula and y is in its range; such a graph represents a function y with formula $f(x)$ provided that any line parallel to the ordinate axis intersects the graph at one point at most. A graph in a form of a curve, not a straight line, might exhibit one or more turning points at which the graph alters from rising to falling, or vice versa, thus exhibiting a maximum or a minimum, which might be local or global. The exponential function, in which an independent variable appears as a power of a number -- generally e, is

considered to be the most central in mathematics, whereas the natural logarithmic function finds diverse applications, including the study of prime numbers that are important in cryptography and secure communication. Although trigonometric functions have a geometric basis, their relation to exponential and logarithmic functions and to complex numbers gives them broad significance. A graphic representation of a function has several advantages: for this reason we introduce plots of formulae and functions in their context, but leave their extensive discussion to section group 2.1. Many functions, of which we describe only a few in this chapter, are known to *Maple* as having either defined procedures for simplification or one or more operators for evaluation, differentiation and expansion.

To represent a formula or a related function, we might use words in a verbal description, numbers in a table of values, a graph for visual examination or an explicit algebraic form; multiple representations of a particular formula provide insight into its nature. A function is a relation between two sets that associates a unique element of the second set with each element of the first set; for such a relation expressed as $y = f(x)$, y is the value of function f for argument x, or f operates on its operand in an operation to yield a result y. For a function relating two sets, one set S of arguments and another set T of values, S is the domain of that function and T is the codomain , expressed concisely as $f: S \rightarrow T$ or $f: x \rightarrow y$. For s being a subset of S, $f(s)$ is a set of values of $f(x)$ for x within s, and is called the image of s under function f; the image $f(S)$ of the domain is the range of the function. A mapping of x according to function f generates a corresponding value y. Although these relations can be expressed in terms of symbols, they apply equally well to numbers. For an operation to achieve doubling a magnitude and changing a sign, a set {1, 2, 3, 4, ...} -- enclosed within braces by convention -- is mapped into another set {−2, −4, −6, −8, ...}; in this case the domain of the first set is positive integer, and its codomain is even negative integer.

In a context of a single independent variable, a function enables one to relate an independent variable to a corresponding dependent variable, hence mapping one quantity into another. For instance, according to an expression having the form of an equality involving two real variables x and y,

$$y = x^2 + 1$$

the right side of the equality sign contains a formula $x^2 + 1$, in which a value of a single independent variable x determines a dependent variable y that appears on the left side of the equals sign; this expression might be considered to provide a definition of a name y assigned to that formula as an equation rather than an equation. This relation might also be expressed as $y = f(x)$ in terms of a function f for which a formula $f(x) = x^2 + 1$ produces output on adding unity to the square of input variable x; the formula is valid for any value of x, assumed real, and the output is then invariably a positive number. Values of x allowed as input form a domain; corresponding values of y form a codomain or range; for the particular formula $y = x^2 + 1$, the domain of x is [−∞, ∞] whereas the codomain of y is [1, ∞]. The graph of that function f in plane xy comprises those points (x, y) such that x is the domain of f and $y = f(x)$. For two arithmetic operations from functions f and g, the domain of x for a sum $(f + g)(x)$, difference $(f - g)(x)$ or product $(f\,g)(x)$ is a domain common to both $f(x)$ and $g(x)$, but for a quotient $(f/g)(x)$ the domain excludes points at

which $g(x) = 0$. Two functions f and g are mutually <u>inverse</u> if $f(g(x)) = x$ for each x in the domain of g and if $g(f(x)) = x$ for each x in the domain of f; such mutually inverse functions or their formulae have graphs symmetric about a line of <u>unit</u> <u>slope</u>. For two functions f and g, their combined effect on a particular argument or operand is $(f + g)(x) = f(x) + g(x)$,

$(f - g)(x) = f(x) - g(x)$, $(fg)(x) = f(x) \, g(x)$ and $\left(\dfrac{f}{g}\right)(x) = \dfrac{f(x)}{g(x)}$ providing that $g(x) \neq 0$ in the

latter case; for instance, for the composition $(f \circ g)(x)$ evaluates function f at $g(x)$ with $g(x)$ as the <u>domain</u> of f. For the composition of two functions f and g with composition operator @, $(f@g)(x)$ $= f(g(x))$, for which the domain of $f@g$ comprises those values of x in the domain of $g(x)$ for which $g(x)$ is in the domain of $f(x)$. A function is <u>one to one</u> if any line parallel to the abscissal axis intersects its graph at one point at most; a function f has an inverse if and only if f be one to one.

If the scope of x be extended to include <u>complex numbers</u>, the output is in general <u>complex</u>. In this way a function resembles a machine in having three key attributes -- a name, an input and an output; a name is a label attached to a formula to identify it, and the formula of that name operates on input data to produce output. In common situations in which input and output are numbers, typically with attached units in a chemical context, the formula, prescription or <u>algorithm</u> that describes how output is produced from input is called a function. In practical use of a function, we should be aware of values of its <u>arguments</u> as independent variables that have an appropriate domain and <u>sign</u> for a problem of interest. A function might comprise a single <u>operator</u>, such as the <u>sine</u> function, and multiple such operators and other <u>algebraic</u> operations, such as sine plus <u>cosine</u>. In chemical situations in which we typically encounter multiple independent variables as arguments of a particular function, variables in a formula on the right side of an equation or <u>assignment</u> might be numerous. We adopt this intuitive definition of function as it relates readily to <u>expressions</u>, formulae and equations that we meet in chemistry.

For $y \propto x^n$ in which α signifies <u>proportionality</u>, dependent variable y varies directly with independent variable x as in x^n raised to power n, whereas for $y \propto x^{(-n)}$ dependent variable y varies inversely proportionally to x^n; for $y \propto w\,x$, dependent variable y varies jointly with independent variables w and x, whereas for $y \propto \dfrac{w}{x}$ dependent variable varies directly with w and inversely with x.

A mathematical <u>model</u> describes mathematically a chemical, physical or other phenomenon. Common models have these types:

- <u>linear</u> formula, if the relation between one variable and another can be depicted as a straight line, of form $y = m\,x + b$;

- <u>power</u> formula, if the relation between two variables involves one variable taken to a particular power, of form $y = x^j$;

- <u>polynomial</u> formula, if the relation between an independent variable and its dependent variable involves a finite sum of terms containing the independent variable to disparate powers, of form

$y = \sum a_j x^j$ in which there is no particular rule for the occurrence of a term x^j to have a coefficient $a_j \neq 0$;

- rational formula, if the relation between an independent variable and its dependent variable involves a ratio of polynomials, of form $y = \dfrac{f(x)}{g(x)}$;

- algebraic formula, if the relation between an independent variable and its dependent variable is constructed with algebraic operations on polynomials, of form $y = f(x^a + b)$ in which f might imply a suare root for instance;

- trigonometric formula, if the algebraic function involves a trigonometric operator, of circular or hyperbolic kind, with an independent variable as operand, of form $y = \sin(a x + b)$;

- exponential formula, if the power formula involves a constant with a variable as power, of form $y = a^x$;

- logarithmic formula, if the formula includes a logarithmic operator, of form $y = \log(x)$;

- transcendental formula, for a formula of other than algebraic type, of which a trigonometric, exponential or logarithmic formula is a special case.

Among some functions of special types, a constant function $y = f(x)$, for which $f(x)$ is identically equal to a constant value for all x in the domain of definition $[a, b]$, has an equation $y = $ constant for x in $[a, b]$. A step function defined on an interval $[a_0, a_n]$ comprising sub-intervals or partitions $[a_0, a_1), [a_1, a_2), \dots [a_{n-1}, a_n]$, each of which except the latter is a half-open interval, has associated with each sub-interval a constant c_k; such a step function is thus a succession of constant functions, each on its sub-interval. A function absolute value has a form $f(x) = |x|$; its geometric form comprises a line $y = x$ for $x \geq 0$ and $y = -x$ for $x < 0$. Some properties of $|a|$ are $|a| \geq 0, |a| = |-a|, |a| = a$ for $a \geq 0$ or $-a$ for $a < 0, \pm a \leq |a|, |a| - |b| \leq |a + b| \leq |a| + |b|$, $|a| - |b| \leq |a - b| \leq |a| + |b|, |ab| = |a||b|$ and $\left|\dfrac{a}{b}\right| = \dfrac{|a|}{|b|}$.

An even function $y = f(x)$ is defined for both positive and negative x such that $f(-x) = f(x)$; the geometrical implication is that the graph of this function shows symmetry about the ordinate axis such that the graph for negative x is a reflexion across the ordinate axis of a graph for positive x, such as for $|x|$ or x^2. An odd function $y = f(x)$ is defined for both positive and negative x such that $f(-x) = -f(x)$; the geometrical implication is that the graph of this function is obtained first on reflexion across the ordinate axis and then on reflexion across the abscissal axis. An odd function must satisfy $f(0) = 0$ because $-f(0) = f(0)$. Most functions are neither even nor odd, but the product of two even functions, or of two odd functions, is an even function, whereas the product of an even function and an odd function yields an odd function. A function $f(x)$ bounded on an interval has values therein neither exceeding some value V nor less than some value v for x within that interval; values V and v are called lower bound and upper bound respectively, and might occur only at the end points of that interval. The graph of a convex function has a property that a chord

joining any two points A and B thereon invariably lies *above* the graph of the function contained between those two points. The graph of a <u>concave</u> function has a property that a <u>chord</u> joining any two points A and B thereon invariably lies *below* the graph of the function contained between those two points. A <u>polynomial</u> function of <u>degree</u> k is an algebraic expression of form

$$y = a_k x^k + a_{k-1} x^{(k-1)} + \dots + a_1 x + a_0$$

in which k is a positive integer; this polynomial, such as <u>linear</u> for degree 1, <u>quadratic</u> for degree 2, <u>cubic</u> for degree 3, <u>quartic</u> for degree 4 and <u>quintic</u> for degree 5, is defined for all x. A <u>rational function</u> is expressible as a quotient of two polynomials, such as

$$\frac{f(x)}{g(x)} = \frac{a_0 + a_1 x + a_2 x^2 + a_3 x^3}{b_0 + b_1 x + b_2 x^2} .$$

The domain of a rational function comprises all real numbers except roots of $g(x)$ for which $g(x) = 0$; the intercepts on the abscissal axis occur at points at which $f(x) = 0$. Multiple branches of a graph of such a rational function occur if any real root of $g(x)$ exist; at such a point the curve is discontinuous, and approaches an <u>asymptote</u> from either side of that point.

An <u>algebraic function</u> $y = f(x)$ might be transformed into a polynomial, or <u>multinomial</u>, involving both variables x and y, the greatest powers of which both exceed unity; examples are

a <u>monomial</u> $y = +\sqrt{x}$ for $x \geq 0$ that becomes $f(x, y) = y^2 - x = 0$,

and

$$y = x^3 + 2\sqrt{x} - 3 \text{ that becomes } f(x, y) = y^2 - 2 y x^3 + 6 y + x^6 - 6 x^3 + 9 - 4 x = 0,$$

although not all algebraic functions are expressible in this manner and such a transformation might introduce <u>extraneous roots</u>.

A function is <u>transcendental</u> if it be not algebraic, such as $y = x - \cos(x)$; a <u>transcendental number</u> is a <u>root</u> of a <u>transcendental equation</u>, such as a root of $x - \cos(x) = 0$. The <u>signum function</u> signum(x) has a value $+1$ when the sign of x is positive, -1 when the sign of x is negative, and 0 when $x = 0$; except the case $x = 0$, signum$(x) = \dfrac{x}{|x|}$. A function f is <u>injective</u> if, for $x_1 \neq x_2$, $f(x_1) \neq f(x_2)$; a condition $f(x_1) = f(x_2)$ implies that $x_1 = x_2$. A function f is <u>bijective</u> if $f(x)$ yield only a single and unique result for each value of x; a plot of that bijective function intersects any horizontal line, with equation $y = c$ for arbitrary c, in at most one point.

In a mathematical context, a sequence is a function f defined for only integer values of its argument and having for its range an arbitrary set; a sequence is hence a function of a particular kind of which the domain is the set of counting numbers or positive integers. If members of a sequence (x_j) have values on some interval I such that $v < u_j < V$ for all value of j, the sequence is described as <u>bounded</u>, with <u>lower bound</u> v and <u>upper bound</u> V, whereas if $x_{j+1} > x_j$ for all j the sequence is described as *strictly monotonically increasing*. Other prospective descriptions of a sequence on some interval are <u>bounded above</u>, <u>bounded below</u>, <u>unbound</u>, <u>monotonic</u>, *strictly monotonically decreasing*, <u>oscillating</u> (alternate members have opposite signs) et cetera. These terms might apply to functions other than sequences.

exponential function

Exponential formulae arise in chemistry in forms 2^x, e^x and 10^x in many contexts, such as

- Beer-Lambert law, $I = I_0 \, 10^{(-\varepsilon c l)}$, in exponential form, relating the intensity of radiation incident I_o on, and transmitted I through, an absorbing medium of concentration c and length l of optical path, with absorption coefficient ε;

- Boltzmann factor, $\mathbf{e}^{\left(-\frac{E}{RT}\right)}$, that occurs in partition functions, with energy E_j, gas constant R and temperature T;

- concentration in reacting systems in which a reactant is subject to loss according to first kinetic order, $c(t) = c_0 \, \mathbf{e}^{(-k_1 t)}$, in which appear concentration c_t at time t, initial concentration c_0 at time $t = 0$, and rate coefficient k_1;

- Arrhenius equation (attributed to van't Hoff), $k = A \, \mathbf{e}^{\left(-\frac{E_a}{RT}\right)}$, relating a rate coefficient k at temperature T to a pre-exponential factor A and activation energy E_a, with gas constant R;

- decay of electronic density with distance r from an atomic nucleus, proportional to $\mathbf{e}^{\left(-\frac{r}{a_o}\right)}$, with Bohr radius a_o as scale factor.

For [positive] [real] [number] b and for each positive real number x, quantity b^x as an [exponential formula] is a unique real number. When x be [irrational], we [approximate] b^x as closely as desired on [evaluating] b^r for which r is a [rational number] sufficiently near number x. For real number b, if $b^x = b^y$ **and** $b \neq 1$, $x = y$. The laws of exponents are

- $a^m \, a^n = a^{(m+n)}$,

- $\left(a^m\right)^n = a^{(m\,n)}$, and

- $(a\,b)^n = a^n \, b^n$.

The properties of rational exponents are applicable to irrational exponents.

In 1676 [Newton] introduced this exponential function $\mathbf{e}^x$ that has a property of an infinite series,

$$\mathbf{e}^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

that converges for all real x, and also for all complex x as discovered by [Euler] 75 years later. The quantity $y = \mathbf{e}^x$ is strictly increasing and continuous for all real argument x. The Euler limit

$$\mathbf{e}^x = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

holds for all real numbers; for all complex numbers z, $\mathbf{e}^z \neq 0$, so vanishes nowhere. In a complex domain, $\mathbf{e}^{(x + 2\pi i)} = \mathbf{e}^x$ for all complex numbers, with $i = \sqrt{-1}$. The value of $\mathbf{e}^1 = \mathbf{e}$ as a decimal

fraction is an irrational number that is non-terminating and non-repeating, as exhibited in these first twenty digits, 2.7182818284590452354.

An expression b^x within these equalities of form

$$y = b^x = e^{(x \ln(b))}$$

and that conforms to the laws of exponents above is called an exponential formula; here b is a base that is a positive number and hence supposed to be a constant, and x is an exponent that is an independent variable, for contrast with x^b in which, for variable x, exponent b is supposed to be a constant. The ultimate formula at the right above reduces the general exponential function to a function of **e**. Bases that we commonly encounter are 2, 10 and **e**. The latter, a transcendental, and hence irrational, number that serves as base of natural logarithms, pervades science and mathematics; for this reason an exponential function with base **e** becomes the canonical exponential function. For all exponential formulae b^x independent of base b, the point (0,1) is an intersection with the ordinate axis or an ordinate intercept, but for only $b =$ **e** the slope of the curve at this point equals unity. This natural exponential function is continuous at all points in its domain, which is $[-\infty, \infty]$. Functions of other forms such as trigonometric that have their roots in this exponential function we discuss in chapter 2.

logarithmic function

A logarithm is the power x to which a number b as base of that logarithm must be raised to equal a particular value y. A logarithmic function is thus an inverse of an exponential function; inverting an equation displayed above for a general exponential function yields

$$x = \log_b(y)$$

A natural logarithm, described by Napier and generally denoted ln (from french, *logarithmic **n** aturel*), has base e, a notation supposed to honour Euler, and represented in text as **e**, whereas a common logarithm, described by Briggs and denoted generally denoted log by chemists, has base 10. Their properties are similar, but, for a formula containing a logarithm with a particular argument, the value of this formula depends on the base of the logarithm. With a general log arithm to base b other than 10 we need not be concerned; the logarithms of interest are essentially thus $\log_{10}(x)$ and $\log_e(x)$, of which the latter is commonly expressed as $\ln(x)$. Independent of the value of a base, a or b, within an appropriate range, for real positive x and y and $y \neq 0$, the laws of logarithms are

- $\log_b(b) = 1$

- $\log(1) = 0$

- $\log(x) + \log(y) = \log(x\,y),$

- $\log(x) - \log(y) = \log\!\left(\dfrac{x}{y}\right),$

- $\log(x^n) = n \log(x),$

- $b^{\log_b(x)} = x$, and

- $\log_a(x) = \dfrac{\log_b(x)}{\log_b(a)}$

   For a logarithm of a number expressed as a real number with embedded decimal point, the digits preceding that point constitute the characteristic of that logarithm and the digits following that point constitute the mantissa. For a briggsian logarithm of a number greater than unity, the characteristic is positive and its value is the number of digits preceding the decimal point of the number, minus unity; for a briggsian logarithm of a number less than unity, the characteristic is negative and its value is the number of zero digits following the decimal point of the number, plus unity. An antilogarithm is a number corresponding to a given logarithm.

   Logarithmic and exponential functions are hence interrelated in that one is the inverse of the other, as shown above. In general, if function f be defined for a given domain $d(f)$ and a range of values of $f(x)$ be specified with $r(f)$, and if function g have domain $r(f)$ and range $d(f)$, for any x in $d(f)$ there is an unique value of $f(x)$ for which $g(f(x)) = x$. A function that satisfies these requirements is termed *monomorphic* or *single-valued*, or a function "1:1": each such function has an inverse function, which is also 1:1. For instance, for the canonical exponential function $d(\exp)$ $= [-\infty, \infty]$, and $r(f) = [0, \infty]$, and $d(\ln) = [0, \infty]$ with $r(\ln) = [-\infty, \infty]$. For any $r > 1$, the natural logarithm of r, or $\ln(r)$, is definable such that it becomes the area of a region between lines $x = 1$ and $x = r$ and between the abscissal axis, pertaining to x and on which $y = 0$, and a curve representing a graph $y = \dfrac{1}{x}$.

   To alter the base to b, we use $\log_b(x) = \dfrac{\ln(x)}{\ln(b)}$, or, for a base b of logarithms less than unity, $\log_b(x)$ becomes simplified to $-\dfrac{\ln(x)}{\ln(b)}$. Logarithmic equations are susceptible to extraneous solutions.

   Logarithmic formulae arise in many chemical contexts, such as

- $pH = -\log_{10}(a_{H+})$ in terms of activity of hydrogen ion H^+;

- Debye-Huckel limiting law, $\log(\gamma_{+-}) = -A|z_+ z_-|I$, for ionic strength I in terms of mean activity coefficient γ_{+-}, charges z_+ and z_- on positive and negative ions; coefficient A is a collection of factors that takes into account the temperature and the solvent;

- Clausius-Clapeyron equation, $\ln\left(\dfrac{P_2}{P_1}\right) = \dfrac{\Delta H_{vap}}{R}(\dfrac{1}{T_1} - \dfrac{1}{T_2})$, relating the vapour pressures P_1 at temperature T_1 and P_2 at T_2 to the enthalpy of vapourization ΔH_{vap};

- Nernst equation, $E_{cell} = E^o - \dfrac{R\,T}{z\,F}\ln(K_{eq})$, relating electromotive force or cell potential E_{cell} to the standard potential E^o for that cell, with gas constant R, temperature T, number z of electrons transferred in the cell reaction, Faraday constant F, and equilibrium quotient K_{eq};

- Beer-Lambert law, $\ln\left(\dfrac{I_o}{I}\right) = A = \varepsilon\, c\, l$, relating absorbance A to a ratio of intensity of radiation incident I_o on, and transmitted I, through an absorbing medium of concentration c and length l of optical path, with absorption coefficient ε;

- Boltzmann relation, $S = k \ln(\Omega)$, relating entropy S to number Ω of microstates that correspond to a particular observed thermodynamic macrostate, with Boltzmann constant k.

   With the preceding summary of mathematical principles, we begin our exploration of mathematical topics pertinent to their implementation and to chemical applications with a survey of basic terms -- grammar and syntax -- in the *language* for numeric and symbolic computation that is *Maple*. As this mature processor for symbolic computation has capabilities much more numerous and diverse than a merely numeric computing language, such as Fortran, so there are correspondingly abundant terms and constructs in this language. An effective use of *Maple* in chemical, or other, applications requires an acquaintance with these terms -- not all of them, but a subset likely to be important for envisaged applications. Even though familiarity with a small subset of *Maple*'s commands and operators, such as those in section 0.21, suffices for many common purposes, acquaintance with a larger subset is helpful so that one can retain a notion of what might be accomplished when a necessity arises; invoking H e l p in the menu bar on a particular topic, as described in section 1.01, can then refresh and expand one's knowledge of a less familiar operator that can be accordingly applied as required. The usage of commands and operators becomes meaningful within mathematical contexts that exhibit their properties and limitations. Although such terms and contexts might, at first glance, seem remote from chemistry, subsequent and genuinely chemical applications depend directly upon them. A chemist -- at whatever level -- benefits from becoming at least acquainted with many terms, so to facilitate progress towards solution of relevant problems when an important chemical application arises.

   We explain in section group 1.1 how to perform simple arithmetical operations with *Maple*, and introduce commands to treat numbers in various collections; in section group 1.2 we manipulate algebraic quantities and equations, and we solve equations, prefatory to working with elementary functions in section groups 1.4 and 1.5 and with operations of calculus in subsequent chapters. In section group 1.3 we distinguish between a formula and a function, and introduce a simple method to form a function for a particular application. Section group 1.4 describes the properties of exponential functions, and section group 1.5 logarithmic functions. In so proceeding, we introduce many commands and operators, in contrast with few in some subsequent chapters; a working acquaintance with commands or statements of types described here enables one to undertake significant calculations for chemical applications with advanced mathematical methods but few additional commands.

   \>

1.01 seeking help

   One can elicit information on the use of *Maple* in several ways. Throughout the text of this interactive electronic book, hyperlinks to *Maple*'s help system are activated on placing the cursor over a word printed in green letters and underlined (underscored), and then clicking with the left lever of the mouse; for instance, here is a hyperlink to help itself. Information about a particular

topic is obtained on clicking on `Help` and stating the topic in `Topic search` in that menu above the display, or on depressing concurrently both keys "*Alt*" and "h" for help and then both keys "Alt" and "t"; alternatively, for the classic worksheet, placing the cursor on a term in a line of input and depressing function key **F1** elicits help on that topic. These procedures to obtain information on diverse topics relating to operation of *Maple* in this interactive form are applicable to practically all operators, functions and statements pertinent to almost all topics in *Maple*. Although printed books and manuals are generally available in libraries for consultation, these sources might be inaccurate if they fail to correspond to the version of *Maple* in use. Other sources of help are available through internet from www.mapleapps.com and other sites, at which there are compilations of commonly asked questions; for instance, search for "*Maple* FAQ" at www.google.com or other search engine. Information about the use of *Maple* and examples and helpful procedures for various purposes have been generously provided by Professor R. B. Israel at his site on internet www.math.ubc.ca/~israel/advisor, although information there pertains to superseded releases of *Maple*.

An index of topics on which help is available is invoked, perhaps slowly and in a separate window, on executing this command,

> **help(index);**

and an index of topics within categories **expression**, **function**, **misc**, **module**, **packages**, **procedure** and **statement** appears on summoning a particular index with one or other command as follows. This help page presents a list of Maple operators that are available for use in forming an expression; clicking on any item in that list functions as a hyperlink to the pertinent help page devoted to that item.

> **?index[expression]**
> **?index, expression**

Merely specifying a question mark in a command line opens the page on the Maple help system.

> **?**

A list of statements, of which an assignment is one type, is available, and help on a statement of any type is found likewise on invoking a hyperlink therein.

> **? index, statement**

A list of functions in *Maple*'s standard library, excluding those in packages that must be invoked explicitly, becomes presented in hyperlink form on executing this command.

> **? index, functions**

Examples of *Maple* statements and their use in worksheets accessible within *Maple* are available for many topics. The best way to discover for which topics examples are available in such worksheets is to seek `Help` and `Topic Search` under topic `example` through the menu bar at the top of the display; below `example` itself in a list are specified those topics for which such worksheets are provided.

> **?examples, index**

For *Maple* as release 9.5 or subsequent, the following command initiates entry into *Maple*'s dictionary of mathematical terms.

> **?MathDictionary**

Other helpful information is available in an introduction that in turn contains hyperlinks to a tour -- or even a quick tour -- for new users. Also available under menu `Help` in the menu bar is a

glossary that defines or explains many terms or names associated with these *Maple* worksheets.

    With a single question mark at the beginning of a line, a page of help on a specified topic is opened with all sections expanded, for instance for trigonometric function *sine*. In that help page there are hyperlinks to calling sequence, parameters, description and examples.

> **?sin**

For topics that might be discussed within this book, a search of the **Table of Contents** preceding chapter 0 likely reveals sections in which the prospective information might be located.

>

summary of chapter 1

    In this chapter we learn how, with symbolic computation, to perform arithmetic, in section group 1.1, and algebra, in section group 1.2, to form our own functions in section group 1.3, and to work with exponential functions in section group 1.4 and logarithmic functions in section 1.5. Most operations appearing for illustration here we might perform manually with little effort, although, as expressions become complicated, we appreciate how a symbolic processor such as *Maple* can spare us tedious manipulation that might be generally incidental to chemical significance. More important than particular arithmetical and algebraic operations is an acquisition of at least an inkling of a manner in which we can do arithmetic and mathematics with a computer; although we must conform to the conventions of the design of a particular symbolic processor, a benefit of this subservience is that we thereby become master of a great mathematical capability, applicable to solve problems of chemical, technical or other nature in time to come. When we combine this knowledge of generally primitive operations expounded in chapter 1 with progressively higher mathematical and statistical capabilities developed in seven succeeding chapters in part I, we form a strong basis upon which to attack sophisticated chemical problems in chapters in part II.

    Besides properties and capabilities of various commands and operators introduced in this chapter, one must be aware of a distinctive property of a computer programme for symbolic computation such as *Maple*, namely that a symbol such as *x* or *y* can signify two disparate meanings: a symbol might denote an independent variable that has no value other than itself, and that might accordingly be called atomic (type atomic) or *kernel*, or a symbol might denote a name of a quantity as a dependent variable that has a value in terms of either a number or an expression involving other and atomic variables; such symbols might coexist in general usage at any point in a calculation. Other essential practices that one learns early on acquaintance with *Maple* are a necessity to end each statement with colon **:** or semicolon **;**, a distinction between operators for equality **=** and assignment **:=**, and the need of an explicit operator ***** for multiplication -- implicit multiplication is not practicable with *Maple*. Many errors of a novice user of *Maple* involve precisely these aspects.

chapter 2   Plotting, geometry, trigonometry and functions

2.0 overview and principles

    In this chapter, after we introduce the powerful plotting capabilities of *Maple*, we use graphs to depict the properties of formulae, functions and geometrical objects. We relate trigonometry and

complex numbers, and undertake some complex analysis. Much of the discussion of plotting in section group 2.1 and of procedures in section 2.601 is concerned with the implementation of mathematics with processor *Maple*, but such information is invaluable for the purpose of supporting a profound understanding of mathematical principles and concepts and their implementation for applications in chemistry. Although in chapter 1 there was minimal usage of plots, the reason was not their limited value but rather that their active use involves a substantial complement of commands and operators that become more meaningful when an acquaintance with basic mathematics has been assured. In combination with the arithmetic, algebra and elementary functions expounded expansively in chapter 1, this chapter constitutes a strong foundation for higher mathematics to follow, including calculus, linear algebra, differential and integral equations, and statistical applications, and those topics in turn for the applications in chemistry in part II.

plotting

   A graph is a drawing that exhibits geometrically a relation between quantities in various sets, or between numbers, by means of lines, points and other features plotted with respect to coordinate axes. Graph theory is an entire abstract branch of mathematics concerned with application of planar graphs and their generalizations, points or vertices and line segments that connect vertices, applied in the study of topology and combinatorial analysis and cognate topics, of which we discuss chemical applications in chapter 11. For chemical purposes we here restrict our attention merely to plotting points, curves and objects in two dimensions, or within pseudo-three-dimensional displays, to illustrate the properties of, and the relations among, quantities of interest. Graphs of other forms, such as a bar chart or *pie* chart, have applications in the display of numerical data, for instance for statistical purposes.

   The graph of an equation in two variables is the set of all points of which the coordinates satisfy the equation. An approach to describe a function -- some f(x) as a formula, in terms of a mapping from numbers in one set x, with or without units, as input, to another number y, as output and with or without its corresponding units, provides no simple visualization of a way in which output relates to input; employing graphical representations as plots, we improve our understanding of functions of one or two variables.   For a particular numerical value of x as a single independent variable in some formula f(x), we evaluate numerically f(x) to form an ordered couple, or duple, expressed symbolically as (x, f(x)), signifying values for (abscissa, ordinate), according to terms specifying coordinates attributed to Leibniz. One method to describe the position of a point in a plane is to use as reference two intersecting lines, not necessarily perpendicular, called axes, as a frame of reference for that point; the point of intersection of these reference lines is called the origin. In a vertical plane, one reference line is conventionally horizontal, called the abscissal axis, typically accorded a symbol x in an abstract mathematical or geometric application but designated any appropriate symbol as an independent variable in a chemical or physical context; another line is conventionally vertical, called the ordinate axis, and typically accorded a symbol y corresponding to x, or another appropriate symbol, as a dependent variable; this system of reference is known as a cartesian or rectangular system of coordinates, after Descartes who founded analytic geometry among other mathematical innovations. A point P becomes coordinated to the reference frame by means of two directed line segments, one parallel

to each axis, that intersect at that point; the distance and sense along the abscissal axis corresponds to the value of an abscissa of the point, whereas the distance and sense along the ordinate axis corresponds to the value of its ordinate. We mark on a sheet of traditionally ruled graph paper a point such that the numerical value of x indicates the distance of that point from one reference line and the numerical value of $f(x)$ indicates the distance of that point from another reference line perpendicular to the former reference line; that point hence represents an ordered couple. With further points formed from other ordered couples obtained on incrementing or decrementing x by a constant amount and evaluating the corresponding value of $f(x)$, we see a pattern in which we can join each two adjacent points with a short and smooth curve, or even just a straight line if points are close together. The total curve represents a graph of formula $f(x)$ over a certain range, or partial domain, of x. A point $(x, f(x))$ is on a particular curve if its coordinates satisfy the equation of that curve, which would be $y = f(x)$ explicitly or $F(x, y) = 0$ implicitly. A locus is a path traced by a point that moves according to a given condition; the locus of an equation is a geometric figure of which any point thereon satisfies that equation; the locus of an equation, or inequality, is defined as the totality of all points of which the coordinates satisfy that equation, or inequality, and only those points. Some graphs are recognisable to pertain to a well known geometric feature, such as a straight line or a circle, whereas another graph might not be an entirely smooth unbroken curve but possess a discontinuity. The laborious exercise of constructing by hand such a graph we avoid by utilising *Maple*'s facilities, for which a short command -- or even merely applying a computer mouse -- suffices to generate a meaningful plot. In preliminary instances before this chapter during our survey of use and availability of important functions contained in *Maple's* library, we employ *Maple*'s facility **smartplot** to avoid encumbering prematurely our explanation of functions with details of producing graphical displays; we here proceed to describe powerful commands for plotting in various forms.

To obtain, from a graph of $f(x)$, the graph of $f(x) + a$, we translate a units parallel to the ordinate axis in an increasing value of ordinate; the graph of $f(x) - a$ we translate a units parallel to the ordinate axis in decreasing value of ordinate; the graph of $f(x + a)$ we translate a units parallel to the abscissal axis in a decreasing value of abscissa; the graph of $f(x - a)$ we translate a units parallel to the abscissal axis in an increasing value of abscissa; the graph of $-f(x)$ we reflect at the abscissal axis, and the graph of $f(-x)$ we reflect at the ordinate axis. Combinations of these operations that are not reverse of each other are practicable. A graph in plane xy represents a formula $y = f(x)$ of a function f provided that any vertical line, of form $x = a$ for any number a, intersects that graph in at most one point; some functions have a complicated dependence on x such that multiple intersections occur along a vertical line.

As a geometric figure is a graphic realization of a formula, a curve provides a general geometric representation of a formula $y = f(x)$; for inversion to be unique there must exist a mapping one to one of a value of x to a value of y: either $f(x)$ must be strictly monotonic within its domain of definition or otherwise it must be expressible piecewise as functions in a set, each of which must be strictly monotonic on its appropriately chosen domain. The concept of a parametric representation eliminates the necessity of subdivision of the domain and allows even curves with loops. Instead of considering an explicit functional form f to relate one variable x to

another variable y, or an implicit relation such as $F(x, y) = 0$, we consider both x and y separately as two formulae in terms of an auxiliary or parametric variable such as t, so $x = u(t)$ and $y = v(t)$, with t within a domain $[\alpha, \beta]$; such a representation is convenient for an expression of coordinates of an object undergoing curvilinear motion. In some cases the ranges of dependent variables are more readily determined in parametric form than in an explicit relation. In some relations, for conventional variables independent x and dependent y, the latter would be a multiply valued function of the former, whereas in a parametric form both $x(t)$ and $y(t)$ are single-valued functions of parametric variable t. Translations of a locus are readily implemented when one employs a parametric representation. For a case $y = t^2$ and $x = t$ for t in a domain $[-\infty, \infty]$, the parametric variable can be eliminated between the two formula to yield $y = x^2$, for which the geometric representation is a parabola that lies in the upper half-plane and is symmetric about the ordinate axis with its vertex passing the origin. In other circumstances, one can not eliminate the parametric variable, as expected because a parametric representation is more general than an explicit representation. When x and y be trigonometric functions of a third variable t, and with axes x and y oriented mutually perpendicularly, values of x and y obtained in this manner yield two trigonometric curves at right angles; such figures are called Lissajous figures. A parametric representation of a particular curve is in general not unique: altering the parametric variable yields an alternative algebraic representation that has the same geometric representation.

As an alternative to cartesian coordinates in a plane, polar coordinates, typically expressed as (r, θ), become defined relative to an origin generally taken as a pole from which extends in one direction, typically that corresponding to the positive direction of the abscissal axis in cartesian coordinates, a half-line called the polar axis; the other coordinate is the polar angle through which the polar axis must be rotated, in a counter-clockwise sense, to coincide with a line segment from the origin to a particular point. The coordinates of the pole are $(0, \theta)$ for arbitrary θ; for any other point the coordinates are $(r, \theta + 2\,n\,\pi)$ with integer n, hence of uncountable number. In polar coordinates a curve is defined as the locus of points that satisfy an equation explicitly as $r = f(\theta)$ or its inverse, or implicitly as $F(r, \theta) = 0$. For the translation from cartesian coordinates (x, y) in a plane to polar coordinates (r, θ), the pertinent relations are

$$r = \sqrt{x^2 + y^2}, \quad \cos(\theta) = \frac{x}{\sqrt{x^2 + y^2}}, \quad \sin(\theta) = \frac{y}{\sqrt{x^2 + y^2}},$$

so $\tan(\theta) = \dfrac{y}{x}$ with a prospective ambiguity because of principal values. For some problems, expression in terms of polar coordinates rather than cartesian coordinates might produce a simplification.

For three spatial dimensions, the most common systems of coordinates are cartesian, for which a point is specified as (x, y, z), or spherical polar, for which a point is specified as (r, θ, ϕ). A surface of an object in three dimensions is effectively a two-dimensional object.
geometry.

We treat descriptive geometry, rather than formal axiomatic geometry that is less applicable to such chemical applications as molecular structure. For this purpose we require four concepts, of which the latter three pertain specifically to geometry. As explained and applied in section 1.118,

a set is considered to be a collection, possibly infinite, of distinct numbers or objects, that becomes an entity in its own right, and with an identity dependent upon only its members. A point, which is a basic element in axiomatic geometry, in a cartesian space is an element that is located according to a single n-tuple of coordinates; n, the number of dimensions of a formal space under consideration, is typically two or three for general conditions that allow plotting, but informally a point is a geometrical element having no dimensions. In cartesian geometry a line is a straight geometric figure having extension in only one dimension, known as length -- which is infinite -- but no thickness; a plane is a geometrical figure described as a flat surface, thus having extension in only two dimensions described as length and width -- each of which is infinite so that the plane has no edges -- but no thickness. According to axiomatic geometry these four concepts are undefined terms in a sense that each is assumed rather than provable. A line is assumed also to constitute points in a set; a line segment is a part of a line lying between two of its points, whereas a ray is a half-line extending from a particular point. In a cartesian space of two dimensions, any two distinct points define a line of which the segment between them is the least path; in three or more dimensions the direction of a line is given by its direction cosines. An angle is defined according to the rotation of a segment of a straight line about an axis perpendicular to that line, and forms thus a geometric figure formed between two distinct rays or line segments with a common point or vertex, or by regions of two distinct planes that extend from a common line; the fundamental unit of measurement of an angle is radian. A bisector of an angle divides the interior of that angle into two equal parts. According to an euclidean space, two distinct points uniquely determine a line. Points in any set constitute a geometric figure, of which a line or a line segment is one particular type. Three distinct points not collinear define a triangle, for which those points serve as vertices and segments of lines between each couple of points serve as edges, and analogously for polygons with more numerous vertices; a polygon is a closed plane figure bounded by three or more straight line segments that terminate at the same number of vertices and that intersect at only those vertices. The sum of interior angles of a polygon with n edges is $2(n-2)\pi$ rad, but the sum of exterior angles is 2π rad independent of the number of edges. A regular polygon has all edges of equal length; the apothem of a regular polygon is the distance from the centre to any vertex, and the area of such a polygon is a product of half the length of an apothem and the perimeter. A convex polygon has no interior angle greater than π rad.

Two triangles are congruent if

- two angles and a side have the same values, or

- three sides have the same lengths, or

- two sides and the included angle have the same values.

Many relations between the sides and angles of a triangle involving trigonometric functions appear in the discussion of trigonometry below.

A simplex is a geometrical figure with line segments as edges, or sides, and vertices for which both edges and vertices number one more than the dimensions of a pertinent space; for a space of two dimensions the simplex is hence a triangle, or for a space of three dimensions a tetrahedron, not necessarily regular. Any three distinct points in a plane, not collinear, define a circle. A classic geometric problem was an attempt to square a circle, i.e. to find, using only a straight edge

and compass, a square with sides of length $\sqrt{\pi}$ of area equal to that of a circle of unit radius.

A polygon is characterized with its convexity or type of non-convexity. For a convex polygon, any line drawn through the polygon, and not tangent to an edge or corner, meets its boundary exactly twice; all interior angles are consequently less than 180°, or, equivalently, any line segment with end points on the boundary passes through only interior points between its endpoints. For a non-convex polygon, a line might be found that meets its boundary more than twice; equivalently, there exists a line segment between two boundary points that passes outside the polygon. For a simple polygon, the boundary of the polygon does not cross itself; all convex polygons are simple. A term complex is sometimes used in contrast to simple, but this usage risks confusion with the idea of a complex polygon as one that exists in a complex Hilbert plane comprising two complex dimensions. For a concave polygon that is hence non-convex and not simple, an interior angle exceeds 180°. For a star-shaped polygon, all interior is visible from a single point, without crossing an edge; the polygon must be simple, and might be convex or concave. For a self-intersecting polygon, the boundary of the polygon crosses itself; a term coptic for this condition seems not to be widely used. A star polygon self-intersects in a regular way. Equiangular implies all equal corner angles. Cyclic implies that all corners lie on a single circle, called the circumcircle. Isogonal or vertex-transitive implies that all corners lie within the same symmetry orbit; the polygon is also cyclic and equiangular. Equilateral implies that: all edges have the same length; a polygon with five or more sides can be equilateral without being convex. Tangential implies that all sides are tangent to an inscribed circle. Isotoxal or edge-transitive implies that all sides lie within the same symmetry orbit; the polygon is also equilateral and tangential. Regular implies that the polygon is both cyclic and equilateral; equivalently, it is both equilateral and equiangular. A non-convex regular polygon is called a regular star polygon.

Rectilinear implies that a polygon has sides that meet at right angles, i.e., all interior angles are $\frac{\pi}{2}$ or $\frac{3\pi}{2}$ rad, or equivalently 90° or 270°. If a polygon be monotonic with respect to a given line L, every line orthogonal to L intersects the polygon not more than twice. An isometry of the euclidean plane is a transformation of the plane that preserves distance and that comprises four types -- translation, rotation, reflexion and glide reflexion.

A polyhedron is a closed solid geometrical figure, or its surface, that is bounded by at least four polygons not coplanar, so that pairs of faces meet at an edge and three faces meet at a vertex; such a figure is hollow. Five regular polyhedra -- a tetrahedron with four equilateral triangular faces, a cube with six square faces, an octahedron with eight equilateral triangular faces, a dodecahedron with twelve equilateral pentagonal faces and an icosahedron with twenty equilateral triangular faces -- have equal lengths of edges and angles between faces at vertices. For any polyhedron, regular or irregular, in three spatial dimensions, a general relation between the numbers of faces, edges and vertices is

$$vertices + faces - edges = 2.$$

A prism is a polyhedron with two parallel and congruent faces, called bases, which make all other faces parallelograms; its volume is a product of the area of its base and the perpendicular distance between the planes of the bases. A pyramid is a polyhedron of which one face, considered to be

the base, is a polygon and other lateral faces triangular with a common vertex, called the apex. A cylinder is a special case of a prism with a circular base; a cone is a special case of a pyramid with a circular base. For a hollow right circular cone with an horizontal base, an horizontal slice above the base produces a circle and a vertical slice produces an hyperbola; a slice at an acute angle with the vertical axis produces an ellipse if the slice does not pass the base or a parabola otherwise: these geometrical figures in two dimensions are thus conic sections of which circle and ellipse are closed curves and hyperbola and parabola open. A double cone has hence two nappes that meet at a vertex. A right circular cone constitutes a surface swept with a straight line with one point, the vertex, fixed as the line sweeps around a circular path, thus with two nappes.

For an angle of which the vertex is at the centre of a circle, this central angle intercepts an arc on the circumference of that circle, and the arc subtends that central angle. The ratio of the arc to the radius of the circle is a measure of the extent of the angle subtended by that arc, with unit radian; 1 radian is a measure of a central angle that intercepts an arc of a circle equal to its radius. A sector of a circle is a region in a plane with boundaries comprising two radii and the intercepted arc. Relative to an origin and a horizon as a base, an angle of elevation implies a rotation of a line segment in a counterclockwise direction, whereas an angle of depression implies a corresponding rotation in a clockwise direction.

The location of a point in a plane is related to a system of coordinates, generally cartesian coordinates that imply an ordered couple of real numbers in $\mathbf{R}^2$, and analogously for location of points in three spatial dimensions with an ordered triple of real numbers in $\mathbf{R}^3$. A location within a plane is thus specified by reference to two number lines, called axes, which are at a right angle to one another according to conventional cartesian coordinates; in a vertical plane in two dimensions, one axis generally drawn horizontally serves as abscissal axis, commonly associated with letter x, and another axis corresponding vertically serves as ordinate axis, commonly associated with letter y. For a particular point within that plane, the real number along each axis that defines a location becomes one member of that couple, with the number pertaining to the abscissal axis preceding the number pertaining to the ordinate axis. For a point within three spatial dimensions, the vertical axis is commonly associated with letter z. Although graphic depiction of points in spaces of dimension greater than three is impracticable, the pertinent algebraic operations, or vector operations (cf section group 6.2), are readily extensible to arbitrary dimensions. For two points in plane xy with cartesian coordinates (x_1, y_1) and (x_2, y_2), the distance between the two points has the magnitude

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2},$$

and the midpoint between those two points on a segment of a straight line is

$$(\frac{x_2 + x_1}{2}, \frac{y_2 + y_1}{2});$$

these formula are extensible to multiple dimensions in an obvious manner. The slope of the straight line passing those two points in two spatial dimensions is

$$m = \frac{y_2 - y_1}{x_2 - x_1},$$

which is a ratio of the rise in the numerator and the run in the denominator. The slope of a line

parallel to the abscissal axis is accordingly zero, whereas the slope of a line parallel to the ordinate axis is undefined. The equation of that line passing the two specified points is accordingly

- $\dfrac{y - y_1}{y_2 - y_1} = \dfrac{x - x_1}{x_2 - x_1}$ in a form with two points (x_1, y_1) and (x_2, y_2), or

- $y - y_1 = m\,(x - x_1)$ in a form with slope m and one point (x_1, y_1), or

- $y = m\,(x - a)$ in a form with slope m and abscissal intercept a, or

- $y = m\,x + b$ in a form with slope m and ordinate intercept b, or

- $\dfrac{x}{a} + \dfrac{y}{b} = 1$ in a form with abscissal intercept a and ordinate intercept b.

Two lines are <u>parallel</u> only if their slopes are equal -- $m_1 = m_2$, so the lines lack an <u>intersection</u>, whereas two lines are mutually <u>perpendicular</u> or <u>orthogonal</u> if their slopes conform to $m_1 = -\dfrac{1}{m_2}$.

The <u>inclination</u> of a line is the angle of counter-clockwise rotation in an <u>interval</u> $[0, \pi[$ from the positive abscissal axis to that line, whereas the <u>declination</u> is the corresponding angle of rotation in a clockwise sense.

   <u>Symmetry</u> is an important concept in geometric constructions. Two points are <u>symmetric</u> with respect to a line if that line is a <u>perpendicular</u> <u>bisector</u> of a segment of a straight line joining those points. A graph is symmetric with respect to a line if all points of that graph occur in pairs symmetric with respect to that line. An equation in x and y has its graph symmetric with respect to axis x if that equation is unaffected on replacing y by $-y$ -- such as $y^2 = x + 1$, with respect to axis y if that equation is unaffected on replacing x by $-x$ -- such as $y = x^4 + 2\,x^2 + 1$, and with respect to the origin if that equation is unaffected on replacing both x by $-x$ and y by $-y$, such as $x^2\,(1 - y^2) = 1$; the former operation is equivalent to <u>reflexion</u> at axis y, the next operation analogously reflexion at axis x, and the latter operation <u>inversion</u> at the origin. Analogous arguments apply to symmetry of geometrical constructions in a physical space with three dimensions.

   In evaluating a point (x, y) of which these coordinates satisfy a particular equation, we consider only real values. When a variable appears in an equation to an even power, a solution for that variable might involve a square root (or other even root). A condition that a negative number has no real square root might then limit the extent of a curve. For instance for $\sqrt{x^2 + y^2} = 4$, solution for x yields $\pm\sqrt{4 - y^2}$; for $|\,y\,| > 2$, the quantity under the surd sign has a negative value; the extent of the curve along axis x is thus limited to an interval $[-2, 2]$, and likewise along axis y by symmetry. The points at which a particular curve <u>intersects</u> an axis of a coordinate system is called an <u>intercept</u>; in two dimensions an <u>abscissal</u> axis for an abstract or pure algebraic case typically bears a label x, and the <u>ordinate</u> axis analogously y. In this case, we find an intercept of axis x on setting $y = 0$, and a y intercept analogously on setting $x = 0$, readily using commands **solve** or **fsolve** as discussed in section 1.208; such a value of $(x, 0)$ or $(0, y)$ might be a

<u>stationary</u> or <u>critical point</u> that one locates on plotting with **`algcurves[plot_real_curve]`**, as described in section 2.103. For a vertical plane containing a system of axes, an equation of a horizontal line passing point (a,b) is $y = b$, whereas the equation of a vertical line passing that point is $x = a$.

As a point $P(x, y)$ moves in a plane along a particular curve farther from the origin, the least distance between that point and a fixed straight line might <u>tend to</u> zero; such a line would then become an <u>asymptote</u> of that curve. In a case of a <u>ratio</u> $y = \dfrac{Q(x)}{S(x)}$ of <u>polynomials</u> $Q(x)$ and $S(x)$ that lacks a <u>common factor</u>, if $x = c$ be a <u>root</u> of $S(x)$ in the <u>denominator</u>, as the x coordinate of a tracing point $P(x, y)$ approaches $x = c$, $`(x\text{-}c)` \to 0$ and $y \to \infty$; hence a vertical line $x = c$ becomes an asymptote for that curve, and such a value of x becomes a <u>pole</u> of <u>rational</u> expression y. Such a rational function has then a <u>discontinuity</u> along axis x; a few commands for locating such a discontinuity are introduced in section 3.103. Likewise, a ratio of polynomials in y in a form $x = \dfrac{Q(y)}{S(y)}$ might exhibit horizontal asymptotes, or discontinuities along axis y.

According to <u>euclidean geometry</u>, any point on a line might be selected as the <u>origin</u> of a system of coordinates with coordinate 0, and any other point might be selected as the unit point with coordinate 1. Every point on a line has thus a real number as its coordinate, and every real number has a point as its graph. As an euclidean line involves real numbers, points in a set can be placed in unitary <u>correspondence</u> with real numbers in their set. For a line segment AB, a point thereon that is not an end point, such as A or B, is an <u>interior point</u>. The <u>length</u> of a line segment is a measure of that segment; two line segments that have the same length are <u>congruent</u> line segments. Two line segments that have a common end point form a plane angle with line segments as sides and a common end point as a vertex. If angle α have as its measure 0, no matter whether <u>radian</u> or <u>degree</u>, the two segments are collinear; for $0 < \alpha < \dfrac{\pi}{2}$ rad $= 90\,^{o}$, this angle is

<u>acute</u>; for $\alpha = \dfrac{\pi}{2}$ rad $= 90\,^{o}$, a <u>right angle</u>, and the line segments are <u>perpendicular</u>; for $\dfrac{\pi}{2}$ rad $=$ $90\,^{o} < \alpha < \pi$ rad $= 180\,^{o}$, an <u>obtuse</u> angle; for $\alpha = \pi = 180\,^{o}$, a <u>straight</u> angle, and for π rad $=$ $180\,^{o} < \alpha < 2\,\pi$ rad $= 360\,^{o}$, a <u>reflex</u> angle. A reflex angle is hence the larger of two unequal angles between two distinct line segments that meet at a vertex. Any two angles with equal measure are congruent angles. A general <u>closed</u> geometrical figure with coplanar and non-intersecting edges is a <u>polygon</u>, of which a particular figure with three coplanar sides is a <u>triangle</u>, with four coplanar sides is a <u>quadrilateral</u> or <u>tetragon</u>, with five coplanar sides is a <u>pentagon</u>, with six coplanar sides is a <u>hexagon</u>, et cetera. A <u>rectangle</u> denotes the periphery of a <u>rectangular</u> figure, and analogously for other polygons.   A <u>regular</u> polygon has equal interior angles and equal lengths of adjacent edges. A quadrangle is a plane figure comprising four points each of which is joined by at least two lines to two other points; the line segments between vertices might intersect so that this figure is not a quadrilateral or tetragon, and hence not a polygon; a quadrangle is <u>convex</u> and hence a quadrilateral if both diagonals lie inside, <u>re-entrant</u> if one lie outside and <u>crossed</u> if both lie outside. A <u>salient</u> angle is less than π rad $= 180\,^{o}$, and an

interior angle of a polygon is salient if its vertex point outwards.

A conic section is a curve formed where a plane intersects a right circular cone -- of type circle, ellipse, parabola or hyperbola. A circle constitutes the set of all points in a plane on its circumference or periphery equidistant from a fixed point called the centre; the distance from centre to circumference is the radius. The equation of a circle in cartesian coordinates in standard form is

$$(x - h)^2 + (y - k)^2 = r^2,$$

with the centre at (h, k) and radius r; the diameter $d = 2\,r$, the length of the circumference is $\pi\,d$ and the area of the planar surface within the circle is $\dfrac{\pi\,d^2}{4}$ or $\pi\,r^2$. A sector of a circle is a region bounded by two radii and the intercepted arc; the area of this sector is $\dfrac{1}{2}\,r^2\,\theta$, in which angle θ is measured in radians. A torus is a geometric figure in three dimensional space that, in its most symmetric form consists of a circular tube, radius R, with a circular cross section, r; the area of its surface is $A = 4\,\pi^2\,r\,R$ and its enclosed volume is $V = 2\,\pi^2\,r\,R$.

An ellipse constitutes the set of all points on its periphery or circumference of which the sum of distances from two fixed points, called foci, is constant; the two foci define a line called the focal or major axis, and the centre of the ellipse is midway between these foci; the vertices of an ellipse lie at the intersections of this axis with the periphery. With a system of cartesian coordinates, the equation of an ellipse in standard form is

$$\frac{(x - h)^2}{a^2} + \frac{(y - k)^2}{b^2} = 1;$$

with the centre of the ellipse located at the origin such that $h = 0$, $k = 0$, the foci are located at $(\pm c ,0)$ with $c^2 = a^2 - b^2$, and the vertices are located at $(\pm a,0)$. The length of the major axis between these vertices is $2\,a$, the length of the minor axis perpendicular to the major axis is $2\,b$, and the ellipse is symmetric to reflexion across both these axes; the eccentricity of the ellipse, which is here a measure of its deviation from a circular shape, is

$$e = \frac{c}{a} = \frac{\sqrt{a^2 - b^2}}{a}.$$

For semi-major axis a and semi-minor axis b, parametric equations to define an ellipse are $x = a\,\sin(\theta)$, $y = b\,\cos(\theta)$, or equivalently $x = a\,\cos(\theta)$, $y = b\,\sin(\theta)$ for $0 < \theta < 2\,\pi$.

A parabola constitutes the set of all points in a plane equidistant from a fixed line, the directrix, and a fixed point, the focus, not on that line, and this parabola is symmetric to reflexion across its axis that is perpendicular to the directrix and contains the focus; the vertex of a parabola occurs at the intersection of the curve with that axis. A canonical equation of a parabola is

$$y^2 = 4\,a\,x,$$

which yields a curve symmetric about axis x with its vertex at the origin, its focus at $(a,0)$ and a distance $2\,a$ between focus and directrix. An equivalent relation $y = 4\,a\,x^2$ is symmetric about axis y; for an equation $y = a\,x^2 + b\,x + c$ completing the square enables rewriting in a form $y = a\,(x - h)^2 + k$, for which an axis of symmetry is $x = h$; the parabola opens upward if $a > 0$ or

downward if $a < 0$.

An hyperbola constitutes the set of all points in a plane with a constant difference of length between two fixed points, the foci; an hyperbola comprises two branches asymptotic to two intersecting fixed lines. A standard equation to describe an hyperbola is

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1,$$

for which the transverse axis coincides with axis x and the conjugate axis coincides with axis y. The distance between the vertices on the transverse axis, the length of the transverse axis, is $2\,a$, the length of the conjugate axis is $2\,b$, the centre is midway between these two vertices, and the asymptotes are $y = \pm \frac{a}{b} x$; the foci are located at $(\pm c, 0)$ and the eccentricity is $e = \frac{c}{a}$, with $c^2 = a^2 + b^2$. A directrix is defined for also an ellipse and an hyperbola but not for a circle.

By means of polar coordinates, a curve of a conic section -- ellipse, parabola, hyperbola -- is describable as the locus of a point that moves so that a ratio e of its distances from a fixed point and a fixed line remains constant, in which e is the eccentricity as defined above; for an ellipse the range of e is the closed interval $]0,1[$, for a parabola $e = 1$, and for an hyperbola $e > 1$. An ellipse and an hyperbola are also definable as the loci of a point that moves so that the sum and difference, respectively, of its distances from two fixed points remain constant. Fermat showed that every equation of first or second degree in one independent variable is reducible to that of a line or a of a conic section. For a quadratic equation reduced to a form $x^2 + a\,x + b = 0$, a graphical solution is obtained on plotting two points at $(0,1)$ and $(-a, b)$ and drawing a circle with the distance between these two points as diameter; two intercepts of the circle with the abscissal axis are the roots of this equation.

For centred conic sections with $a > 0$, $b > 0$ and $c > 0$, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ is an equation in normal form describing the locus of points defining an ellipse with major axis x if $a > b$ or major axis b if $a < b$, $\frac{x^2}{a^2} + \frac{y^2}{b^2} = -1$ definiing an imaginary ellipse, and $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 0$ defining a double point; $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ describing an hyperbola with two lobes along axis x, $\frac{y^2}{b^2} - \frac{x^2}{a^2} = 1$ with two lobes along axis y, $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$ a double line bisecting axes x and y. For a non-centred conic section $y = a\,x^2$ is an equation in normal form describing the locus of point defining a parabola symmetric about positive axis y, $x = a\,y^2$ is an equation in normal form describing the locus of point defining a parabola symmetric about positive axis x, $y^2 = 0$ defines a double line along axis x, $y^2 = a^2$ defines two lines as $y = \pm\,a$, and $y^2 = -a^2$ defines two imaginary lines.

In three spatial dimensions, a quadric surface is a graph of an equation that is quadratic in coordinates x, y and z of cartesian type, for instance; such an equation has a general form

$$A\,x^2 + B\,y^2 + C\,z^2 + D\,x\,y + E\,y\,z + F\,z\,x + G\,x + H\,y + I\,z + J = 0$$

in which $A, B, ..., J$ are parameters of which the relative values determine the shape of the surfaces, which might be a cylinder elliptic, hyperbolic or parabolic, an ellipsoid, an elliptic paraboloid, an elliptic cone, an hyperbolic paraboloid, or an elliptic hyperboloid of one or two sheets.

Likewise in normal form, an equation $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 1$ defines in three spatial dimensions a

surface designating an ellipsoid, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = -1$ defines an imaginary ellipsoid, and

$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} + \dfrac{z^2}{c^2} = 0$ defines an origin; $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$ defines an hyperboloid of one sheet,

$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 0$ defines a double cone, and $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} - \dfrac{z^2}{c^2} = 1$ defines an hyperboloid of two

sheets, all symmetric about axis z. Furthermore, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 2\,c\,z$ defines an elliptic paraboloid,

$\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 2\,c\,z$ defines an hyperbolic paraboloid having a col, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$ defines an elliptic

cylinder, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = -1$ defines an imaginary elliptic cylinder, $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 0$ defines a degenerate

elliptic cylinder that coincides with axis z, $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$ defines an hyperbolic cylinder, $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 0$

defines two intersecting planes, $x = 2\,c\,y^2$ defines a parabolic cylinder, $x^2 = a^2$ defines two parallel planes at $x = \pm\,a$, and $x^2 = 0$ defines a double plane.

All planar geometry considered above corresponds to the euclidean system attributed to Euclid, who composed the first systematic discussion of geometry. According to this system as stated by Playfair, for a given line and a separate point, there is at most one line through that point that is parallel to the given line. That postulate is inapplicable to other geometries, such as the hyperbolic plane in which, according to Poincare's disc model on the interior of a circle, lines are represented by arcs of circles that are orthogonal to the boundary circle, plus diameters of the boundary circle. In this model, the distance between points P and Q within the circle is

$\mathrm{d}(PQ) = \left| \ln\!\left(\dfrac{|XP|\,|YQ|}{|XQ|\,|YP|} \right) \right|$; here P lies on an intersection of two arcs that are orthogonal to each

other at that intersection, and points X and Y denote the intersections of the arc containing Q at the bounding circle; $|XP|$ denotes the euclidean distance between points X and P. A characteristic property of hyperbolic geometry is that the angles of a triangle add to less than a straight angle (π radians in euclidean geometry); in the limit as the vertices go to ∞, there are even ideal hyperbolic triangles in which all three angles are zero. A non-euclidean geometry can be understood on picturing the drawing of geometric figures on curved surfaces, for example, the surface of a sphere

or the inside surface of a bowl.

trigonometry

In contemporary mathematics, trigonometry is a branch thereof linked closely with algebra, but it originated as an evolution of geometry, and, in particular, of measurement of triangles. We focus on the latter development because the structure of chemical matter is generally described at an atomic level in terms of lengths between two atomic centres and angles involving at least three atomic centres; trigonometry provides us with tools to express these molecular properties. We introduce common trigonometric functions by means of a geometric construction; their relations with other algebraic quantities and other trigonometric functions emerge subsequently.

Rotation of a straight line about a point in two dimensions from an initial ray to a terminal ray causes it to sweep through an angle that we measure in various units; two directions of rotation are clockwise and anticlockwise or counterclockwise, of which the former is conventionally taken as a positive direction and the latter accordingly as a negative direction. An angle comprises two rays or line segments with a common end point, called a vertex, and the two line segments are called sides. Three units of rotation applicable in science and engineering are degree, radian and *grad*. According to tradition originating with or before the ancient Babylonians, a full angle is equal to $360\,^o$, so a straight angle $180\,^o$ and a right angle $90\,^o$, whereas an engineering unit called *grad* has a full angle equal to 400 grad, a straight angle 200 grad and a right angle 100 grad, relative to full angle equaling 2π rad, a straight angle π rad and a right angle $\dfrac{\pi}{2}$ rad in SI unit. The practical units of angular measurement hence include degree, such that $180\,^o = \pi$ rad, or grad, such that 100 grad = a right angle or $90\,^o$ or $\dfrac{\pi}{2}$ rad. The standard position of an angle has its vertex at the origin of a system of cartesian coordinates in a plane and the initial ray along the positive abscissal axis; the standard position of an angle is defined according to the quadrant -- I, II, III or IV,

```
                        ordinate axis
                             +
                             |
              II             |        I
      --  _____|_____   +   abscissal axis
                             |
              III            |        IV
                             |
                             --
```
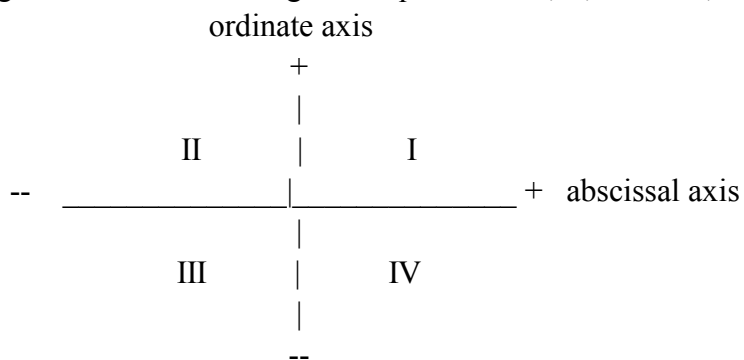
depending on the angle being acute, obtuse, reflex but less than $\dfrac{3\pi}{2}$ rad or reflex but greater than $\dfrac{3\pi}{2}$ rad respectively -- in which the terminal ray lies. The reference angle for an angle in standard position is the acute angle between its terminal ray and the abscissal axis. According to convention, rotating a straight line until it returns to its initial orientation sweeps through a *full angle*; a *straight angle* is half that full angle, and a right angle is half the latter. A triangle that contains a right angle is a right triangle, whereas a triangle with all three sides of equal length,

thus with all three angles equal to $\frac{\pi}{3}$ rad or 60 ^{o}, is equilateral, with two sides of equal length, thus with two equal angles opposite those sides, is isosceles, and with all sides of disparate lengths scalene. For a right triangle, according to a theorem of Pythagoras, the square of the length, c, of the side opposite the right angle, the hypotenuse, is equal to the sum of the squares of the lengths, a,b, of the other two sides, so $c^2 = a^2 + b^2$; conversely, if the lengths of sides of a triangle conform to a relation that the square of one side equals the sum of squares of the other two sides, the figure is a right triangle. Two triangles with equal angles and corresponding sides proportional are similar. For a right triangle, the sum of the other two angles must be $\frac{\pi}{2}$ rad, and one of those two angles is the complement of the other. The natural, and SI, unit of a plane angle is radian, abbreviated rad, or of a solid angle in three dimensions is steradian, abbreviated sr.

Circular trigonometric functions include sine, cosine and tangent, their respective reciprocals cosecant, secant and cotangent, and their respective inverses arcsine, arccosine, arctangent, arccosecant, arcsecant and arccotangent. For complementary angles of which a sum is $\frac{\pi}{2}$ rad, the sine of one angle becomes the cosine of the other, and the cofunction of an angle is the original function of the complementary angle analogously for the other circular functions; the value of a circular trigonometric function depends on only the angle that determines the terminal ray. Two angles in standard position are coterminal when they have the same terminal ray; coterminal angles have the same values of trigonometric functions. Hyperbolic trigonometric functions correspondingly include hyperbolic sine, hyperbolic cosine and hyperbolic tangent, their respective reciprocals hyperbolic cosecant, hyperbolic secant and hyperbolic cotangent and their inverses hyperbolic arcsine, hyperbolic arccosine, hyperbolic arctangent, hyperbolic arccosecant, hyperbolic arcsecant and hyperbolic arccotangent. Hyperbolic functions sinh and cosh are related to a curve for a formula $x^2 - y^2 = 1$, called a *unit hyperbola*, analogously as functions sine and cosine are related to a formula for a *unit circle*, $x^2 + y^2 = 1$. The domain of function sine is d(sin) $= [-\infty, \infty]$ and the range or codomain is r(sin) $= [-1, 1]$, but there are uncountably many values of θ for which sin(θ) has the same value. In relation to mapping of a function to another function or of one set to another set, if two sets be associated in such a way that separate members of a domain are coupled with separate members of their codomain, although not all members of the codomain need be members of a specified range, the mapping is injective. So that sine be an injective function, we restrict by convention its domain to $\left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$; all points in this interval are uniquely connected with points in range $[-1, 1]$. A domain for function cosine is analogously restricted to $[0, \pi]$; tangent is likewise restricted to an open interval $] -\frac{\pi}{2}, \frac{\pi}{2} [$, but in this case, as end points are excluded, a value of tan(θ) also remains finite. Other ranges of values of θ for domains of these functions are selectable as long as a monomorphic nature of each function is preserved. All these circular trigonometric functions are periodic in that a period $p > 0$ exists such that f($x + p$) = f(x). For sine and cosine functions, and their reciprocals, the period is 2 π rad,

whereas for tangent function and its reciprocal cotangent the period is π rad. For such a periodic function the amplitude is half the difference of maximum and minimum values. Two acute angles are complementary when their sum is $\dfrac{\pi}{2}$ rad; for such complementary angles the value of a circular trigonometric function is equal to that of the complementary trigonometric function of the complementary angle, and vice versa -- for $\theta < \dfrac{\pi}{2}$, $\sin(\theta) = \cos\left(\dfrac{\pi}{2} - \theta\right)$, et cetera, as shown at the bottom of the table below. An angle is in a *standard position* if the vertex between its two defining rays is at the origin and the initial side coincides with the abscissal axis; for angle θ not a multiple of $\dfrac{\pi}{2}$, the *reference angle* associated with θ is the acute angle, with positive measure, formed by the abscissal axis and the terminal side of angle θ.

<div align="center">properties of circular trigonometric functions of θ</div>

----

sine -- sin, domain all real numbers, codomain $[-1, 1]$, period 2π, amplitude 1, sinusoidal graph

cosine -- cos, domain all real numbers, codomain $[-1, 1]$, period 2π, amplitude 1, sinusoidal graph

tangent -- tan, domain all real numbers except $\pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2} \dots$, codomain $[-\infty, \infty]$, period π,

asymptotes $\theta = \pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2}, \dots$, θ intercepts $0, \pm\pi, \pm2\pi \dots$, other intercept 0

cotangent -- cot, domain all real numbers except $0, \pm\pi, 2\pi \dots$, codomain $[-\infty, -1], [1, \infty]$, period π,

asymptotes $\theta = 0, \pm\pi, \pm2\pi, \dots$, θ intercepts $\pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2}, \dots$, other intercept undefined

secant -- sec, domain all real numbers except $\pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2} \dots$, codomain $[-\infty, -1], [1, \infty]$, period 2π,

asymptotes $\theta = \pm\dfrac{\pi}{2}, \pm\dfrac{3\pi}{2}, \dots$, θ intercept none, other intercept 1

cosecant -- csc, domain all real numbers except $0, \pm\pi, 2\pi \dots$, codomain $[-\infty, -1], [1, \infty]$, period 2π,

asymptotes $\theta = 0, \pm\pi, \pm2\pi, \dots$, θ intercepts none, other intercept undefined

----

$$\sin(\theta) = \cos\left(\frac{\pi}{2} - \theta\right), \ \tan(\theta) = \cot\left(\frac{\pi}{2} - \theta\right), \ \sec(\theta) = \csc\left(\frac{\pi}{2} - \theta\right)$$

$$\sin\left(\frac{\theta}{2}\right)^2 = \frac{1-\cos(\theta)}{2}, \quad \cos\left(\frac{\theta}{2}\right) = \frac{1+\cos(\theta)}{2}, \quad \tan\left(\frac{\theta}{2}\right) = \frac{\sin(\theta)}{1+\cos(\theta)}$$

properties of $a\sin(b\,\theta + c)$ and $a\cos(b\,\theta + c)$ -- period $\dfrac{2\,\pi}{b}$, amplitude $|a|$, phase shift $-\dfrac{c}{b}$

Each circular trigonometric function is related to another through these formulae, in which for arbitrary value of angle θ the sign of the root is given according to the quadrant of the angle:

$$\sin(\theta) = \pm\sqrt{1-\cos(\theta)^2}, \quad \sin(\theta) = \pm\frac{\tan(\theta)}{\sqrt{1+\tan(\theta)^2}}, \quad \sin(\theta) = \pm\frac{1}{\sqrt{1+\cot(\theta)^2}}, \quad \sin(\theta) = $$

$$\frac{1}{\csc(\theta 0)}$$

$$\cos(\theta) = \pm\sqrt{1-\sin(\theta)^2}, \quad \cos(\theta) = \pm\frac{1}{\sqrt{1+\tan(\theta)^2}}, \quad \cos(\theta) = \pm\frac{\cot(\theta)}{\sqrt{1+\cot(\theta)^2}},$$

$$\cos(\theta) = \frac{1}{\sec(\theta)}$$

$$\tan(\theta) = \pm\frac{\sin(\theta)}{\sqrt{1-\sin(\theta)^2}}, \quad \tan(\theta) = \pm\frac{\cos(\theta)}{\sqrt{1-\cos(\theta)^2}}, \quad \tan(\theta) = \frac{1}{\cot(\theta)}$$

with corresponding formulae for the reciprocals of these functions.

Of hyperbolic trigonometric functions, both *sinh* and *tanh* are monomorphic or "1:1", whereas *cosh* is *bimorphic* or *double-valued* or "2:1": for this reason the domain of *sinh* and *tanh* is $[-\infty, \infty]$ whereas $[0 \le x]$ for *cosh*. Domains of inverse circular and inverse hyperbolic functions also are defined so as to be monomorphic or to display behaviour 1:1.

Although we introduce circular trigonometric functions here with angles as arguments, a real variable might equally serve as a variable, which makes a function containing such a trigonometric function periodic; the smallest positive number τ for which the periodic function $f(x+\tau) = f(x)$ is the period. The sum or difference of two periodic functions is also periodic. For a general sine formula of type $f(x) = \alpha\sin(\omega x + \phi)$ containing parameters α, ω and ϕ as constants and the former two also positive constants, the amplitude α is half the range of $f(x)$, the period is $\dfrac{2\,\pi}{\omega}$ and the *phase shift* is $-\dfrac{\phi}{\omega}$. Replacement of a real variable with an imaginary variable in a circular trigonometric function annuls its periodic properties and converts it to an hyperbolic trigonometric function.

Although for the sine of an angle θ raised to a power n we might write $\sin^n\theta$, we must compute this quantity as $\sin(\theta)^n$, i.e. first finding the sine of the angle and taking the result to power n, and likewise for other trigonometric functions. Likewise a notation $\sin^{(-1)}\theta$ implies not the reciprocal of $\sin(\theta)$ but its inverse, $\arcsin(\theta)$. The formula $\sin(\theta)$, which is defined for all real θ, lacks an inverse, but with the domain of function sine restricted to $\left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right]$, $\sin(\arcsin(\theta)) = \theta$ for all θ in an interval $[-1, 1]$; likewise $\arcsin(\sin(\theta)) = \theta$ and $\arctan(\tan(\theta)) = \theta$ for all θ in an interval

$\left[-\dfrac{\pi}{2}, \dfrac{\pi}{2} \right]$, but for $\arccos(\cos(\theta)) = \theta$ the interval of θ is $[0, \pi]$. A function that has, as a value, a principal value of a function with many values is indicated conventionally by writing it with an initial majuscule or capital letter, such as in $\mathrm{Sin}^{(-1)}(\theta)$, but for *Maple* arcsin(θ) has that significance.

For a triangle with a side of length a opposite an angle of extent α, a side of length b opposite an angle of extent β and the other side of length c opposite an angle of extent γ, the law of sines is

$$\frac{\sin(\alpha)}{a} = \frac{\sin(\beta)}{b} = \frac{\sin(\gamma)}{c}$$

Use of this law to solve a general triangle requires at least the length of one side and the extent of the opposite angle and one other side or angle, but two possible solutions might arise. For the same triangle, the cosine law is for the side of length c,

$$c^2 = a^2 + b^2 - 2\,a\,b\,\cos(\gamma)$$

and analogously for any other side. Use of this law to solve a general triangle requires either the lengths of three sides or of two sides and the extent of the included angle between those two sides. According to the tangent rule,

$$\tan\!\left(\frac{\beta - \gamma}{2}\right) = \frac{b - c}{b + c}\,\cot\!\left(\frac{\alpha}{2}\right)$$

or, with its cyclic permutations among a, b, c and α, β, γ,

$$\frac{a - b}{a + b} = \frac{\tan\!\left(\dfrac{\alpha - \beta}{2}\right)}{\tan\!\left(\dfrac{\alpha + \beta}{2}\right)} = \frac{\tan\!\left(\dfrac{\alpha - \beta}{2}\right)}{\cot\!\left(\dfrac{\gamma}{2}\right)}$$

the former case involves three angles and two sides, or just two sides and two angles in the latter case at the left; these formulae are also useful in solving plane triangles. According to Heron's formula, the area of a triangle equals half the product of the lengths of two sides multiplied by the sine of the included angle:

$$A = \frac{1}{2}\,a\,b\,\sin(\gamma), \text{ or } A = \sqrt{s\,(s - a)\,(s - b)\,(s - c)} = \rho\,s$$

in which $s = \dfrac{a + b + c}{2}$ is half the perimeter of the triangle and ρ is the radius of the inscribed circle,

$$\rho = (s - a)\tan\!\left(\frac{\alpha}{2}\right) = (s - b)\tan\!\left(\frac{\beta}{2}\right) = (s - c)\tan\!\left(\frac{\gamma}{2}\right)$$

$$\rho = s\,\tan\!\left(\frac{\alpha}{2}\right)\tan\!\left(\frac{\beta}{2}\right)\tan\!\left(\frac{\gamma}{2}\right) = 4\,r\,\sin\!\left(\frac{\alpha}{2}\right)\sin\!\left(\frac{\beta}{2}\right)\sin\!\left(\frac{\gamma}{2}\right),$$

with radius r of the circumcircle of a triangle,

$$r = \frac{a}{2\,\sin(\alpha)} = \frac{b}{2\,\sin(\beta)} = \frac{c}{2\,\sin(\gamma)}$$

and

$$s = 4\,r\,\cos\!\left(\frac{\alpha}{2}\right)\cos\!\left(\frac{\beta}{2}\right)\cos\!\left(\frac{\gamma}{2}\right)$$

Each triangle has also three escribed circles, each of which is a circle tangent to one side and to the extensions of the other two sides of a triangle; the center of each circle is an excentre with an exradius ρ_a or ρ_b or ρ_c to the perimeter.

$$\rho_a = s\,\tan\!\left(\frac{\alpha}{2}\right) = \frac{a\,\cos\!\left(\frac{\beta}{2}\right)\cos\!\left(\frac{\gamma}{2}\right)}{\cos\!\left(\frac{\alpha}{2}\right)}$$

$$\rho_b = s\,\tan\!\left(\frac{\beta}{2}\right) = \frac{b\,\cos\!\left(\frac{\alpha}{2}\right)\cos\!\left(\frac{\gamma}{2}\right)}{\cos\!\left(\frac{\beta}{2}\right)}$$

$$\rho_c = s\,\tan\!\left(\frac{\gamma}{2}\right) = \frac{c\,\cos\!\left(\frac{\alpha}{2}\right)\cos\!\left(\frac{\beta}{2}\right)}{\cos\!\left(\frac{\gamma}{2}\right)}$$

A triangle has also three medians, each of which is a straight line joining one vertex of a triangle to the midpoint of the opposite side, each of length

$$s_a = \frac{1}{2}\sqrt{b^2 + c^2 + 2\,b\,c\,\cos(\alpha)}$$

$$s_b = \frac{1}{2}\sqrt{a^2 + c^2 + 2\,a\,c\,\cos(\beta)}$$

$$s_c = \frac{1}{2}\sqrt{a^2 + b^2 + 2\,a\,b\,\cos(\alpha)}$$

three bisectors of angles from a vertex to the opposite side, each of length

$$w_\alpha = \frac{2\,b\,c\,\cos\!\left(\frac{\alpha}{2}\right)}{b + c}$$

$$w_\beta = \frac{2\,a\,c\,\cos\!\left(\frac{\beta}{2}\right)}{a + c}$$

$$w_\gamma = \frac{2\,a\,b\,\cos\!\left(\frac{\gamma}{2}\right)}{a + b}$$

and three altitudes perpendicular from a vertex to the opposite side of a triangle, each of length

$$h_a = b\,\sin(\gamma) \;=\; c\,\sin(\beta)$$
$$h_b = a\,\sin(\gamma) \;=\; c\,\sin(\alpha)$$
$$h_c = a\,\sin(\beta) \;=\; b\,\sin(\alpha)$$

The stature of a triangle over side a is
$$h_a = b \sin(\gamma) = c \sin(\beta)$$
In terms of that height the surface area of a triangle is
$$A = \frac{1}{2} h_a \, a$$
and the latter formulae are permutable for the other sides.

With $\alpha + \beta + \gamma = \pi$ rad, other formulae for a plane triangle are
$$\sin\left(\frac{\gamma}{2}\right) = \sqrt{\frac{(s-a)(s-b)}{ab}}$$
$$\cos\left(\frac{\gamma}{2}\right) = \sqrt{\frac{s(s-c)}{ab}}$$
$$\tan\left(\frac{\gamma}{2}\right) = \frac{\sin\left(\dfrac{\gamma}{2}\right)}{\cos\left(\dfrac{\gamma}{2}\right)}$$
$$\frac{a+b}{c} = \frac{\cos\left(\dfrac{\alpha-\beta}{2}\right)}{\cos\left(\dfrac{\alpha+\beta}{2}\right)} = \frac{\cos\left(\dfrac{\alpha-\beta}{2}\right)}{\sin\left(\dfrac{\gamma}{2}\right)}$$
$$\frac{a-b}{c} = \frac{\sin\left(\dfrac{\alpha-\beta}{2}\right)}{\sin\left(\dfrac{\alpha+\beta}{2}\right)} = \frac{\sin\left(\dfrac{\alpha-\beta}{2}\right)}{\cos\left(\dfrac{\gamma}{2}\right)}$$

Mollweide's formulae, of which cyclic permutations among a, b, c and α, β, γ furnish four analogous relations,
$$\frac{a+b}{c} = \frac{\cos\left(\dfrac{\alpha-\beta}{2}\right)}{\sin\left(\dfrac{\gamma}{2}\right)}$$
$$\frac{a-b}{c} = \frac{\sin\left(\dfrac{\alpha-\beta}{2}\right)}{\cos\left(\dfrac{\gamma}{2}\right)}$$

this tangent formula,
$$\tan(\gamma) = \frac{c \sin(\alpha)}{b - c \cos(\alpha)} = \frac{c \sin(\beta)}{a - c \cos(\beta)}$$
and a projection formula,
$$c = a \cos(\beta) + b \cos(\alpha)$$
According to the application of selected items among the above, given

- one side c and adjacent angles α and β,

- two sides a and b and angle γ between them,

- all three sides a, b, c, and

- two sides a and b and one opposite angle α,

one can determine uniquely the other angles, sides and area of the plane triangle.

Direction ratios are numbers proportional to direction cosines; for direction ratios a, b, c in a space $\mathbf{R}^3$ of three dimensions,

$$\cos(\alpha) = \frac{a}{\sqrt{a^2 + b^2 + c^2}}, \quad \cos(\beta) = \frac{b}{\sqrt{a^2 + b^2 + c^2}}, \quad \cos(\gamma) = \frac{c}{\sqrt{a^2 + b^2 + c^2}}$$

The cosine of angle θ between two lines with direction angles α_1, β_1, γ_1 and α_2, β_2, γ_2 is

$$\cos(\theta) = \cos(\alpha_1) \cos(\alpha_2) + \cos(\beta_1) \cos(\beta_2) + \cos(\gamma_1) \cos(\gamma_2)$$

For a plane defined by equation $A x + B y + C z = D$, coefficients A, B and C are direction ratios for any normal, or line perpendicular, to that plane. The distance from that plane to a point $P_1 = (x_1, y_1, z_1)$ is

$$\frac{A x_1 + B y_1 + C z_1 - D}{\sqrt{A^2 + B^2 + C^2}}$$

The equations of a straight line that is such a normal passing point P_1 are

$$\frac{x - x_1}{A} = \frac{y - y_1}{B} = \frac{z - z_1}{C}$$

The cosine of angle θ between two planes of which normals have direction ratios A_1, B_1, C_1 and A_2, B_2, C_2, or between two lines with these direction ratios, is

$$\cos(\theta) = \frac{A_1 A_2 + B_1 B_2 + C_1 C_2}{\sqrt{A_1^2 + B_1^2 + C_1^2} \sqrt{A_2^2 + B_2^2 + C_2^2}}$$

These relations are useful in calculation of interatomic distances, interbond angles and related quantities.

trigonometry and complex number

A complex number of unit modulus is expressible as $\cos(\theta) + i \sin(\theta)$, which is in turn related to a complex exponential quantity through Euler's formula,

$$\mathbf{e}^{(i\,\theta)} = \cos(\theta) + i \sin(\theta)$$

with $i = \sqrt{-1}$. Raising both sides of this formula to power n yields

$$(\mathbf{e}^{(i\,\theta)})^n = (\cos(\theta) + i \sin(\theta))^n$$
$$= \mathbf{e}^{(i\,n\,\theta)} = \cos(n\,\theta) + i \sin(n\,\theta)$$

which constitutes de Moivre's formula that is valid for all complex x and all n, positive or negative, integer or rational or real. A polar representation of complex number $z = x + i y$ in its cartesian

form is $z = r \, (\cos(\theta) + i \sin(\theta))$; therein modulus $|z| = r = \sqrt{x^2 + y^2}$ represents the distance between the origin and point (x, y) in the complex plane, and, in that complex plane, axis x becomes the real axis and axis y becomes the imaginary axis. In polar coordinates r, θ are related to rectangular coordinates x, y through $x = r \cos(\theta)$ and $y = r \sin(\theta)$, in which $\theta = \arg(z)$ with θ as argument, amplitude or phase of z is represented geometrically as the angle between the half line defined by positive axis x and a line segment joining the origin to point (x, y). Because trigonometric functions are periodic, θ or $\arg(z)$ has uncountably many values, one differing from another by a multiple of $2\,\pi$; its principal value is the particular value that lies in interval $]-\pi, \pi]$.

To calculate that principal value, we apply $\tan(\theta) = \dfrac{y}{x}$, which for known x and y yields a value in

interval $]-\dfrac{\pi}{2}, \dfrac{\pi}{2}]$; because the latter interval fails to coincide with the preceding interval, if (x, y) lie in quadrant II we add π to $\arctan\left(\dfrac{y}{x}\right)$, whereas if (x, y) lie in quadrant III we subtract π from

$\arctan\left(\dfrac{y}{x}\right)$. Some particular principal values of $\arctan\left(\dfrac{y}{x}\right)$ for (x, y) are for (x, x), 0; for $(x, -x)$, π;

for $(-x, x)$, $\dfrac{3\,\pi}{4} + 2\,n\,\pi$ with integer n, but for $(0,0)$ θ is undefined.

The distance r between two points representing z_1 and z_2 in the complex plane is expressed as
$$r = |z_2 - z_1| = |x_2 - x_1 - i\,(y_2 - y_1)| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$
With z_0 complex and r real, relation $r = |z - z_0|$ defines the locus of points z in the complex plane to represent a circle of radius r centred at z_0.

A geometric interpretation of addition of two complex numbers involves the parallelogram law, according to which the diagonal of the resultant parallelogram defines a boundary between two triangles, one side of each of which is defined by one addend of the sum; because the sum of lengths of any two sides of a triangle must be greater than or equal to the length of the other side, we obtain these triangle inequalities:
$$|z_1 + z_2| \le |z_1| + |z_2|,$$
$$\left||z_1| - |z_2|\right| \le |z_1 - z_2|,$$
which are combined in the following relation.
$$\left||z_1| - |z_2|\right| \le |z_1 + z_2| \le |z_1| + |z_2|$$
This result is extensible to complex numbers of finite number according to a generalized triangle inequality,
$$\left|\sum_{j=1}^{n} z_j\right| \le \sum_{j=1}^{n} |z_j|$$

We define an euclidean norm such that
$$N(x + i\,y) = x^2 + y^2 = (x + i\,y)\,(x - i\,y);$$
multiplication by a fixed number $z_0 \ne 0$ multiplies all lengths by $\sqrt{N(z_0)}$, which defines an

euclidean similarity. When $N(z_0) = 1$, $z \to z_0\, z$ becomes an *euclidean congruence* or *isometry*. For a complex unit u, which implies complex number u to have <u>modulus</u> unity, a *map* $z \to u\, z$ corresponds to a *rotation*, and $z \to u\, \bar{z}$ constitutes a *reflexion*. Specifically, for $u = 1$, reflexion occurs about axis x; for $u = -1$, reflexion occurs across axis y; for $u = i$, rotation occurs by $\dfrac{\pi}{2}$ rad anticlockwise, which also corresponds to reflexion across a diagonal bisecting quadrants I and III, or at angle $\dfrac{\pi}{4}$ anticlockwise to axis x, and for $u = -i$, rotation occurs by $-\dfrac{\pi}{2}$, which also corresponds to reflexion bisecting quadrants II and IV or at angle $-\dfrac{\pi}{4}$, so clockwise to axis x.

Two geometric operations pertain to properties of groups: a <u>special orthogonal group</u> SO_2 that comprises all multiplications $z \to u\, z$ by complex unit u corresponds to a circle of real angles θ, considered <u>mod</u> $2\,\pi$, whereas a *general orthogonal group* GO_2 includes the former with an additional map $z \to u\, \bar{z}$.

sequence and series

Among <u>mathematical</u> <u>expressions</u> that have diverse and important applications in a chemical context, a <u>sequence</u> implies a <u>function</u> defined on a set J of all <u>positive</u> <u>integers</u> as domain; for $f(j) = x_j$ with j in J, x_1, x_2, x_3 ... or (x_j) denotes the sequence f, of which the <u>values</u> as <u>elements</u> x_j are called <u>terms</u> of that sequence, and j serves as a <u>counting index</u>. Examples of such sequences are, numerical, $1, \dfrac{1}{2}, \dfrac{1}{3}, \dfrac{1}{4}, ...$ and, algebraic, $1, x, x^2, x^3, ...$ The terms of a sequence need not be distinct, as in <u>Fibonacci's sequence</u> -- 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, ..., and an <u>infinite</u> sequence contains elements of <u>uncountable</u> number. Alternatively, a sequence of numbers comprises a set in a definite order and fomed according to a specific rule. Of the two parts of a <u>recursive definition</u> of a sequence, the first part specifies the first term or terms of that sequence, and the second part indicates how each term is calculated from the preceding term or terms. For instance, a first term $x_0 = 1$ and a relation $x_n = 2\,x_{n-1} + 3$ generates a numerical sequence 1, 5, 13, 29, 61, 125, ..., but, for those six terms of that sequence, recursive definitions of uncountable number generate those particular values.

A sequence (x_j) in a <u>metric space</u> X <u>converges</u> if there exist a point p such that for each <u>epsilon</u> or $\varepsilon > 0$ there is an integer n for which $j \geq n$ implies that in X a distance $d(x_j, p) < \varepsilon$, or for which p is a <u>limiting</u> value of (x_j); if there exist no such limit, (x_j) <u>diverges</u>. Such a criterion for convergence depends not only on (x_j) but also on X: for instance the sequence $(\dfrac{1}{j})$ for positive integer j converges in $\mathbf{R}^{\,l}$ to 0, but fails to converge in the set of all positive real numbers. The <u>set</u> of all points x_j with $j = 1, 2, 3, ...$ is the <u>range</u> of (x_j), which might be finite or infinite; the sequence (x_j) is <u>bounded</u> if its range be bounded. A sequence has *increasing* terms if $x_{n+1} > x_n$ for all counting numbers n, *non-decreasing* terms if $x_{n+1} \geq x_n$, *decreasing* terms if $x_{n+1} < x_n$ and non-decreasing terms if $x_{n+1} \leq x_n$; a sequence possesses an <u>upper bound</u> if, for any j, $x_j < U$, or a

lower bound if $x_j > L$. A bounded and monotonic sequence has a limit, and a monotonic sequence has a limit only if it be bounded; for that limit L of a sequence, for a given $\varepsilon > 0$ there exists a counting number N such that, for all $N > j$, $|x_j - L| < \varepsilon$. If sequence $\{x_j\}$ have limit ξ and sequence $\{y_j\}$ have limit υ, $\{c\,x_j\}$ has limit $c\,\xi$ for a constant c, $\{x_j \pm y_j\}$ has a limit $\xi \pm \upsilon$, $\{x_j\,y_j\}$ has a limit $\xi\,\upsilon$, and $\{\dfrac{x_j}{y_j}\}$ has a limit $\dfrac{\xi}{\upsilon}$ provided that $\upsilon \neq 0$ and that no $y_j = 0$.

An arithmetic progression comprises a sequence of numbers with a common interval between consecutive terms; the general formula for the sum of an arithmetical progression having n terms is $S_n = \dfrac{n}{2}(f + l)$, with first term f and last term l A geometric progression comprises a sequence of numbers with a common ratio between consecutive terms, such as 1, 2, 4, 8, 16, ...; term n of that progression is $a\,r^{(n-1)}$ with ratio r and first member a. The terms between any two specified terms in a progression are the means of that progression, arithmetic mean or means for an arithmetic progression or geometric mean or means for a geometric progression.

A sequence u_1, u_2, u_3, ... has a limit if, for every preassigned positive number ε, one can discover a number N such that $|u_n - L| < \varepsilon$ for all integers $n > N$. If a sequence have a limit L, this condition is expressed as $\lim\limits_{n \to \infty} u_n = L$; if no limit exist, this condition is expressed as $\lim\limits_{n \to \infty} u_n = \infty$. If $\lim\limits_{n \to \infty} u_n$ and $\lim\limits_{n \to \infty} v_n$ exist,

- $\lim\limits_{n \to \infty} u_n + v_n = (\lim\limits_{n \to \infty} u_n) + (\lim\limits_{n \to \infty} v_n)$ and $\lim\limits_{n \to \infty} u_n - v_n = (\lim\limits_{n \to \infty} u_n) - (\lim\limits_{n \to \infty} v_n)$;

- $\lim\limits_{n \to \infty} u_n\,v_n = (\lim\limits_{n \to \infty} u_n)(\lim\limits_{n \to \infty} v_n)$

- $\lim\limits_{n \to \infty} \dfrac{u_n}{v_n} = \dfrac{\lim\limits_{n \to \infty} u_n}{\lim\limits_{n \to \infty} v_n}$ provided that $\lim\limits_{n \to \infty} v_n \neq 0$;

- if $\lim\limits_{n \to \infty} v_n = 0$ and $\lim\limits_{n \to \infty} u_n \neq 0$, $\lim\limits_{n \to \infty} \dfrac{u_n}{v_n}$ does not exist;

- if $\lim\limits_{n \to \infty} v_n = 0$ and $\lim\limits_{n \to \infty} u_n = 0$, $\lim\limits_{n \to \infty} \dfrac{u_n}{v_n}$ might exist;

- $\lim\limits_{n \to \infty} v_n^{\,k} = (\lim\limits_{n \to \infty} v_n)^k$ for any real number k.

A sequence is bounded if, for positive number N independent of n, $|u_n| \leq N$. A sequence increases monotonically if $u_{n+1} \geq u_n$ or decreases monotonically if $u_{n+1} \leq u_n$. A bounded and monotonically increasing or decreasing sequence has a limit, but a sequence might have a limit without being monotonically increasing or decreasing.

A series is defined as a sum of sequential terms of countable or uncountable number, such as this infinite numerical series,

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots + \frac{1}{2^k} + \ldots$$

or this finite symbolic series,

$$a_0 + a_1 + a_2 + \ldots + a_n$$

in contrast to this sequence,

$$1, \ \frac{1}{2}, \ \frac{1}{4}, \ \frac{1}{8}, \ \ldots \ \frac{1}{2^k}, \ \ldots$$

or this,

$$a_0, \ a_1, \ a_2, \ a_3, \ \ldots, \ a_n$$

each of which is merely an ordered set of such terms; each term in such a sequential sum is defined according to a formula or prescription. Such a sequential sum might be finite or infinite, according to whether it terminates at a particular term or continues indefinitely. We represent the former series as a sum with a Sigma or Σ notation as $\sum_{k=1}^{\infty} \frac{1}{2^k}$ and the latter series as $\sum_{j=1}^{n} a_j$ in

contrast with the latter sequences $\{\frac{1}{2^k}\}$ and $\{a_j\}$. The domain of an infinite sequence is the set of natural numbers. An alternative method to generate this sequence involves use of a recurrence relation as a prescription, according to which each successive term y_j is obtained from the

preceding term y_{j-1}; in the case of $\sum_{k=1}^{\infty} \frac{1}{2^k}$ or $\{\frac{1}{2^k}\}$, for counting index j beginning at unity the relation is

$$y_j = 1 \text{ if } j = 1 \text{ or } \frac{y_{j-1}}{2} \text{ otherwise}$$

As a generator of sums, we consider $(a + 1)^n$; for positive integer n, $n \geq 0$, the sum has a finite number of terms and the coefficients of a and b involve binomial coefficients $\begin{bmatrix} n \\ k \end{bmatrix} = \dfrac{n!}{k! \, (n-k)!}$

.

$$(a+1)^n = \begin{bmatrix} n \\ 0 \end{bmatrix} a^n + \begin{bmatrix} n \\ 1 \end{bmatrix} a^{(n-1)} + \begin{bmatrix} n \\ 2 \end{bmatrix} a^{(n-2)} + \ldots = \sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} a^{(n-k)}$$

Such sums are generated readily as Taylor series for arbitrary values of exponent n. The coefficients follow regular patterns

$n = -3$ 1, −3, 6, −10, 15, ...
$n = -2$ 1, −2, 3, −4, 5, ...
$n = -1$ 1, −1, 1, −1, 1, ...
$n = 0$ 1, 0, 0, 0, 0, ...
$n = 1$ 1, 1, 0, 0, 0, ...
$n = 2$ 1, 2, 1, 0, 0, ...
$n = 3$ 1, 3, 3, 1, 0, ...

A sequence, in symbolic form with term y_j such as

$$y_1, y_2, y_3, \ldots$$

represents a function; its domain is specified as either a subset of positive integers j or all positive integers. For a <u>finite</u> series, there is an upper bound of j that enumerates the terms,

$$a_0 + a_1 + a_2 + \ldots + a_n$$

in this case for $j = n$. This series

$$2^0 + 2^1 + 2^2 + 2^3 + 2^4 + 2^5 + \ldots$$

is a constant series, such that a <u>partial sum</u>, of arbitrary length, plus unity has also a value $2^{(n+1)}$, in which n is a positive integer:

$$\sum_{j=0}^{n} 2^j = 2^{(n+1)} - 1$$

This sequence,

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots$$

known as an <u>harmonic progression</u> because the reciprocals of terms in the sequence conform to an <u>arithmetic progression</u>, yields an <u>harmonic series</u>;

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \ldots$$

the individual terms form a sequence, above, that <u>converges</u> to a limiting value $\frac{1}{\infty}$ = zero, but the sum of this harmonic series <u>diverges</u>: decomposing this series into a sum of partial sums,

$$1 + \frac{1}{2} + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \ldots$$

of which each partial sum of terms within parentheses is greater than $\frac{1}{2}$, as $\frac{1}{3} + \frac{1}{4} > \frac{1}{4} + \frac{1}{4}$,

$$\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} > \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$ et cetera demonstrates that this total sum increases without limit. The concept of convergence of a sequence is thus distinct from the concept of the convergence of a series. If a series $\sum_{j=1}^{\infty} a_j$ converge, $\lim_{j \to \infty} a_j = 0$. Even if a series converge, the rate of convergence might vary greatly: for instance, this series

$$\frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + \ldots + \frac{1}{r^2} + \ldots$$

converges slowly, requiring more than 500 terms to achieve precision in the third digit, whereas an alternative series

$$\frac{\pi}{2} = \frac{1}{1} + \frac{1(1)}{1(3)} + \frac{1(1)(2)}{1(3)(5)} + \frac{1(1)(2)(3)}{1(3)(5)(7)} + \ldots$$

achieves convergence to the third digit in only ten terms. The latter two series are instances of

non-negative series, for which each term is non-negative. If series $\sum_{j=1}^{\infty} a_j$ have sum α and $\sum_{j=1}^{\infty} b_j$ have sum β, $\sum_{j=1}^{\infty} c\, a_j = c\, \alpha$ for constant c and $\sum (a_j \pm b_j) = \alpha \pm \beta$.

A geometric series has this form,

$$a + a\, r + a\, r^2 + a\, r^3 + a\, r^4 + \dots$$

has a finite sum, or converges to a limiting value, if $\lim_{n \to \infty} \left| \dfrac{a\, r^n}{a\, r^n - 1} \right| < 1$ or $|r| < 1$; this constraint on values of r for which the series converges defines its interval of convergence, which is twice the radius of convergence of a power series. The partial sum of the first n terms is

$$S_n = a + a\, r + a\, r^2 + \dots + a\, r^{(n-1)} = \frac{a\, (r^n - 1)}{r - 1}$$

and the sum of an infinite geometric series is

$$S = \frac{a}{1 - r}$$

For an alternating series, the signs of consecutive terms alternate between positive and negative; the maximum possible error when an alternating series is truncated is the first omitted term.

A formula or function of a particularly general and useful kind is a *functional series* in which terms contain a variable with non-negative integers as powers; the variable might be a simple symbol, here x,

$$f(x) = c_0 + c_1\, x + c_2\, x^2 + c_3\, x^3 + \dots$$

with coefficients c_j and $j = 0, 1, 2, 3 \dots$, or be itself an expression involving a function, such as e^x in this series:

$$g(x) = c_0 + c_1\, e^x + c_2\, e^{(2x)} + c_3\, e^{(3x)} + \dots$$

A power series can be expanded about a centre other than the origin of a system of coordinates, such as a in

$$h(x) = c_0 + c_1\, (x - a) + c_2\, (x - a)^2 + c_3\, (x - a)^3 + \dots$$

in contrast with power series $f(x)$ above for which the expansion centre is implicitly zero. A power series hence contains terms containing a variable to ascending positive powers, either with an explicit centre of expansion as above or in an expanded form. A polynomial is a formal power series that has, different from zero, coefficients of only finite number; a monomial is an expression comprising a single term, such as $3\, c\, x^2$, and formally each such term constitutes a basis element such that a linear combination of these monomials in sums and differences becomes a polynomial. Other basis elements might be Lagrange functions of form $\dfrac{x - x_j}{\prod_{j <> k} (x_j - x_k)}$, or

Bernstein polynomials of form $_n C_k\, x^k\, (1 - x)^{(n-k)}$ in which appears the binomial coefficient $_n C_k$

$$= \frac{n!}{k!\,(n-k)!}$$ defined in section 1.116.

A polynomial, or multinomial, formula is a mathematical expression that comprises a sum of terms of finite number, each term of which is a product of a constant and one, or more, variables or indeterminates raised to a power that is a non-negative integer; a simple polynomial of form

$$c_0 + c_1 x + c_2 x^2 + c_3 x^3$$

in variable x and with coefficients c_j, $j = 0, 1, 2, 3...$, is thus a finite power series. The degree of a polynomial in a particular variable is the largest power to which that variable is raised. An infinite series, such as f(x) above, is a functional series that has a form resembling that of a polynomial except that it terminates at no particular power of the independent variable, x in the above examples; a consequence of this lack of termination is a requirement to specify a domain that includes only those real numbers for which f(x) remains finite. For instance the power series

$$f(x) = 1 + x + x^2 + x^3 + ...$$

yields no finite result for $x > 1$ or for $x < -1$: in the former case the value of the sum increases without limit, whereas in the latter case the sum oscillates between positive and negative values of large magnitudes as further terms become included. Power series are useful not only in chemistry because they provide an alternative means to represent transcendental functions such as exponential and trigonometric functions, because they allow a simple means to investigate the properties of complicated functions for small or large values of an independent variable, and because they allow a representation of data without requiring any particular chemical or physical basis of such representation; as instances of the latter, the thermal capacity $C_p(T)$ at constant pressure of a substance as a function of temperature T or the electric dipolar moment p(R) of a free diatomic molecule as a function of internuclear separation R might represent data for those properties that results from fits of experimental or theoretical evaluations at various values of the respective independent variables. The disadvantages of these representations are that they are unreliable not only beyond the domain of definition from fitted data -- extrapolation, but also even within that domain except at or near a fitted value.

The fundamental theorem of algebra states that a polynomial of degree n has at least one root and possibly n roots, which are in general complex but which might be purely real or purely imaginary depending on the degree and the values of coefficients. A general equation of a polynomial of degree n, written in order of descending values of powers of variable x, is

$$P(x) = a_n x^n + a_{n-1} x^{(n-1)} + a_{n-2} x^{(n-2)} + \; ... \; + a_1 x + a_0 = 0$$

in which each coefficient a_j is a real number with $a_n \neq 0$. This function $y = P(x)$ is continuous throughout the entire domain of x. Its limit as $x \to \infty$ is $+\infty$ if $a_n > 0$ or $-\infty$ if $a_n < 0$; its limit as $x \to -\infty$ is $+\infty$ for $a_n > 0$ and n even or $a_n < 0$ and n odd, or $-\infty$ for $a_n > 0$ and n odd or $a_n < 0$ and n even. If n be odd, the graph of $y = P(x)$ intersects the x-axis at least once; the point of intersection corresponds to a solution of an equation $P(x) = 0$. If n be even and $a_n > 0$, $P(x)$ has a global minimum: there exists a point $x = a$ at which f(a) $\leq$ f(x) for all x; if n be even and $a_n < 0$, $P(x)$ has a global maximum. For $n \geq 2$, function $y = P(x)$ has at most $n - 1$ local extrema, which are alternately local minima and local maxima. For $n \geq 3$, the graph of $y = P(x)$ has at most $n - 2$

<u>points of inflexion</u>. If the roots of $P(x)$, which are particular values of x that satisfy this general but <u>conditional</u> equation, be $r_1, r_2, \ldots, r_n$,

$$P(x) = a_0 (x - r_1)(x - r_2) \ldots (x - r_n)$$

These roots have <u>symmetric</u> formulae such that

$$-\frac{a_{n-1}}{a_n} = \sum_{i=1}^{n} r_i,$$

$$\frac{a_{n-2}}{a_n} = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} r_i r_j \right),$$

$$-\frac{a_{n-3}}{a_n} = \sum_{k=1}^{n} \left(\sum_{j=1}^{n} \left(\sum_{i=1}^{n} r_i r_j r_k \right) \right),$$

$$\ldots,$$

$$r_1 r_2 \ldots r_n = \frac{(-1)^n a_0}{a_n}$$

If m roots have the same value r, r is a <u>multiple</u> or <u>repeated root</u> of <u>order</u> m; hence $P(x) = (x - r)^m Q(x)$. All multiple roots are zeros of the <u>greatest common divisor</u> of $P(x)$. If r be a root of $P(x)$, so that $P(r) = 0$, $P(r)$ contains a <u>factor</u> $x - r$ such that <u>division</u> of $P(x)$ by $x - r$ yields a formula $Q(x)$ of degree decreased by unity. When $P(a)$ and $P(b)$ have opposite <u>signs</u> for real a and b, a real root lies between a and b; an <u>interpolated</u> value

$$c = a - \frac{P(a)(b - a)}{P(b) - P(a)}$$

provides an estimate of that real root, and repetition of this calculation with $P(c)$ produces closer limits. For $\dfrac{b - a}{P(b) - P(a)}$ to become replaced with a <u>reciprocal</u> of a <u>derivative</u> constitutes <u>Newton's method</u> to find roots, which is implemented in *Maple*'s procedures to <u>solve</u> equations; this formula we present explicitly in section 3.308. For a general polynomial $P(x)$, if an approximate complex root can be found, with Newton's method one generates improved estimates of that root, subject to certain conditions. For a polynomial formula $P(x)$ such as that above with real numbers for all coefficients, if complex number $a + b i$ with $i = \sqrt{-1}$ and $b \neq 0$ be a root of that equation $P(x) = 0$, its complex conjugate $a - b i$ is also a root of that equation: such conjugate roots thus invariably appear in couples. For a polynomial formula $P(x)$ with real coefficients written with x in order of either ascending or descending powers, a variation of sign occurs where two successive coefficients have opposite signs; for a polynomial formula $P(x)$ such as that above with real numbers for all coefficients, for the equation $P(x) = 0$ the number of positive roots is either equal to the number of variations of sign of $P(x)$ or less than that number by an even integer, and the number of negative roots is either equal to the number of variations of sign of $P(-x)$ or is less than that number by an even integer. For a polynomial formula $P(x)$ such as that above with rational numbers for all coefficients, if $b + c \sqrt{d}$ in which b, c, d be rational and $\sqrt{d}$ irrational be a root of that equation, another root of that equation is $b - c \sqrt{d}$. For a polynomial formula $P(x)$ such as

that above but with coefficients restricted to integers, a theorem on rational roots asserts that, for a rational number $\dfrac{p}{q}$ in which p and q have no common factor other than ± 1 and that is a root of the equation $P(x) = 0$ of that polynomial, p is a factor of a_0 and q is a factor of a_n.

   Polynomial functions with special properties arise in solutions of differential equations; orthogonal polynomials associated with Chebyshev, Hermite, Laguerre and Legendre that have particular chemical applications appear in subsequent chapters. Many a special function, defined as a non-elementary transcendental function, arises typically as a solution of an integral or a differential equation, and has thereby no simple algebraic representation; examples of special functions already encountered in section 1.117 are gamma Γ and beta B functions, which we relate to integrals in section 4.304, whereas Bessel and hypergeometric functions arise from differential equations. Such functions we encounter also in subsequent chapters, whereas in this chapter our attention is devoted to geometry, trigonometry, series, complex functions and related topics.

   For a numerical representation of a function, such as numerical results obtained in a laboratory, a numerical value of a dependent quantity might be desired at a value of independent variable between those for which measurements are made, an operation known as interpolation. Two simple methods to accomplish this objective are linear interpolation, appropriate to only a linear scale, and graphical interpolation, neither of which might be accurate in a general case. Three algebraic methods are polynomial interpolation, in which known points are fitted to a polynomial of sufficient degree, Thiele interpolation, in which points are fitted to a continued fraction of sufficient levels, and a spline function, according to which the measured points in contiguous sets are fitted with segmented polynomials of small degree. For a spline fit, multiple successive points, ordered according to abscissal value, are approximated with a polynomial of selected degree -- commonly three; the point of minimum abscissa is discarded and a further point is added at the other side, and the fit is repeated. These local polynomials are subject to a condition that, at a point at which two meet, the slopes must be identical; use of only a small element from each constrained polynomial yields an impression of a smooth curve. Such a spline fit is useful when many measurements are available to be fitted according to an empirical, rather than a theoretical description, and are differentiable throughout the entire domain of measurement; such splines might be suitable for both smoothing of data and interpolation. If points be subject to random perturbation, a statistical fitting method is applicable; otherwise a spline function might be most reliable. A purely numerical approach involving no explicit algebraic form is based on an interpolation formula for forward differences attributed to Newton and Gregory, as follows:

$$\mathrm{f}(p) = f_0 + p\,\Delta f_0 + \frac{p\,(p-1)\,\Delta^2 f_0}{1\,(2)} + \frac{p\,(p-1)\,(p-2)\,\Delta^3 f_0}{1\,(2)\,(3)} + \ldots$$

in which f_0 is the value of the dependent variable at the measured point nearest the unknown value, p is the fraction of the interval between known points containing the unknown point, Δf_0 is the first difference between values of dependent variable at the two known points enclosing the unknown point, $\Delta^2 f_0$ is the first difference between first differences of values of dependent variables at three known points enclosing the unknown point -- or the second difference, et cetera. According to gaussian interpolation one fits a polynomial of degree n to $n + 1$ points, which we

illustrate with a cubic polynomial to fit four points (x_j, y_j) for $j = 0 .. 3$

$$y = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)} y_0 + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)} y_1$$
$$+ \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)} y_2 + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_3)} y_3$$

This equation fulfills a condition of being a cubic equation that passes the specified points: for instance, when $x = x_2$, only the third term differs from identically zero and has identically a value y_2. Interpolation, with either Newton's or Gauss's method, is unwise for an order greater than cubic because the process might become unstable, yielding sinuous and unrealistic curves, particularly if errors or uncertainties impose an irregular or unsmooth progression of data points; in such a case interpolation with a spline formula, whereby continuity of derivatives is preserved at each point as well as an exact fit there, except zero curvature at the ends of the curve, is likely to prove more satisfactory. For a natural spline, as coded in *Maple*'s procedures, the derivatives at the end points are set to zero, rather than being set to the derivatives in the adjacent intervals. Linear interpolation is practical also in multiple dimensions and can substantially decrease duration of calculations when a function of multiple variables must be evaluated many times; interpolation of greater order, such as parabolic, is also practicable for multidimensional tables, but linear extrapolation generally suffices.

A recurrence relation or difference equation is an equation that expresses one term of a sequence or series as a formula or function involving preceding members of that sequence, of form $x_{k+j} = f(k, x_k, ..., x_{k+j-1})$, thereby providing a recursive definition of that sequence; the number of terms $x_k, ..., x_{k+j-1}$ is the order of this recurrence relation that requires the same number of initial values to define all members of that sequence or series. In some cases, such a relation is solvable to yield an explicit expression for an arbitrary member of that sequence. Such relations arise also in the solution of differential equations; cf section group 7.2.

A recurrence relation might be the outcome of a practice of mathematical induction, according to which, for each natural number n, there holds a proposition P such that the base clause P_1 is true for $n = 1$ and, for a greater integer k, if P_k be true, the proposition is true for P_{k+1}; the conclusion is that the proposition is true for all integers.

An argument of a power series, as illustrated above, might be a simple variable x or an expression, called a functional, involving an elementary function. A functional is a function of which its domain comprises functions in a set; $\cos(n\,x)$ is hence a functional of a function $n\,x$. A series of alternative form such as a Fourier series that can represent a periodic function $F(x)$ of repeating length L, of form for a cosine series,

$$F(x) = a_0 + a_1 \cos\left(\frac{\pi x}{L}\right) + a_2 \cos\left(\frac{2\pi x}{L}\right) + a_3 \cos\left(\frac{3\pi x}{L}\right) + ...$$

involves not a variable to various powers but cosine functions, of general form $\cos\left(\frac{n\pi x}{L}\right)$ in which coefficients n of the variable in the functional take values of successive integers. A Fourier sine series contains sine functions instead of cosine, as above, whereas the most general Fourier

series contains both cosine and sine terms, or exponential terms because Euler's identity

$$e^{(i\theta)} = \cos(\theta) + i \sin(\theta),$$

in which $i = \sqrt{-1}$, connects exponential and trigonometric functions; a Fourier series is thus equivalently expressed as

$$F(x) = \sum_{k=-\infty}^{\infty} c_k \, e^{\left(\frac{i k x}{L}\right)}$$

If $F(x)$ be a real function, factors of sine and cosine in each term of a trigonometric Fourier series are also real, but factors c_k in an exponential Fourier series are generally complex. When a summation becomes replaced by integration, the result might be an integral transform, such as a Fourier transform or a Laplace transform.

In a functional series, a factor of a coefficient in each term that contains an independent variable is called a basis function, which might be a simple variable to a power, such as x^k in a polynomial or Maclaurin's series or $(x-h)^k$ in a Taylor's series with expansion centre h,

$$a_0 + a_1 (x-h) + a_2 (x-h)^2 + a_3 (x-h)^3 + ...$$

with $h = 0$ for a Maclaurin's series but $h \neq 0$ for a Taylor's series, or a functional of that variable, such as $e^{(kx)}$ or $\cos(kx)$. The Taylor's series as an expansion of a formula as a power series, for instance in x, has a positive radius R of convergence, which means that the series converges absolutely for $|x| < R$ with $R > 0$. Every power series $\sum_n c_n x^n$ has a radius of convergence: for $R = 0$, the series converges for only $x = 0$; for $R \to \infty$, the power series converges absolutely for all x, whereas for $0 < R < \infty$ the series converges absolutely for $|x| < R$, diverges for $|x| > R$, and might converge for $|x| = R$. Although requiring no particular theoretical derivation, a power series might be generally useful within a finite range of its argument, but, for purpose of extrapolation beyond a range in which it is defined, it is unreliable because of likely rapid divergence.

convergence of series

Among tests to determine whether an infinite series converges are the following:

- if another series, obtained on taking an absolute value or magnitude of each term in an infinite series of interest, converge, the original infinite series is absolutely convergent;

- if an infinite series with only positive terms converge, a finite series of the same terms is convergent;

- if in a series each term have a magnitude smaller than that of a corresponding term in another series that is known to converge, the former series is convergent; if in a series each term have a larger magnitude than a corresponding term in another series that is known to diverge, the former series is divergent;

- a series with terms of alternating signs is conditionally convergent if successive terms approach zero and if magnitudes of successive terms decrease monotonically; if the same series with an absolute value or modulus of each term also converge, the original series is absolutely convergent; the terms of an absolutely convergent series might be arranged in any order without

affecting the convergence, but, depending on the ordering of terms of a conditionally convergent series, the series might converge or diverge;

- if successive terms of a partial sum approach a limit other than zero or approach no limit, the series diverges;

- for a series with only positive terms, for a limiting ratio of successive terms a_j and a_{j+1},

$$r = \lim_{j \to \infty} \frac{a_{j+1}}{a_j} < 1$$

the series converges; if this ratio $r > 1$, the series diverges; if $r = 1$, this test is inconclusive -- the series might converge or diverge; if the ratio approach no limit but not increase without bound, this test is also inconclusive;

- for an infinite series comprising a sum of terms a_j, if the limit of the jth root of absolute value of a_j according to

$$\lim_{j \to \infty} \left| a_j \right|^{\left(\frac{1}{j} \right)} < 1$$

this series is absolutely convergent; if this limit > 1, the series is divergent; if this limit $= 1$ the series might be absolutely convergent, conditionally convergent or divergent;

- another test involving an improper integral is presented in section 4.405.

Within a domain of convergence, power series can be added or multiplied, terms might be permuted, and one can differentiate or integrate term-wise as desired; upon differentiation or integration the domain of convergence does not alter.

For a power series in x such as $\sum_{j=0}^{} c_j x^j$ or in $(x - a)$ such as $\sum_{j=0}^{} c_j (x - a)^j$ with constant a, the set of values of x for which that power series converges is called its interval of convergence, which might become evaluated with the ratio test supplemented with other tests applied at the bounds of the interval.

For two arbitrary sequences $\{a_j\}$ and $\{b_j\}$ and with $\sum_{j=1}^{n} a_j = A_n$, according to Abel's formula for partial summation a further sum is

$$\sum_{j=m}^{n} a_j b_j = \left(\sum_{j=m}^{n} A_j (b_j - b_{j+1}) \right) + A_n b_{n+1} - A_{m-1} b_m$$

A series that is not rapidly convergent for all values of its expansion variable has dubious value to represent a formula or function, and ought generally to be avoided. A Fourier series in x is uniformly convergent for all real values of x.

As a strategy to test a series for convergence or divergence, one might classify a series according to its form.

- A p series of form $\sum \dfrac{1}{k^p}$ is convergent for $p > 1$ and divergent for $p < 1$.

- A geometric series of form $\sum a\, r^n$ is convergent for $|r| < 1$ and divergent for $|r| > 1$; some manipulation might be applicable to convert a series to this form.

- For a series of form resembling a p or geometric series, a comparison test is applicable for a series with positive terms, or for $\sum |a_j|$ when negative terms also occur.

- When $\lim\limits_{k \to \infty} a_k \neq 0$, a test for divergence becomes applicable.

- For a series of form $\sum (-1)^k a_j$ with $k = j \pm 1$, a test for an alternating series is applicable.

- For a series containing a factorial quantity or product including a constant raised to power n, the ratio test is applicable, but, when $\left| \dfrac{a_{k+1}}{a_k} \right|$ --> 1 as $k \to \infty$ for a p series or rational or algebraic formulae of k, the ratio test is inapplicable.

- For a term of form $a_k = b_k{}^k$, the root test might be appllicable.

- For a term of form $a_j = \mathrm{f}(j)$ with $\displaystyle\int_1^\infty \mathrm{f}(x)\, dx$ readily evaluated, the integral test becomes applicable; cf. section 4.405.

A [Pade](#) approximant is a rational formula $\dfrac{P_n(x)}{Q_m(x)}$ of a particular type whereby the coefficients in [polynomials](#) $P_n(x) = \displaystyle\sum_{k=0}^{n} a_k\, x^k$ of [degree](#) n in the [numerator](#) and $Q_m(x) = \displaystyle\sum_{k=0}^{m} b_k\, x^k$ of degree m in the [denominator](#), hence $n + m + 1$ parameters in total, are chosen to reproduce exactly the values of derivatives of the fitted formula $\mathrm{f}(x)$ up to a particular order; commonly $b_0 = 1$. To calculate coefficients a_k and b_k, one expands first formula $\mathrm{f}(x)$, if type other than polynomial, in a [Taylor series](#), typically of order $n + m$, then solves the identity $f_{n+m}(x)\, Q_m(x) = P_n(x)$ for coefficients a_k, $k = 0 \,..\, n$ and b_k, $k = 1 \,..\, m$. The advantages of a Pade approximant over a mere polynomial are that, whereas a polynomial must diverge eventually outside a domain of definition from either experimental data or theoretical form, this rational function formula might be constrained to obey limiting or [asymptotic](#) properties through the choice of n and m.

function of a complex variable

Analogous to the abstraction of a [real number](#) to an [algebraic](#) [variable](#) such as x, we abstract the notion of a [complex number](#) $a + b\, i$ to a [complex variable](#), generally denoted $z = x + i\, y$, with its [real](#) and [imaginary](#) parts represented by real variables x and y, respectively. A [calculus](#) based on complex variables yields a [branch](#) of [analysis](#) called [complex analysis](#), which has applications in diverse areas of science and engineering: in physics, complex numbers serve to describe the

behaviour of an electromagnetic field; for atomic systems complex numbers and complex functions play a central role.

A complex equation is generally equivalent to two real equations. Replacing a real variable x in a function f(x) by $z = x + i\,y$, with $i = \sqrt{-1}$, to form f(z) creates a function of a complex variable. Assuming a <u>function</u> f(x) of a single <u>real</u> <u>variable</u> x, and denoting that function y such that $y = $ f(x), for various real values assigned to x we plot specific points in plane xy to obtain a <u>graph</u> of that function. For a <u>complex</u> variable of form $x + i\,y$, which we conventionally denote as z so that $z = x + i\,y$, a function of z conventionally denoted $w = $ f(z) is in general also complex; z then signifies a complex <u>independent variable</u> and w signifies a corresponding complex <u>dependent variable</u>. For example, with f(z) $= z^2$, expansion yields f(z) $= x^2 - y^2 + i\,2\,x\,y$. Although we can plot a point z in an Argand diagram for any particular numerical values of x and y measuring x along the real axis and y along the imaginary axis, we can not plot directly f(z) with x and y because f(z) has in general both real and imaginary parts, requiring four dimensions in total. We hence represent values of f(z) on a separate plane; with $w = $ f(z) having real part u and imaginary part v as in $w = u + i\,v$; we thus obtain two complex variables -- $z = x + i\,y$ and $w = $ f(z) $= u + i\,v$. In contrast, w(x, y) $= x^2 + y^2$, which is equal to $z\,\bar{z}$, is not a legitimate complex variable w(z) because it has no imaginary part; $\bar{z}$ is considered to be not a function of z. To derive a general condition for w(x, y) $= $ w(z), we express x and y in terms of z and $\bar{z}$ as

$$x = \frac{1}{2}(z + \bar{z}), \qquad y = \frac{1}{2}(z - \bar{z})$$

An arbitrary function in $w = $ f(z) $= $ u(x, y) $+ i\,$v(x, y) hence becomes expressible as $w = $ w($z, \bar{z}$). A condition that $w = $ w(z) with no dependence on $\bar{z}$ implies that the derivative of w with respect to $\bar{z}$ be zero, which yields the <u>Cauchy-Riemann equations</u>,

$$\frac{\partial}{\partial x}u = \frac{\partial}{\partial y}v, \quad \frac{\partial}{\partial y}u = -\left(\frac{\partial}{\partial x}v\right)$$

which apply to the real and imaginary parts of function w(x, y) for w to be a function of complex variable z. If u and v have continuous partial derivatives in some region with respect to x and y, w(z) is there an <u>analytic function</u> of z, or a <u>holomorphic</u> function for distinction from a <u>real</u> analytic function.

A point P with coordinates (x, y) in plane z becomes transformed into a corresponding point P' in plane w through the <u>transformation</u> <u>relation</u> or transformation function $w = $ f(z); the location P' depends on both the initial location P and the transformation relation, according to which operation, called the <u>mapping</u> of P onto P', P' is generated as the <u>image</u> of P. A <u>fixed point</u> of a transformation is a point that remains <u>invariant</u>. A segment of a straight line in plane z, defined between two specified points, becomes mapped to a corresponding segment of a straight line in plane w only if the transformation is a linear relation, of form $w = $ f(z) $= a\,z + b$. A transformation of form $w = a\,z + b$ with parameters a, b real or complex numbers a few types:

- for $a = 1$ and complex b, the result of transformation of a straight line is a <u>translation</u> of that straight line, by b;

- for a real and $b = 0$, the result of transformation of a straight line is a magnification for $a > 1$ or compression for $a < 1$ of that straight line, by $|a|$;

- for a complex and $b = 0$, the result of transformation of a straight line is a [rotation](#) of that straight line, by the argument of a or [argument](#)(a);

- for various combinations of the above cases, the result of transformation of a straight line is an appropriate combination of translation, magnification or compression, and rotation of that straight line.

To evaluate the effect of a linear transformation, we express each factor in a [polar](#) [representation](#): for $w = a\,z$, $a = \alpha\,e^{(i\,\phi)}$ and $z = r\,e^{(i\,\theta)}$; the product $w = a\,z = \alpha\,e^{(i\,\phi)}\,r\,e^{(i\,\theta)} = \alpha\,r\,e^{(i\,(\theta + \phi))}$. For $\alpha > 1$, radius r becomes dilated to $\alpha\,r$, whereas for $\alpha < 1$, radius r becomes contracted in the same way. This transformation rotates point z by angle ϕ about the origin.

For a non-linear transformation, the result depends on the nature of that transformation and whether the original function in plane z passes the origin. For a transformation z^n, a length r becomes r^n and the angle becomes $n\,\theta$.

The transformation $\dfrac{1}{z}$ transforms lines in plane z to lines in plane w and circles in plane z to circles in plane w. A line not passing the origin in plane z is transformed into a circle passing the origin in plane w, and a circle passing the origin in plane z becomes a line that does not pass the origin in plane w. A line passing the origin in plane z is transformed into a line through the origin in plane w, whereas a circle that does not pass the origin in plane z is transformed into a circle that does not pass the origin in plane w.

Whereas, according to classical analytic geometry, an equation of a locus in two spatial dimensions is expressible as a relation between x and y, that relation is expressible likewise in terms of a complex number z and its complex conjugate $\bar{z}$. A parametric equation $z = a + b\,t$ plots as a straight line in the complex plane for complex numbers a and b with $b \neq 0$ and parametric variable t taking all real values. Two equations $z = a + b\,t$ and $z = a' + b'\,t$ represent the same line only if $a' - a$ and b' are real multiples of b; the lines are parallel when b' is a real multiple of b and equally [directed](#) for b' a positive multiple of b.

[Conformal](#) mapping from plane z to plane w implies that angles between lines in plane z are preserved in both magnitude and sense of direction in their images in plane w. For a transformation to be conformal requires that $w = \mathrm{f}(z)$ must be a [regular](#) function of z: it must be defined and have single values, and the slope of the curve depicting the function must not be zero at a point of such intersection. For a general linear transformation of form $w = a\,z + b$, because $\dfrac{\partial}{\partial z}\,w = a \neq 0$, a mapping of such a form has no critical point and hence provides a conformal mapping of the entire plane z. For a particular case in which $a = 0$, the mapping $w = z + b$ leaves invariant the shape and size of a curve in plane z and translates that curve in plane w by a distance $|b|$ so that the origin in plane z coincides with point $(-b, 0)$ in plane w. For the complementary

particular case in which $b = 0$, the mapping $w = a\,z$, which becomes in polar form $w = \rho\,r\,\mathbf{e}^{i(\Theta + \theta)}$ from $z = r\,\mathbf{e}^{(i\,\theta)}$ and $a = \rho\,\mathbf{e}^{(i\,\Theta)}$, has the effect of multiplying the modulus of z by a constant factor ρ, thus for dilation with $\rho > 1$ and contraction with $\rho < 1$, and to increase the argument of z by a constant angle Θ, so to rotate about the origin by the same angle Θ. The general linear transformation thus becomes described geometrically as a dilation or contraction, a rotation and a translation in combination. The effect of a mapping $w = z^n$ for integer n is to transform the segment $0 \leq \arg(z) < \dfrac{2\,\pi}{n}$ onto the entire plane w, with a cut along axis u because the origin of plane w is a critical point at which $\dfrac{\partial}{\partial z}\,w$ vanishes. The inversion mapping $w = \dfrac{1}{z}$ of points z with respect to the unit circle is followed by their reflexion in the real axis; such an inversion maps points interior to the unit circle about the origin in plane z onto the exterior of the circle about the origin of plane w. A straight line parallel to axis x or y in plane z becomes under inversion a circle in plane w that passes the origin and has its centre on axis u or v, whereas a straight line not passing the origin in plane z transforms into a circle through the origin in plane w but a straight line passing the origin in plane z becomes also a straight line through the origin in plane w. A mapping of form $w = \dfrac{a\,z + b}{c\,z + d}$ is called a *bilinear* or linear fractional transformation of which a general linear transformation or inversion is a special case; its application maps straight lines and circles into straight lines and circles, but not necessarily in the same order.

Conformal mapping is applicable in a solution of problems involving Laplace's equation in two independent variables, for the conduction of heat, the flow of a fluid or an electrostatic potential.

A computational procedure resembles a simple function in mapping input variables into output variables, but offers great flexibility and scope through its extent being limited by only the ingenuity of the programmer and the resources of the computer on which it runs. Such procedures comprise sequences of repetition and conditional constructs with internal assignments and tests to implement arbitrarily complicated operations on the input data or variables. Language *Maple* itself comprises almost innumerable procedures for arithmetic, algebraic and graphical purposes; a user can extend the capability of the language for particular purposes nearly without limit, apart from the computer resources.

The topics within these chapters 1 and 2 collectively constitute what might be called pre-calculus, even though some differential or integral properties are germane to the discussion of functions of various kinds and of geometry and trigonometry. Many commands and operators within *Maple* that we discuss and apply herein and for which these topics serve as a vehicle for introduction are nevertheless essential for an effective and efficient application of symbolic computation in an attack of diverse problems in chemistry.

\>

— summary of chapter 2

Beyond elementary functions exponential and logarithmic explained in chapter 1, we encounter in this chapter circular and hyperbolic trigonometric functions and their inverses. We discuss also

polynomial and rational formulae and functions according to both algebraic and graphical properties; algebraic properties include the relation between functions and conversions to expressions of other kinds, such as series. To be operational, most functions require an input -- numerical or symbolic -- and yield an output that might be an object of a kind different from that of the input. Elementary functions of types exponential, logarithmic and trigonometric are also transcendental functions. For comparison, a special function, i.e. not elementary, is function Γ that is also a transcendental function. Series of a few kinds are introduced, and animated plots illustrate vividly how properties of these quantities depend on the number of terms in a series. All these quantities provide a basis for symbolic mathematical treatments in succeeding chapters. Numerous functions are available in *Maple* in a list formed here; help on a topic therein is obtained on invoking this hyperlink.

> **? index, function**

Information is available on procedures and programming in *Maple*; help on a particular topic in presented lists is obtained on invoking this hyperlink.

> **? index, procedure**

Apart from an automated mechanism available through smartplot, diverse and powerful plotting facilities are available in *Maple*; we explain plots of diverse types in two and three dimensions, and indicate how one can exploit options of plotting to enhance the impact of a plot on a viewer.

Although most mathematical topics discussed in this chapter might nominally be introduced in secondary school, and most corresponding operations might be conducive to manual execution, our coverage and discussion are generally extensive, although not intended to be comprehensive. Entire courses on complex analysis, number theory, abstract algebra et cetera that are of interest to students specialist in mathematics might have their initial point on this basis. For a student whose interest is primarily chemistry, all this material provides a strong foundation on which to construct calculus, linear algebra and other higher mathematics, according to topics in succeeding chapters for which an application to chemistry in its various branches becomes readily discernible.

>

chapter 3 Differentiation

3.0 overview and principles

Our ability to describe how concentrations of chemical compounds vary with time expands our understanding of chemical systems. Calculus is the mathematical study of change, in the same way that geometry is the study of shape and algebra is the study of operations and their application to solving equations. Although algebra, geometry and trigonometry are useful to describe relations among static quantities, these tools lack concepts appropriate to describe a temporal variation. To describe motion, as a temporal variation of spatial coordinates, or the progress of a chemical reaction, as a temporal variation of the concentration of reactants, the infinitesimal calculus provides additional tools through differentiation and integration; like addition and subtraction, these operations are mutually opposite: what one operation does, the other reverses the effect. Isaac Newton developed calculus from a point of view of a derivative as a rate of change whereas Leibniz developed calculus in terms of differential quantities.

Chemical processes involving a temporal variation of a property of a chemical system are a principal concern of chemists, for instance the variation with time of concentration of reactants and products in chemical reactions or the rate of emission of radiant energy in a photoluminescent process. The temporal rate of variation of concentration of a species attracts particular chemical attention, as such rates are characteristic of the chemical nature of the constituents of a reacting mixture. Such a rate one treats in a quantitative mathematical manner as a derivative of a quantity, such as concentration, with respect to time; an alternative term for a derivative is a *differential quotient*, or perhaps differential coefficient. In this chapter we explain how to perform operations involving derivatives, or differential calculus, involving a single independent variable in either a formula of algebraic form or even purely numerical data. We express formally a fundamental definition of a limit: already introduced informally in a discussion of limiting behaviour of a formula or function under some specified conditions, this concept underpins a definition of not only a derivative but also an integral. We proceed to define and to evaluate derivatives through a basic definition of a limit, and view applications of derivatives with the aid of plots. We introduce differential quantities and employ them in a tangent approximation for chemical applications.

limit

A limit of a function such as f(x) is its value to which an approach becomes increasingly near as its independent variable x approaches a stated value, such as $x = p$. For a real function f(x) there exists a limit L at a point p, at which f(x) be defined, if f(x) satisfy this condition:

for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|$ f(x) $- L | < \varepsilon$ for all x such that $| x - p | < \delta$.

Here ε is a conventional symbol for a small but rigorously positive quantity associated with formula f(x) whenever its independent variable x is within another small but rigorously positive quantity δ of a stated value p. If f(x) be continuous at p, this condition is expressed as

$$\lim_{x \to p} f(x) = f(p)$$

For a value k to be a limit of a formula f(x) as $x \to \infty$, expressed as

$$\lim_{x \to \infty} f(x) = k$$

there must exist a large integer N such that

$$| f(x) - k | < \varepsilon \text{ for all } x > N.$$

A discontinuity occurs at a point, or value of an independent variable, at which a value of a formula differs from its limit as the value of the independent variable approaches that limit, or at which a formula or function is undefined. A singular point, or singularity, is a point at which a curve representing a formula or function lacks an unique smooth tangent; a point at which a curve crosses itself is such a singularity.

- If a function f(x) be defined at a point $x = p$ -- so that $f(x)\big|_{x=p} = f(p)$ exists,

- if a limit, $\lim_{x \to p}$ f(x), exist on approach to p from both greater and smaller values of x, and

- if that limit equal f(p),

$$\text{i.e. } \lim_{x \to p} f(x) = f(p),$$

that function f(x) is continuous at $x = p$. The limit of a sum is a sum of separate limits, a

distributive property,
$$\lim_{x \to p} f(x) \pm g(x) = \lim_{x \to p} f(x) \pm \lim_{x \to p} g(x);$$
the limit of a product is a product of limits of the factors,
$$\lim_{x \to p} f(x) g(x) = (\lim_{x \to p} f(x))(\lim_{x \to p} g(x)),$$
including a special case in which one factor is a constant,
$$\lim_{x \to p} c f(x) = c (\lim_{x \to p} f(x));$$
and the limit of a quotient is a quotient of separate limits,
$$\lim_{x \to p} \frac{f(x)}{g(x)} = \frac{\lim_{x \to p} f(x)}{\lim_{x \to p} g(x)},$$
provided that a limit in a denominator be not zero. If both numerator and denominator have zero limits at $x = p$, the quotient of the limits becomes $\frac{0}{0}$, so undefined, but a limit of this quotient of functions might still exist. The limit of a formula to some power is
$$\lim_{x \to p} f(x)^n = (\lim_{x \to p} f(x))^n,$$
for n some real positive or negative number, and for a logarithm of a formula,
$$\lim_{x \to p} \ln(f(x)) = \ln(\lim_{x \to p} f(x))$$

difference quotient and differential quotient

For a dependent variable y that has a functional dependence $f(x)$ on a single independent variable x, a derivative, conventionally written $\frac{dy}{dx}$, is a limit of a quotient of a difference $(y + \Delta y) - y = f(x + \Delta x) - f(x)$ with an increment Δx in x, as that increment tends to zero:
$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$$
We might equally write a decrement instead of an increment, as in
$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{f(x - \Delta x) - f(x)}{\Delta x}$$
We thus distinguish between an *instantaneous* rate of variation of y with x, evaluated with the derivative on the left side, and a *mean* rate of variation of y with x over some finite interval Δx, as in $\frac{f(x + \Delta x) - f(x)}{\Delta x}$, before evaluating the limit on the right side. For the difference quotient $\frac{\Delta y}{\Delta x}$, the corresponding differential quotient is $\frac{dy}{dx}$ or $\frac{\partial}{\partial x} y$, and differential dy is the differential of y that belongs to differential dx. For a particular point with coordinates (x_0, y_0) on a continuous curve corresponding to formula $y = f(x)$, an increment Δx yields a corresponding change Δy, such that
$$\frac{\Delta y}{\Delta x} = \frac{y - y_0}{x - x_0} = \frac{f(x) - f(x_0)}{x - x_0} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = \tan(\alpha),$$

in which the <u>secant</u> is a straight line passing points (x, y) and (x_0, y_0) and the corresponding difference quotient makes an angle α with axis x. In the limit in which $\Delta x \to 0$, the secant becomes a <u>tangent</u> of which the differential quotient indicates the gradient.

derivative

   <u>Differentiation</u> is a process of evaluating such a derivative of a formula with respect to some variable therein. Differentiation of a <u>first derivative</u> or a derivative of first <u>order</u>, such as above, produces a <u>second derivative</u>, or a derivative of second order, expressed as

$$\text{standard notation:} \qquad \frac{d}{dx}\left(\frac{dy}{dx}\right) = \frac{d^2 y}{d x^2},$$

$$\textit{Maple notation:} \qquad \frac{\partial}{\partial x}\frac{\partial}{\partial x}y = \frac{\partial^2}{\partial x^2}y,$$

and so forth. Here the first part $\dfrac{d}{dx}$, for which *Maple* might deploy letter d in a special font in both numerator and denominator as $\dfrac{\partial}{\partial x}$, is a differential <u>operator</u> implying differentiation, once, of a formula on which it operates, with respect to independent variable x; in this case the formula evaluated from $\dfrac{dy}{dx}$ or $\dfrac{\partial}{\partial x}y$ on which it operates is a first derivative of y with respect to x.

According to SI convention and standard mathematical notation, letter 'd' in those derivatives is written in roman font, not *italic* font like that for the independent or dependent variable, but that notation is impracticable here in this *Maple* worksheet. The latter operation, i.e. differentiating a derivative, yields a second derivative, for which notation $\dfrac{\partial^2}{\partial x^2}$ or $\dfrac{d^2}{dx^2}$ implies analogously differentiation of a formula twice with respect to x; hence operation of $\dfrac{\partial}{\partial x}$ on $\dfrac{\partial}{\partial x}y$ yields $\dfrac{\partial^2}{\partial x \partial x}y$, expressed in *Maple* also as $\dfrac{\partial^2}{\partial x^2}y$, not $\left(\dfrac{dy}{dx}\right)^2$ expressed in *Maple* as $\left(\dfrac{\partial}{\partial x}y\right)^2$. An alternative notation for a first derivative of a formula f(x) with respect to x is f $'(x)$, and for its second derivative is f $''(x)$. As mentioned above, Newton's view of a derivative was a rate of change, implying time to be the independent variable; Newton represented a first derivative such as $\dfrac{dx}{dt}$ with a notation $\dot{x}$, hence with a point above the dependent variable. A geometric definition of a derivative as a tangent to a curve is explored in the sections below.

   As an example of differentiation we extract a first derivative of a simple <u>quadratic</u> formula

$$y = \text{f}(x) = 3\,x^2 + 5\,x + 2$$

With increment Δx in independent variable x and Δy as the corresponding increment in dependent variable y, we form an expression in terms of increments in both x and y,

$$y + \Delta y = \text{f}(x + \Delta x) = 3\,(x + \Delta x)^2 + 5\,(x + \Delta x) + 2$$
$$= 3\,x^2 + 6\,x\,\Delta x + 3\,\Delta x^2 + 5\,x + 5\,\Delta x + 2 ;$$

after expansion, we subtract the original expression $y = f(x)$ to obtain

$$\Delta y = f(x + \Delta x) - f(x) = (3x^2 + 6x\,\Delta x + 3\,\Delta x^2 + 5x + 5\,\Delta x + 2) - (3x^2 + 5x + 2)$$
$$= 6x\,\Delta x + 3\,\Delta x^2 + 5\,\Delta x$$

after cancellation of like terms with positive and negative signs. After dividing both sides by Δx supposed to have a finite magnitude, we obtain

$$\frac{\Delta y}{\Delta x} = 6x + 3\,\Delta x + 5.$$

According to a definition of a derivative, we form a limit of both sides of this equality as $\Delta x \to 0$,

$$\frac{\partial}{\partial x} y = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$
$$= \lim_{\Delta x \to 0} 6x + 3\,\Delta x + 5$$
$$= 6x + 5 ,$$

because the second term $3\,\Delta x$ in a limit on the right side of the equality sign vanishes in the limit as $\Delta x \to 0$. The first derivative of a formula named y,

$$y = f(x) = 3x^2 + 5x + 2$$

with respect to x is hence

$$\frac{\partial}{\partial x} y \equiv f'(x) = 6x + 5 .$$

Three theorems that are important in the theory of differentiation are Rolle's theorem and two theorems on mean values. According to Rolle's theorem, for a formula $f(x)$ that is continuous on a closed interval $[a, b]$, thus including the end points, of x and that has a first derivative everywhere in the open interval $]a, b[$ (thus excluding the end points), if

$$f(x)\big|_{x=a} = f(x)\big|_{x=b}$$

there exists at least one point at $x = c$ at which

$$\left(\frac{d}{dx} f(x)\right)\bigg|_{x=c} = 0.$$

If for a function with formula $f(x)$ that is continuous on interval $[a, b]$ and with formula $f(a)$ at $x = a$ and $f(b)$ at $x = b$ a product $f(a)\,f(b) < 0$, there exists at least one value $x = c$ within that interval for which $f(c) = 0$. That continuous function increases on the same interval if $f(a) < f(b)$, or $\dfrac{d}{dx} f(x) \geq 0$, and decreases on that interval if $f(a) > f(b)$, or $\dfrac{d}{dx} f(x) \leq 0$. For the same conditions applicable to two formulae $f(x)$ and $g(x)$ with respect to the same interval, and if also $\dfrac{d}{dx} g(x) \neq 0$ on the open interval, according to Cauchy's mean-value theorem there exists at least one point at $x = c$ at which in a general case

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{\left(\dfrac{d}{dx} f(x)\right)\bigg|_{x=c}}{\left(\dfrac{d}{dx} g(x)\right)\bigg|_{x=c}}.$$

For a particular case of a mean-value theorem in which $g(x) = x$,

$$\frac{f(b) - f(a)}{b - a} = \left(\frac{d}{dx} f(x)\right)\Bigg|_{x = c}$$

geometric aspects of derivatives

   Besides a formal algebraic definition, a first derivative has a geometric definition as the slope of a tangent to a continuous curve representing $y = f(x)$ at any point along that curve with abscissa x. A stationary point on that curve -- an extremum that is either a maximum or a minimum -- and a point of inflexion are then defined in relation to particular values of first, second and third derivatives. When one (or more) derivative of a formula or function has zero value for a particular value of an independent variable, a point on the corresponding curve displays a special property in a geometric sense. Whereas the first derivative represents a slope of a tangent to a curve, or its gradient, the second derivative represents concavity, which is related in turn to curvature. A point at which a first derivative is zero is a *stationary point* or a critical point; such a point at which a tangent has zero slope can mark the presence of an extremum -- a maximum or minimum that is a turning point of a curve -- which might be local or global, or a point of inflexion; likewise, for a second derivative to be zero might, subject to further tests, indicate a point of inflexion, at which a tangent might or might not have zero slope. Near a local maximum of a curve describing a formula or function, *concavity* is defined to be concave *downward*, or in a vicinity of a local minimum, *concave upward*. *Curvature* measures a rate at which the inclination of a tangent to a curve varies relative to a length of arc, which is a segment of a curve; curvature is positive for a curve that is concave upward, or negative for a curve that is concave downward. To test these effects, one evaluates a second derivative of an expression at a chosen turning point: a positive value of the second derivative indicates a local or global minimum whereas a negative value indicates a local or global maximum; a zero value might indicate a point of inflexion, at which concavity changes sign along the curve across the point. A further test for a point of inflexion is to verify that the third derivative has a finite value at a location under consideration. On passage through a turning point from left to right, the slope of a tangent line changes from positive to negative for a maximum, and from negative to positive for a minimum; the slope of a tangent line retains its sign at a point of inflexion of a curve, but that tangent line crosses that curve.

   The algebraic derivative of a formula $f(x)$ with respect to its single independent variable x provides information about the geometric properties of the associated curve in a cartesian plane; because these geometric concepts related to derivatives are important, we rephrase them here. For a formula to assume ascending values on an interval I between x_1 and x_2 signifies that $f(x_1) < f(x_2)$; vice versa, for a formula to assume descending values on the same interval signifies that $f(x_1) > f(x_2)$. In the former case the derivative $\frac{d}{dx} f(x) \geq 0$ on that interval and is not 0 everywhere thereon, whereas in the latter case $\frac{d}{dx} f(x) \leq 0$ and is not 0 everywhere thereon.

Moreover, in the former case the slope m of the curve representing $f(x)$ is likewise positive, or at least non-negative, on that interval, but negative, or at least non-positive, in the latter case. A stationary point or critical point of a curve representing a formula is a point in the domain of that

formula at which the first derivative $\dfrac{d}{dx}\text{f}(x)$ is either zero or does not exist; at such a stationary point, that curve has either at least a local or relative maximum or minimum, either of which is at least a local extremum. If there exist no such extremum of lesser value of f(x), that extremum is a global minimum, or for no other extremum of greater value of f(x) that extremum is a global maximum; a formula f(x) that be continuous on a closed interval must attain an absolute maximum and an absolute minimum within that interval. At such a relative or global maximum, the first derivative and the slope of the tangent line of the curve alter from positive to negative as x varies through that stationary point and the curve is concave downward, whereas at such a relative or global minimum the first derivative and the slope of the tangent of the curve alter from negative to positive as x varies through that stationary point and the curve is concave upward.   At such a stationary point at which the first derivative $\dfrac{d}{dx}\text{f}(x)$ is zero, the second derivative $\dfrac{d^2}{dx^2}\text{f}(x)$ is negative if the formula and curve have there a local or global maximum, or positive if the formula and curve have there a local or global minimum. An extremum of f(x) might also occur at a point at which $\dfrac{d}{dx}\text{f}(x)$ or $\dfrac{d^2}{dx^2}\text{f}(x)$ does not exist. Between intervals of opposite concavity there must exist a point of inflexion at which the second derivative $\dfrac{d^2}{dx^2}\text{f}(x)$, if it exist, is zero and changes sign as x increases through that point; at such a point of inflexion the tangent line crosses the curve, with this tangent line either parallel to the abscissal axis for a horizontal point of inflexion or parallel to the ordinate axis for a vertical point of inflexion. Application of these conditions on the zero values of first and second derivatives of a formula enables one to locate absolute extrema and stationary points on a closed interval.

If x and y are specified as parametric functions of t, provided that $\dfrac{\partial}{\partial t}x \neq 0$,

$$\frac{\partial}{\partial x}y = \frac{\dfrac{\partial}{\partial t}y}{\dfrac{\partial}{\partial t}x}$$

$$\frac{\partial^2}{\partial x^2}y = \frac{\dfrac{\partial}{\partial t}\left(\dfrac{\partial}{\partial x}y\right)}{\dfrac{\partial}{\partial t}x}$$

A formula is differentiable only if its variable can assume values in a continuous range. If a complicated formula defy rearrangement such that a variable considered dependent be isolable on one side of an equality, implicit differentiation is practicable whereby differential operator $\dfrac{\partial}{\partial x}$ is applicable to each term of this formula. A more abstract and general differential operator D is

applicable to a function without explicit reference to a specific independent variable; for instance,
$$D(\sin) = \cos$$
Logarithmic differentiation that implies evaluating derivatives after taking logarithms of each side of an identity is most applicable to expressions involving products or quotients; a logarithmic

derivative is the derivative of a logarithm of a particular formula, $\dfrac{d}{dx}\ln(f(x)) = \dfrac{\dfrac{d}{dx}f(x)}{f(x)}$ according

to the chain rule. Both these operations that might be useful in manual work are superfluous when a symbolic processor is applied; implicit differentiation might be applicable for this purpose.

differentiation of complex formula

In forming derivatives above we implicitly assume real numbers and real formulae. By analogy with a real function of two real variables, a complex formula $w = f(z) = f(x + iy)$ is continuous at $z = z_0$ if

$$\lim_{z \to z_0} f(z) = w_0 \quad \text{and} \quad f(z_0) = w_0 .$$

Like a real function $f(x)$, a complex function $f(z)$ is differentiable, and has a derivative

$$\left(\frac{\partial}{\partial z} w\right)\Big|_{z=z_0} = \left(\frac{d}{dz} f(z)\right)\Big|_{z=z_0}$$ at a point z_0 *provided that* function $f(z)$ is defined in the vicinity of z_0

and that the limiting value of $\dfrac{\Delta w}{\Delta z}$ as $\Delta z \to 0$,

$$\text{i.e.} \quad \frac{dw}{dz}\Big|_{z=z_0} = \lim_{\Delta z \to 0} \frac{f(z_0 + \Delta z) - f(z_0)}{\Delta z} ,$$

exists as a limit in the complex plane independent of the direction of approach to that point, wherever that limit exist. To illustrate this property we consider a derivative of z and of its complex conjugate $\bar{z}$; either derivative exists only if the value of a limit be independent of the direction of approach to that limit. First for $z = x + iy$, we apply the definition of a derivative as a limit at point z_0, and take $w = f(z)$; then $\Delta w = f(z) - f(z_0)$, and

$$\frac{dw}{dz} = \lim_{z \to z_0} \frac{\Delta w}{\Delta z}$$

For z we approach a point $(0,0)$ along axes x and y. For axis x we apply the limit,

$$\frac{z + \Delta z - z}{\Delta z} = \frac{x + iy + \Delta x - (x + iy)}{\Delta x} = \frac{\Delta x}{\Delta x} = 1$$

For axis y we apply the limit analogously,

$$\frac{z + \Delta z - z}{\Delta z} = \frac{x + iy + i\Delta y - (x + iy)}{\Delta y} = \frac{\Delta y}{\Delta y} = 1$$

Accordingly we obtain the same value of the limit on the approach to the origin along either axis. When we apply these limits for $\bar{z}$ we obtain for approach to the origin along axis x,

$$\frac{(\bar{z} + \Delta \bar{z}) - \bar{z}}{\Delta z} = \frac{(x + iy + \Delta x) - (x + iy)}{\Delta x} = \frac{x - iy + \Delta x - (x - iy)}{\Delta x} = \frac{\Delta x}{\Delta x} = 1$$

but for axis y we obtain

$$\frac{\overline{(z + \Delta z)} - \overline{z}}{\Delta z} = \frac{\overline{(x + iy + i\,\Delta y)} - \overline{(x + iy)}}{\Delta y} = \frac{x - iy - i\,\Delta y - (x - iy)}{\overline{\Delta y}} = \frac{-\Delta y}{\Delta y} = -1$$

Because these limits are not identical, the derivative of $\overline{z}$ does not exist, even though the function is continuous everywhere.

Rules of differentiation of a function of a complex variable are the same as for a real variable, except for possible requirements involving the same branch of that function. There exist complex functions for which no derivative exists, even for a quantity as simple as $\overline{z}$, as derived above, because that limit above depends on the direction of approach. *Maple* provides a derivative with respect to only an atomic quantity -- one for which there is no preceding assignment -- that is also a name or a symbolic quantity; for this reason, whether a quantity z be real or complex is immaterial as long as there is no preceding assignment to that quantity. If one must differentiate with respect to a quantity that is not atomic, an appropriate substitution must be made before differentiation and then the reverse substitution afterward.

The combinations $x + y$ and $x - y$ are not independent, but z and $\overline{z}$ are not independent, according to the following argument: z represents the radius vector to some point, but $\overline{z}$ is not determined before the direction of axis x is specified; if both z and $\overline{z}$ be given, the direction of axis x is determined as a line bisecting the angle between vectors z and $\overline{z}$, from which x and y can be found.

numerical differentiation

For numerical data rather than an algebraic formula or function, either a derivative be estimated through ratios of finite differences, or an algebraic formula might be fitted to regularly varying data and subsequently differentiated. In the former case, differences in the following set might serve for the calculation of numerical approximations to derivatives. For points in a set $\{(x_i, y_i)\}$ with x_i at equal intervals Δx, the first difference for point i is

$$\Delta y_i = y_{i+1} - y_i,$$

the second difference is

$$\Delta^2 y_i = \Delta y_{i+1} - \Delta y_i = y_{i+2} - 2 y_{i+1} + y_i,$$

the third difference is

$$\Delta^3 y_i = \Delta^2 y_{i+1} - \Delta^2 y_i$$

et cetera; this scheme for point i involves only points $i + 1$, $i + 2$, ..., but other equivalent schemes are devised for points on both sides of each point in turn.

The simplest method to obtain a numerical first derivative is hence to apply Newton's difference quotient,

$$\frac{dy}{dx} = \frac{f(x + \Delta x) - f(x)}{\Delta x},$$

in which $=$ implies an approximate equality, and x and $x + \Delta x$ are adjacent points for which the corresponding values $f(x)$ and $f(x + \Delta x)$ are known; the slope of this secant line differs from the slope of the tangent line by an amount that is approximately proportional to Δx. As $\Delta x \rightarrow 0$, the slope of the secant line approaches the slope of the tangent line, and the accuracy of the

approximation increases. An alternative formula involving two known points is this approximate equality

$$\frac{dy}{dx} = \frac{f(x + \Delta x) - f(x - \Delta x)}{2 \Delta x} ;$$

in this case the errors of first order cancel, and the slope of the secant lines differ from the slope of the tangent line by an amount that is approximately proportional to Δx^2; for this reason, for small values of Δx, this approximation to the slope of the tangent line is more accurate than the preceding expression that involves an increment at only one side of a given point. Although the derivative is being evaluated at x, the value of $f(x)$ is not involved. A method of greater order involves five points in this approximate equality,

$$\frac{dy}{dx} = \frac{-f(x + 2 \Delta x) + 8 f(x + \Delta x) - 8 f(x - \Delta x) + f(x - 2 \Delta x)}{12 \Delta x} ,$$

in which the error is proportional to $\dfrac{\Delta x^4}{30}$.

Based on a formula for interpolation attributed to Gregory and Newton, first derivative $\dfrac{dy}{dx}$ at point i with coordinates (x_i, y_i) is also evaluated from

$$\left.\frac{dy}{dx}\right|_{x = x_i} = \frac{1}{\Delta x} (\Delta y_i - \frac{1}{2} \Delta^2 y_i + \frac{1}{3} \Delta^3 y_i - \dots)$$

and second derivative $\dfrac{\partial^2}{\partial x^2} y$ at point i is evaluated from

$$\left.\left(\frac{\partial^2}{\partial x^2} y\right)\right|_{x = x_i} = \frac{1}{(\Delta x)^2} (\Delta^2 y_i - \Delta^3 y_i + \frac{11}{12} \Delta^4 y_i - \dots)$$

Although these operations might be automated through the use of a spreadsheet, likely a value of the desired first or second derivative might have superior accuracy from the use of a fitted spline formula and algebraic differentiation.

Taylor series

A Taylor's series comprises a sum of terms in each of which a derivative of successively increasing order evaluated at a particular point is coefficient of an expression or variable with an exponent of the same degree:

$$f(x) = \frac{f(x)|_{x = x_0}}{0!} + \frac{\left.\dfrac{d\,f(x)}{dx}\right|_{x = x_0}(x - x_0)}{1!} + \frac{\left.\dfrac{d^2 f(x)}{d\,x^2}\right|_{x = x_0}(x - x_0)^2}{2!}$$

$$+ \frac{\left.\dfrac{d^3 f(x)}{d\,x^3}\right|_{x = x_0}(x - x_0)^3}{3!} + \frac{\left.\dfrac{d^4 f(x)}{d\,x^4}\right|_{x = x_0}(x - x_0)^4}{4!} + \dots ;$$

if $x_0 = 0$, the series is a Maclaurin's series. The radius r of convergence of a Taylor's series is the distance from the expansion centre x_0 to the nearest singularity of $f(x)$, and the circle of

convergence is a circle of radius r of convergence centred at $x = x_0$; such a circle exists in the complex plane. For instance, for a function $f(x) = \dfrac{1}{1 + x^2}$ there exist poles at $x = \pm i$, and a Taylor series for this function converges for only $|x| < 1$. These series are expressible also in terms of operator D.

Two geometric properties of a curve at a particular point (x_0, y_0) are a tangent line and a normal line. For a tangent line defined with a formula $y = m\,x + b$, its slope m is just the first derivative, so $m = \dfrac{dy}{dx}$, of the formula for the curve evaluated at that point; its intercept b on the ordinate axis is then evaluated on substituting coordinates x_0 and y_0 and that value of m into a defining relation, $y = m\,x + b$. For a normal line defined likewise as $y = m\,x + b$, its slope m is just the negative reciprocal of the first derivative of the formula for the curve, so $m = \dfrac{-1}{\dfrac{\partial}{\partial x} y}$; its intercept b on the ordinate axis is then again evaluated on substituting coordinates x_0 and y_0 and that value of m into a defining equation; if $\dfrac{\partial}{\partial x} y = 0$ at that point (x_0, y_0), the normal line is parallel to axis y and has equation $x = x_0$.

root of equation with Newton's algorithm

Here we apply calculation of derivatives to a general problem -- to find a real root of an equation in one unknown quantity of form $f(x) = 0$. For this purpose Gauss based an approach on Newton's algorithm; at each stage of an iterative process, an approximate value of x from the preceding stage becomes its value to be corrected with a ratio $\dfrac{f(x)}{\dfrac{d}{dx} f(x)}$:

$$x_{j+1} = x_j - \frac{f(x_j)}{\left(\dfrac{d}{dx} f(x) \right)\Big|_{x = x_j}} .$$

Convergence to a solution x_c occurs as $j \to \infty$ providing that

- that $f(x)$ has two continuous derivatives,

- that $\left(\dfrac{d}{dx} f(x) \right)\Big|_{x = x_j} \neq 0$, and

- that an initial estimate x_0 is sufficiently near x_c.

In a particular case that the first derivative $\left(\dfrac{d}{dx} f(x) \right)\Big|_{x = x_j}$ at a particular point x_j is small, convergence might be difficult to attain. Under that particular condition an alternative method might be practicable, involving this iterative formula with a second derivative first,

$$x_{j+1} = x_j - \sqrt{\dfrac{-2\, f(x_j)}{\left(\dfrac{d^2}{dx^2} f(x)\right)\Big|_{x=x_j}}}$$

and then reverting to the above formula when near x_{j+1} the first derivative becomes larger.

For a product of two functions $f(x)$ and $g(x)$, derivatives have the following form.

$$\frac{d}{dx}(f(x)\, g(x)) = \left(\frac{d}{dx} f(x)\right) g(x) + f(x)\left(\frac{d}{dx} g(x)\right)$$

$$\frac{d^2}{dx^2}(f(x)\, g(x)) = \left(\frac{d^2}{dx^2} f(x)\right) g(x) + 2\left(\frac{d}{dx} f(x)\right)\left(\frac{d}{dx} g(x)\right) + f(x)\left(\frac{d^2}{dx^2} g(x)\right)$$

with derivatives of greater orders conforming to a binomial series.

differential

Following <u>Leibniz</u>, before this point we regard *Maple*'s notation $\dfrac{\partial}{\partial x} y$ to denote a single entity implying a derivative of an <u>expression</u> or formula named y with respect to its single independent variable x, according to a definition above. Another view of this derivative is a ratio of two separate quantities, a <u>differential</u> dy of dependent variable y and a differential dx of independent variable x, of which this <u>ratio</u> constitutes a derivative; neither differential quantity implies necessarily an <u>infinitesimal</u> change. An increment in independent variable x is denoted Δx, which is expressible as dx; this increment hence becomes known as the differential of x. A differential dy of a given function, such as $f(x)$, is expressed as a product of a derivative of that function $f'(x)$ and a corresponding increment dx of independent variable x.

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

Although dx is thus an increment in x, dy is not in general the corresponding increment in y, expressed as Δy and calculated as

$$\Delta y = f(x + \Delta x) - f(x)\,;$$

only for a linear formula is this derivative $\dfrac{dy}{dx}$ or $\dfrac{\partial}{\partial x} y$ at any point equal to a ratio $\dfrac{\Delta y}{\Delta x}$ with an arbitrary variation Δx of independent variable x in the denominator and thus dy equal to a corresponding <u>variation</u> Δy in the numerator. A tangent to a smooth curve at a particular point is by definition a <u>straight</u> line; the slope of this tangent is a derivative $\left(\dfrac{dy}{dx}\right)_t$ of a linear relation defining that straight line, and is by definition equal to a derivative $\left(\dfrac{dy}{dx}\right)_c$ of a formula defining that curve at that point. That derivative $\left(\dfrac{dy}{dx}\right)_t$ pertaining to the tangent line is equal to a ratio $\dfrac{\Delta y}{\Delta x}$ for variation Δx of arbitrary magnitude and a corresponding variation Δy for that tangent line; the differential dx is exactly equal to the variation Δx, and to the extent that the curve deviates from

that tangent line a corresponding differential dy for the curve deviates from Δy. The *tangent approximation* thus has as its basis the use of a derivative $\left(\dfrac{dy}{dx}\right)_t$, or the slope of a line tangent to a curve at a particular point, and its associated differentials dy and dx, as an approximation of the derivative $\left(\dfrac{dy}{dx}\right)_c$, or the slope of that curve, and the differential dy associated with the tangent line to estimate the true increment $\Delta\,y$ at some distance $\Delta\,x$ away from that point. At a critical point or extremum, $\dfrac{dy}{dx} = 0$, so that $dy = 0$; at such a point one can not use a differential dy to approximate a change $\Delta\,y$ of formula $y = \mathrm{f}(x)$.

   With a relation between independent and dependent variables known in symbolic form, this tangent approximation is generally superfluous because, with a symbolic processor such as *Maple*, one can almost invariably obtain readily an exact variation through a derivative expressed in terms of <u>symbol</u>s, but with numerical data this approximation might have some applicability.

   Differentials pervade physical chemistry, and chemical kinetics in particular. For example, consider a dependence of concentration $[A]$ on duration t of a reaction of first kinetic order according to $[A] = a\,\mathrm{e}^{(-k\,t)}$, in which a is the initial concentration of reactant and k is a rate coefficient. According to that definition, we differentiate to obtain

$$\mathrm{d}\,[A] = -k\,a\,\mathrm{e}^{(-k\,t)}\,\mathrm{d}t,$$

For given t and dt, we can estimate a depletion d $[A]$ of A. In thermodynamics of phase transitions, the Clapeyron equation $\dfrac{\partial}{\partial T}P = \dfrac{\Delta H}{T\,\Delta V}$ describes the effect of pressure on the temperature of a transition in terms of the molar volume change $\Delta\,V$ and the molar enthalpy change $\Delta\,H$ that accompany that transition. Differentials of variables that describe a physical state of a chemical compound are important in thermodynamics; as experimental conditions generally involve multiple independent variables, we consider these applications in chapter 5.

   Differentation constitutes a powerful tool for the solution of problems in diverse fields; many such applications involve rates, such as of chemical reactions, and maxima and minima. The great utility of a symbolic processor, such as *Maple*, is that, to differentiate a complicated function, we need not concern ourselves how differentiation is achieved or with deployment of conventional rules for a product, quotient, chain rule or other particular procedure: three commands -- **diff** for explicit differentiation with respect to a specified variable, **implicitdiff** for implicit differentiation either when a dependent variable is not readily isolable or for formulae expressed parametrically, and **D** as a differential operator -- suffice to activate *Maple* to implement the most appropriate approach for any algebraic expression.

summary of chapter 3

   In this chapter we investigate the nature of a limit or an asymptotic approach to a value, and how a derivative in calculus is based on such a limiting behaviour of a ratio of two small changes. Only a few commands provide tremendous power for operations in differential calculus. We apply differentiation to various formulae and functions, and discover complications caused by

discontinuous conditions. A derivative viewed as a ratio of differential quantities constitutes a simple method to estimate a variation in a value of a function when an independent variable is incremented a given amount, although this method is generally superfluous when one has access to powerful software for symbolic computation that readily enables an exact calculation. Chemical application of these principles arises in estimation of error when one processes experimental data through use of a formula, as demonstrated in examples in chemical kinetics and thermodynamics.

chapter 4 Integration

4.0 overview and principles

In chemistry an integral is widely applicable to transform one formula or function, typically associated with a measured property, into another formula or function. Thermodynamic functions given at a particular temperature considered a standard temperature are converted to values at another temperature through integrals involving thermal capacities. In reaction kinetics, for instance, if one knows a rate of disappearance of a species, by means of an integral one can generate a formula or function to describe how concentration of that species varies temporally; likewise, if one knows a probability per unit volume, or probability density, of finding an electron in an infinitesimal molecular volume containing a given point, one might evaluate a probability of finding an electron in a given volume of interest with a definite integral of that density. Of these two examples, the former involves an indefinite integral whereas the latter involves a definite integral for which the spatial coordinates specify the bounds of integration.

integral

Before we elucidate use of integrals in a chemical context, we consider a derivative, explained in chapter 3, as a basis on which to inter-relate integrals of these two kinds. If we have a prototypical derivative with f(x) a continuous function on a particular interval,

$$\frac{dy}{dx} = f(x),$$

we multiply both sides by a differential dx, to form

$$\frac{dy}{dx}\, dx = dy = f(x)\, dx$$

Because integration is formally an operation inverse to differentiation, the left side of this equation is simply dy that we integrate directly to y, as an infinite sum of infinitesimal quantities; we thus obtain a formula

$$\int dy = y = \int f(x)\, dx$$
$$= F(x)$$

for an indefinite integral on the right side, to which we assign a name F(x); F(x) contains an additive constant because on differentiation of F(x) to yield f(x) any contribution due to that constant vanishes, but such an additive constant is arbitrary unless additional information be available to set it to a particular value. Following an integral sign on the right side, f(x) is an integrand and dx is a corresponding differential quantity that serves as integrating element. For an

indefinite integral $F(x)$ of an integrand $f(x)$,

$$\frac{d}{dx}F(x) = f(x)$$

formula $F(x)$ is called an <u>antiderivative</u> of $f(x)$. The latter equation implies that

$$\frac{d}{dx}\int f(x)\,dx = f(x)$$

An indefinite integral is thus a formula, called an <u>antiderivative</u>, of which the derivative is a given formula; for some derivatives no antiderivative might exist in an explicit algebraic form. An indefinite integral or antiderivative $F(x)$ of a formula $f(x)$ is thus a family of formulae of uncountable number, each of which has $f(x)$ as its first derivative; one member of this family differs from another member by at most an additive constant. A <u>geometric</u> interpretation of this condition is that <u>curves</u> of the integral number <u>uncountably</u> that are related to one another through a <u>parallel</u> <u>displacement</u> along the <u>direction</u> of the <u>ordinate</u> <u>axis</u>.

indefinite integral

If $f(x)$ and $g(x)$ have antiderivatives on an interval, on that interval other formal definitions are a <u>distributive</u> property or rule of linearity,

$$\int [\,f(x) + g(x)\,]\,dx = \int f(x)\,dx + \int g(x)\,dx$$

or for multiple integrating variables in a sum,

$$\int (\,dx + dy + dz + ...) = \int dx + \int dy + \int dz + ...$$

and for constant c,

$$\int c\,f(x)\,dx = c \int f(x)\,dx$$

$$\int c\,dx = c \int dx$$

whereas an indefinite integral of a unit integrand yields this result,.

$$\int dx = x + C$$

in which C is a constant of integration, to be evaluated from external conditions. The latter integral is a special case, for $n = 0$, of the following general rule.

$$\int x^n\,dx = \frac{x^{(n+1)}}{n+1} + C, \qquad \text{for } n \neq -1$$

Integration by parts implies that

$$\int f(x)\,g(x)\,dx = f(x)\int g(x)\,dx - \int [\,\left(\frac{d}{dx}f(x)\right)\int g(x)\,dx\,]\,dx$$

and integration by substitution implies that

$$\int f(g(x))\left(\frac{d}{dx}g(x)\right)dx = \int f(u)\,du\,\Big|_{u=g(x)}$$

definite integral

If formula $f(x)$ be continuous on a finite interval $[a, b]$, a definite integral of $f(x)$ with respect to x between $x = a$ and $x = b$ exists along that line of axis x and has form A according to

$$A = \int_a^b f(x)\, dx = F(b) - F(a),$$

in which end points of an interval of x over which integration is performed, or the bounds of that integral, are specified to have values a and b as indicated below and above an integral sign, respectively; no such bounds appear for an indefinite integral. $f(x)$ is an integrand or expression to be integrated. As $\dfrac{d}{dx} F(x) = f(x)$, the above formula embodies the fundamental theorem of calculus. Conversely, when $f(x)$ is continuous on an interval $[a, b]$, a function F of formula

$$F(x) = \int_a^x f(t)\, dt$$

is differentiable on that interval, and $\dfrac{d}{dx} F(x) = f(x)$. This theorem is expressible in an alternative form, subject to $f(y)$ being continuous,

$$\frac{\partial}{\partial x}\left(\int_a^x f(y)\, dy \right) = f(x)$$

which shows that the derivative of the integral generates the original formula.

A definite integral might be split into multiple contributions; for two contributions,

$$\int_a^b f(x)\, dx = \int_a^c f(x)\, dx + \int_c^b f(x)\, dx$$

provided that $a \le c \le b$. Two further rules are

$$\int_a^b f(g(x))\left(\frac{d}{dx} g(x) \right) dx = \int f(u)\, du \Big|_{u = g(x)} \Big]_a^b = \int_{g(a)}^{g(b)} f(u)\, du$$

$$\int_a^b f(x)\, dx = \int f(g(t))\left(\frac{d}{dt} g(t) \right) dt \Big|_{t = g(x)} \Big]_a^b = \int_{g(a)}^{g(b)} f(g(t))\left(\frac{d}{dt} g(t) \right) du$$

in the central expressions of which the definite integral is to be eventually evaluated as a difference of the indefinite integrals into which is substituted $x = b$ and $x = a$.

If formula $f(x)$ be continuous on finite interval $[a, b]$, definite integral $\int_a^b f(x)\, dx$ also exists.

The definite integral of a continuous function is a differentiable function of its upper limit of integration:

$$\frac{\partial}{\partial x}\left(\int_a^x f(t)\, dt \right) = f(x) - \left(\frac{\partial}{\partial x} f(a) \right) = f(x),$$

because $\dfrac{\partial}{\partial x}$ f$(a) = 0$ or a derivative of a constant is zero. Integration in calculus implies evaluation of such an integral, either definite when a lower bound, such as $x = a$ and upper bound $x = b$, are specified at an integral sign, or indefinite in their absence; as an exception, in some cases a variable such as x might appear as an upper bound. An outcome of this distinction between integrals of two kinds is that one can regard an indefinite integral as an operation on a function f(x) to produce another function F(x); for a definite integral, a result A is just a number, likely with units, or a quantity that evaluates to a number, of which a value depends upon values of a and b.

   If an integrand be odd an odd formula -- such that f$(-x) = -$f(x), its definite integral over an interval $-a$ .. a is zero:

$$\int_{-a}^{a} \mathrm{f}(x)\, dx = 0 \qquad \text{if } \mathrm{f}(-x) = \mathrm{f}(x).$$

If an integrand be an even formula -- such that f$(-x) = f(x)$, its definite integral over an interval $-a$ .. a is twice that over an interval 0 .. a:

$$\int_{-a}^{a} \mathrm{f}(x)\, dx = 2\int_{0}^{a} \mathrm{f}(x)\, dx \qquad \text{if} \quad \mathrm{f}(-x) = \mathrm{f}(x).$$

   To evaluate a definite integral, a direct approach is first to find an expression F(x) in algebraic or symbolic form for an indefinite integral of the same integrand f(x), such that $\dfrac{d}{dx}\mathrm{F}(x) = \mathrm{f}(x)$, and then to calculate the difference between values of that antiderivative substituted with upper and lower bounds of an interval or domain of an integrating variable, i.e.

$$A = \int_{a}^{b} \mathrm{f}(x)\, dx = \mathrm{F}(x)\big|_{x=b} - \mathrm{F}(x)\big|_{x=a}$$

   or

$$A = \mathrm{F}(b) - \mathrm{F}(a).$$

For an integral definite or indefinite a typical approach hence involves finding first an antiderivative F(x), if it exist. For an indefinite integral, F(x) is not an antiderivative function of most general form; for this reason we take care to refer to F(x) as *an* antiderivative function. For an indefinite integral one must add to an antiderivative an arbitrary constant, such as C, as a constant of integration to obtain an antiderivative of most general form; such a constant of integration implies no dependence on an integrating variable x because, for whatever value of such a constant C, a rigorous equality

$$\dfrac{\partial}{\partial x}(\mathrm{F}(x) + C) = \dfrac{d}{dx}\mathrm{F}(x) = \mathrm{f}(x)$$

holds. As each value of C implies a distinct antiderivative function, such functions are uncountable. Each formula for differentiation thereby becomes restated as a formula for antidifferentiation. For a definite integral, a fundamental theorem of integral calculus, to be explained as a limit of a sum, ensures that we express the form of A above in terms of a difference between values of antiderivative function F(x) at upper and lower bounds of an interval over

which definite integration is performed,
$$A = F(b) - F(a),$$
as above, no matter what be the value of a constant that we choose for that antiderivative function, but that value must be the same for $F(a)$ and $F(b)$. The significance of a constant of integration reappears when we consider the solution of differential equations in chapter 7. An alternative term for a bound or end point of an interval of integration is a limit, but the latter term has a significance different from that of a limit of a function as in $\lim\limits_{n \to \infty} \sum\limits_{i=0}^{n} x_i$, as discussed in chapter 3.

According to a definition of an indefinite integral as an antiderivative,
$$\text{if } f(x) = \frac{d}{dx} F(x), \quad \int_a^x f(t)\, dt = F(t)\big|_{t=x} - F(t)\big|_{t=a} = F(x) - F(a),$$
in which a is a constant quantity. Differentiation with respect to x yields
$$\frac{\partial}{\partial x}\left(\int_a^x f(t)\, dt \right) = \frac{\partial}{\partial x}(F(x) - F(a)) = \frac{d}{dx} F(x) = f(x),$$
as shown above. Analogously,
$$\int_x^a f(t)\, dt = F(a) - F(x),$$
so that
$$\frac{\partial}{\partial x}\left(\int_x^a f(t)\, dt \right) = -\left(\frac{d}{dx} F(x) \right) = -f(x)$$
or, with $\dfrac{d}{dx} F(x) = f(x)$,
$$F(x) = \int_a^x f(t)\, dt$$
For an alternative change of variable from x to ξ, with $f(x)$ continuous between end points a and b, with $a = g(\alpha)$ and $b = g(\beta)$, and with $\dfrac{d}{d\xi} g(\xi)$ continuous between α and β such that for ξ between α and β $g(\xi)$ is between a and b,
$$\int_a^b f(x)\, dx = \int_\alpha^\beta f(g(\xi))\left(\frac{d}{d\xi} g(\xi) \right) d\xi$$

Other properties of a definite integral with both $f(x)$ and $g(x)$ continuous between the same end points $x = a$ and $x = b$ and a constant c are
$$\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx$$

$$\int_a^b f(x) + g(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$$

$$\int_a^b c\ f(x)\ dx = c \int_a^b f(x)\ dx$$

$$\int_a^a f(x)\ dx = 0$$

$$\min(f(b-a)) \le \int_a^b f(x)\ dx \ \le \max(f(b-a))$$

$$\text{if } f(x) \le g(x) \text{ on } [a, b],\ \int_a^b f(x)\ dx \le \int_a^b g(x)\ dx$$

To construct a finite increment of function $F(x)$ between values x_1 and x_2 of independent variable x, we integrate with integrand $f(x) = \dfrac{d}{dx} F(x)$,

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\ dx$$

In a chemical context, generally no ambiguity arises about a constant C of integration, as we typically possess information about a system of interest that enables us to evaluate C. For example, in a chemical reaction, in which temporal variation of a concentration is measured that embodies a significance of $f(t)$, we generally know an initial concentration at a particular time t_0, and therefore specify a value of $F(t)$ at $t = t_0$, or $F(t_0)$; a symbol for an independent variable is here t, for time, instead of customary x in abstract algebraic notation.

Methods of integration include integration of partial fractions, integration by parts and substitution; these methods are explained and illustrated in sections 4.203, 4.204 and 4.205, respectively. There is no general method applicable to an arbitrary integrand; for a particular integrand of complicated nature, one might try various methods, but for some integrands, even those of apparently simple form such as $\mathbf{e}^{(-x^2)}$, there is no known algebraic antiderivative. In such cases of a definite integral, one might apply numerical integration or quadrature to obtain a numerical result, provided that no symbols appear in the integrand or end points other than that of the integration variable.

geometric interpretation of integration

An informal geometric definition of a definite integral is that in a graph of a given positive function it represents an area between a curve and the abscissal axis between two specified values of independent variable known as bounds of integration. For bounds of integration $x = a$ and $x = b$, interval $[a, b]$ along abscissal axis x is divided into n equal subintervals of width Δx; a sum of areas of rectangles, each of width Δx and of stature $f(x)$ between abscissal axis x and a curve of that positive function $f(x)$ for some x within that subinterval, as $n \to \infty$, yields a total area of that region and a value of an associated definite integral. Each element of width Δx along the abscissal

axis is associated with a corresponding element of area f(x) Δx of a rectangle bounded above by the curve pertaining to f(x). A geometric definition, due to <u>Riemann</u>, of a definite integral is hence an area that is formed as a *limit of a sum* of areas of contiguous rectangles between the abscissal axis and the curve corresponding to a formula f(x) pertaining to a bounded function *f* taking real values between two points *a* and *b* on that axis as the number of such rectangles increases without limit, so that the width Δx of each rectangle <u>tends to</u> zero:

$$\lim_{n \to \infty} \sum_{j=0}^{n} \mathrm{f}(a + j\,\Delta x)\,\Delta x = \int_{a}^{b} \mathrm{f}(x)\,dx = A$$

<u>Lebesgue</u> <u>integration</u> is a generalization of a Riemann integral to functions that have <u>discontinuities</u>, and a <u>Stieltjes</u> integral is an extension of a Riemann integral that allows integration of a function with respect to another function.

An alternative definition of a definite integral is a difference of two indefinite integrals, each of which is evaluated at an end point *b* or *a* of integration, as mentioned above:

$$\int_{a}^{b} \mathrm{f}(x)\,dx = \int \mathrm{f}(x)\,dx \Big|_{x=b} - \int \mathrm{f}(x)\,dx \Big|_{x=a}.$$

applications of integration

Two common applications of a definite integral involve finding an <u>arithmetic mean</u> value and a <u>root-mean-squared</u> value of a continuous formula in some stated domain of independent variable. An arithmetic <u>mean</u> value of a formula between two bounds corresponds to simply the area under the curve of that formula between those end points divided by their difference,

$$arithmetic\ mean = \frac{1}{b-a} \int_{a}^{b} y\,dx \ .$$

The <u>mean-value theorem</u>, which is an <u>elementary</u> result in mathematical <u>analysis</u>, is a particular application of this definition. A <u>root-mean-squared</u> value of a formula in a stated range between two bounds that define an interval becomes analogously the <u>square root</u> of an integral of that formula <u>squared</u> between those bounds divided by their difference:

$$root\text{-}mean\text{-}squared\ value = \sqrt{\dfrac{1 \displaystyle\int_{a}^{b} y^2\,dx}{b-a}}\ .$$

These quantities are important not only in purely <u>statistical</u> contexts but also, for instance, in relation to properties of molecules in a gaseous sample. Exercises on applications of these formulae to a kinetic-molecular description of gaseous samples arise after section 4.303 because the upper bound of integration being ∞ makes these integrals formally <u>improper</u>.

A property of a <u>plane figure</u> is its <u>centroid</u>. The centroid of a <u>triangle</u> is located at a <u>point</u> at which its medians coincide; a <u>median</u> of a triangle is a <u>segment</u> of a <u>straight</u> <u>line</u> drawn from one <u>vertex</u> to a <u>midpoint</u> of the opposite side. For a continuous object of uniform density in three dimensions, the centroid becomes the centre of mass or centre of gravity, which is evaluated through integrals. The position of a centroid of a plane figure depends on not only its area but also the way that the area is distributed, i.e. the shape of that figure y(x). <u>Cartesian coordinates</u> (x_c, y_c)

of a centroid of a plane figure relative to axes x and y are evaluated with these integrals,

$$x_c = \frac{\displaystyle\int_a^b x\,y\,dx}{\displaystyle\int_a^b y\,dx}$$

and

$$y_c = \frac{\dfrac{1}{2}\displaystyle\int_a^b y^2\,dx}{\displaystyle\int_a^b y\,dx},$$

in which an integral in each <u>denominator</u> is recognisable as the area of the figure that extends from $x = a$ to $x = b$. For a planar figure of uniform density, the centroid coincides with the <u>centre of mass</u>. These coordinates of a centroid are first <u>moments</u> of this figure about the corresponding axes, x or y; moments of greater <u>order</u> are evaluated analogously with integrands involving greater powers. First moments of a figure about mutually <u>perpendicular</u> axes passing the centroid are zero. For a non-continuous <u>figure</u> or object in two or three <u>dimensions</u>, such as a molecule in which mass is concentrated at positions of atomic nuclei, a centroid is evaluated through finite sums; for a molecule the second moment of mass, called the <u>moment of inertia</u>, is important in relation to its spectral properties pertaining to rotational motion, according to which its geometric structure might be characterised in favourable cases.

Geometric applications of a definite integral hence include an area of a figure of a given shape, a volume of a solid of revolution, an area of a surface of revolution, a length of a curve, and centroids of a figure; physical applications of a definite integral include work as an integral of force over some distance or as an integral of pressure over some volume, a centre of mass that is related to a product of uniform mass density and a centroid, moments of inertia of a body of uniform density and a given shape, of which some such applications have also direct chemical pertinence.

special functions

Although a differentiation to yield an explicit derivative is practicable for almost any continuous formula or function in algebraic form, the converse is false. An integration of an algebraic form might commonly yield a function or formula not in algebraic form, such as

$\displaystyle\int \frac{1}{x}\,dx = \ln(x)$ that yields the indicated elementary formula; in many cases the result of such an indefinite integral is a <u>special formula</u> or function that is simply a name given to that integral. Below appear a few such instances.

Among such special functions involving integrals, the <u>gamma function</u> Γ, introduced in section 1.117, is defined as this integral,

$$\Gamma(x) = \int_0^\infty t^{(x-1)} \, e^{(-t)} \, dt$$

which is convergent for all real $x > 0$. For a positive integer n, $\Gamma(n+1) = n!$, a factorial, as introduced in section 1.116. Closely related to function Γ are the *exponential integral*,

$$\mathrm{Ei}(x) = \int_x^\infty \frac{e^{(-t)}}{t} \, dt \qquad \text{for } x > 0,$$

and the *logarithmic integral*,

$$\mathrm{Li}(x) = \int_0^x \frac{1}{\ln(t)} \, dt ,$$

which are connected through $\mathrm{Ei}(x) = -\mathrm{Li}(e^{(-x)})$. Other related integrals are the cosine integral,

$$\mathrm{Ci}(x) = -\int_x^\infty \frac{\cos(t)}{t} \, dt$$

and the sine integral,

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(t)}{t} \, dt,$$

of which applications appear below. Two other forms that yield a factorial for positive integer n are

$$\int_0^1 \ln\!\left(\frac{1}{t}\right)^n dt = \int_0^\infty \frac{t^n}{e^t} \, dt = n!$$

To verify the left integral, for which one might expect a problem for the integrand at both bounds of the integral, we must express $\ln\!\left(\dfrac{1}{t}\right) = -\ln(t)$, and raise that quantity to power n:

> **`Int((-ln(t))^n,t=0..1)  = int((-ln(t))^n,t=0..1) assuming n>0;`**

the result is clearly equal to $\Gamma(n+1) = n\,\Gamma(n) = n!$. This plot of $\Gamma(x)$,
> **`plot(GAMMA(x), x=-5..5, -6..10, discont=true,`**
> **`      title="Gamma function", titlefont=[TIMES,BOLD,14]);`**

>

shows that $\Gamma(x)$ has discontinuities at $x = 0$ and $x =$ any negative integer, but is a continuously increasing function for $x > \dfrac{3}{2}$. This expression of $\Gamma(x)$ as an integral is useful for an evaluation of

other integrals, such as $\displaystyle\int_0^\infty x^6\, \mathrm{e}^{(-x)}\, dx$

```
> Int(x^6*exp(-x), x=0..infinity) = int(x^6*exp(-x),
  x=0..infinity);
```

which is just $\Gamma(7) = 6!$;

```
> GAMMA(7) = 6!;
```

the latter integral is a special case for $n = 6$ of this general integral,

```
> Int(x^n*exp(-x), x=0..infinity) = int(x^n*exp(-x),
  x=0..infinity);
> 
```

which yields function Γ directly, and likewise for analogous integrals that *Maple* evaluates automatically. A *duplication formula* involving Γ functions is

$$\Gamma\!\left(n + \frac{1}{2}\right) = \frac{\sqrt{\pi}\,\Gamma(2\,n)}{2^{(2\,n-1)}\,\Gamma(n)},$$

which takes a simple form for positive integer n. Although *Maple* fails to solve the following integral so as to recognise it as yielding $\Gamma(n + 1)$ or $n!$,

```
> int(ln(1/t)^n, t=0..1) assuming n::posint;
```

it succeeds with this form.

```
> Int(t^n/exp(t), t=0..infinity) = int(t^n/exp(t),
  t=0..infinity);
> rhs(%) = convert(rhs(%), factorial);
> Beta
```

Also introduced in section 1.117, beta function $\mathrm{B}(m, n)$ is defined through this integral,

$$\mathrm{B}(m, n) = \int_0^1 x^{(m-1)}\,(1 - x)^{(n-1)}\, dx$$

which converges for $m > 0$ and $n > 0$. With a substitution $x = \sin(\alpha)^2$, an alternative form is

$$\mathrm{B}(m, n) = 2\int_0^{\frac{\pi}{2}} \sin(\alpha)^{(2\,m-1)}\,\cos(\alpha)^{(2\,n-1)}\, d\alpha$$

Use of the latter form and trigonometric identities yields a reduction formula

$$\mathrm{B}(m, n) = \frac{(m-1)\,(n-1)}{(m+n-1)\,(m+n-2)}\,\mathrm{B}(m-1, n-1)$$

Functions Γ and B are related through

$$\mathrm{B}(m, n) = \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}$$

which is valid for general real values of m and n, and which is relatable in turn to both factorials and binomial coefficients for m and n being positive integers.

The error function erf(x), defined through this integral,

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{(-t^2)} \, dt$$

that occurs in physical and statistical applications, has no explicit algebraic solution, but is readily evaluated in *Maple* for any value of x. Two related special functions with trigonometric functions instead of an exponential function are Fresnel integrals.

Another special function that arises from an integral involving an exponential function is

$$\mathrm{Ei}(x) = \int_{-\infty}^x \frac{e^t}{t} \, dt$$

called the *exponential integral* Ei(x), which has three related functions likewise based on integral: the logarithmic integral Li(x),

$$\mathrm{Li}(x) = \mathrm{Ei}(\ln(x)) \quad \text{or} \quad -\mathrm{Li}(e^{(-x)}) = \mathrm{Ei}(x)$$

which provides an approximation to the number of primes less than or equal to x, the sine integral Si(x),

$$\mathrm{Si}(x) = \int_0^x \frac{\sin(t)}{t} \, dt$$

and the cosine integral Ci(x),

$$\mathrm{Ci}(x) = \gamma + \ln(x) + \int_0^x \frac{\cos(t) - 1}{t} \, dt$$

Elliptic functions that provide a means to evaluate other definite integrals are known of a few variants, complete and incomplete and complementary. An incomplete elliptic integral of the first kind is expressible in trigonometric terms as

$$F(\alpha, \phi) = \int_0^\phi (1 - \alpha^2 \sin(\theta)^2)^{\left(-\frac{1}{2}\right)} \, d\theta$$

which is valid for ϕ in a closed interval $[0, \frac{\pi}{2}]$ and for α in an open interval $]0,1[$. An incomplete elliptic integral of the second kind is expressible in trigonometric terms as

$$E(\alpha, \phi) = \int_0^\phi (1 - \alpha^2 \sin(\theta)^2)^{\left(\frac{1}{2}\right)} \, d\theta$$

with the same ranges for ϕ and α. If $\phi = \frac{\pi}{2}$, these integrals are complete. *Maple* evaluates all these elliptic integrals with arbitrary values of arguments, and provides information on any through the

FunctionAdvisor, for instance this case.
> **FunctionAdvisor(EllipticK);**
Elliptic functions are expressible also in algebraic terms, rather than trigonometric terms, as indicated in the pertinent help pages.
>

geometric applications of definite integrals

A geometric application of a definite integral alternative to finding an area under a curve in two dimensions or the volume of a solid or area of a solid of revolution in three dimensions is to find a length of an arc of a curve; this property is called a *path integral* or line integral, or a curvilinear integral of the second kind. This property has chemical applications whenever one focuses attention on, for example, an appropriately defined path either between two thermodynamically defined states, or a trajectory of minimum energy in a chemical reaction: in these, and other, situations, our typical concern is with a function of more than one independent variable, the properties of which we describe in chapter 5. To introduce this concept, we consider a length of an arc along a curve of function y(x) in a plane. Between two adjacent points on a curve, we take a difference in abscissae of their coordinates to be dx and a corresponding difference in ordinates to be dy; approximating the arc of this curve by a segment of a straight line, we apply a theorem of Pythagoras that the length ds of this hypotenuse approximating the arc is

$$ds^2 = dx^2 + dy^2 \; ;$$

dividing through by dx^2 and taking a square root of each side, we express a derivative $\dfrac{ds}{dx}$ as

$$\frac{ds}{dx} = \sqrt{1 + \left(\frac{dy}{dx}\right)^2} \; .$$

Integrating the left side with respect to x between two points a and b yields, in cartesian coordinates,

$$s = \int_a^b \frac{ds}{dx}\, dx = \int_a^b \sqrt{1 + \left(\frac{d}{dx}\,y(x)\right)^2}\, dx = \int_{t_1}^{t_2} f(x(t), y(t)) \sqrt{\left(\frac{d}{dt}\,x(t)\right)^2 + \left(\frac{d}{dt}\,y(t)\right)^2}\, dt \; ,$$

or in polar coordinates,

$$s = \int_{\theta_1}^{\theta_2} \sqrt{\left(\frac{d}{d\theta}\,r(\theta)\right)^2 + r(\theta)^2}\, d\theta = \int_{t_1}^{t_2} f(r(t), \theta(t)) \sqrt{\left(\frac{d}{dt}\,r(t)\right)^2 + r^2 \left(\frac{d}{dt}\,\theta(t)\right)^2}\, dt \; ;$$

in each case on the right side we show also the result in parametric form, with parameter t. The area bounded by a curve expressed in polar coordinates as $r = f(\theta)$ is

$$area = \frac{1}{2} \int_{\theta_1}^{\theta_2} f(\theta)^2\, d\theta$$

and the slope of a line tangent to that curve is

$$\frac{\left(\dfrac{d}{d\theta}\,f(\theta)\right)\sin(\theta)+f(\theta)\cos(\theta)}{\left(\dfrac{d}{d\theta}\,f(\theta)\right)\cos(\theta)-f(\theta)\sin(\theta)}$$

For $r=f(\theta)$ continuous on an interval $[\alpha, \beta]$, the area enclosed by that curve between those bounds is

$$area=\frac{1}{2}\int_{\alpha}^{\beta}f(\theta)^2\,d\theta$$

As an alternative approach to evaluate the length of a curve, we apply an explicit definition of a curve involving a parametric variable. A curve might be expressible as a graph of an equation or formula, but also in parametric form as equations in a set involving a further variable. For instance, a graph in plane xy of equation

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$$

generates an ellipse, but this geometric figure is also the graph of these parametric equations: $x=a\cos(t)$, $y=b\sin(t)$ with parametric variable t in a domain $[0,2\pi]$. If the <u>domain</u> of function f named y be an <u>interval</u> I, the graph of $y(x)$ in plane xy is also the graph of parametric equations

$$x=t,\ y=f(t)\ .$$

An intersection of graphs in three dimensions of two equations such as $y=f(x)$ and $z=g(x,y)$ is analogously the graph of parametric equations

$$x=t,\ y=f(t),\ z=g(t,\,f(t))\ .$$

Parametric equations are thus widely applicable. A curve in three spatial dimensions is thus a graph of parametric equations

$$x=f(t),\ y=g(t),\ z=h(t)$$

in a set such that f, g and h be continuous on an interval $[a, b]$ of parametric variable t; the points that correspond to a and b constitute the end points of the curve. A simple curve has a property that, with a possible exception of a and b that would make it a <u>simple closed curve</u>, no two numbers in that interval $[a, b]$ determine the same point on the curve. For a closed curve the end points coincide.

A curve is a continuous <u>image</u> of an interval. For a point $P(t)$ in three spatial dimensions with coordinates $f(t)$, $g(t)$, $h(t)$, its distance from a fixed point $P(t_0)$ is

$$d(t)=\sqrt{(f(t)-f(t_0))^2+(g(t)-g(t_0))^2+(h(t)-h(t_0))^2}$$

If $f(t)$, $g(t)$, $h(t)$ be continuous, this distance approaches zero as $t\to t_0$, in which case $P(t)$ is a continuous function of t and the curve is <u>connected</u>.

A formula or function $f(x)$ that is <u>real</u> rather than <u>complex</u> and that can be represented with a <u>convergent</u> <u>power series</u> throughout a vicinity of $x=a$ is called <u>analytic</u> at $x=a$. For a complex <u>variable</u> $z=x+i\,y$, in which $i=\sqrt{-1}$, such that z is associated with a point (x, y) in a <u>complex</u>

a single-valued function $f(z)$ has a derivative $\dfrac{d}{dz} f(z) = f'(z)$ if

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

for which $\Delta z \to 0$ through any complex values. For such a variable z and a fixed point z_0 in any open, simply connected domain R excluding a boundary, a function $f(z)$ is *analytic* in R if any of these four conditions hold.

- $f(z)$ has a derivative $f'(z)$ at each point in R.

- $f(z)$ is integrable in R in a sense that an integral $\int f(z)\, dz = 0$ about every closed path in R; thereby

$$F(z) = \int_{z_0}^{z} f(z)\, dz$$

  is an analytic function of z having only a single value at each point and is independent of path in R.

- $f(z)$ has an expansion as a Taylor series in $(z - z_0)$ to various powers about each point z_0 in R.

- $f(z) = u(x, y) + i\, v(x, y)$, for which both $u(x, y)$ and $v(x, y)$ have continuous partial derivatives -- cf section 5.104 -- that satisfy these differential equations:

$$\frac{\partial}{\partial x} u = \frac{\partial}{\partial y} v, \quad \frac{\partial}{\partial y} u = -\left(\frac{\partial}{\partial x} v\right)$$

  in which u and v are conjugate functions, each satisfying Laplace's equation -- cf example *x6.403*,

$$\left(\frac{\partial^2}{\partial x^2} u\right) + \left(\frac{\partial^2}{\partial y^2} u\right) = 0$$
$$\left(\frac{\partial^2}{\partial x^2} v\right) + \left(\frac{\partial^2}{\partial y^2} v\right) = 0$$

Hence, from conditions in the first derivatives above,

$$\frac{\partial^2}{\partial x^2} u = \frac{\partial^2}{\partial y\, \partial x} v, \quad \frac{\partial^2}{\partial y^2} u = -\left(\frac{\partial^2}{\partial x\, \partial y} v\right)$$

If $f(z)$ be analytic at all points along a circle with centre O, but not at O, $z_0 = O$ is an isolated singular point; this point is a pole of order n if n be the smallest positive integer for which $(z - z_0)^n f(z)$ remains finite. If there exist no such value of n, z_0 is an essential singularity. Polynomial, sine, cosine and exponential formulae are analytic everywhere, and sums, differences, products of polynomial, sine, cosine and exponential formulae are also analytic everywhere; quotients of any two such expressions are analytic at all points at which the denominator is other than zero.

A scalar function Φ with formula $\Phi(x, y)$, continuous second derivatives and satisfying

Laplace's equation,

$$\left(\frac{\partial^2}{\partial x^2}\Phi\right) + \left(\frac{\partial^2}{\partial y^2}\Phi\right) = 0 \ ,$$

on a particular domain is called an harmonic function.

series and integrals

A necessary, but insufficient, condition for convergence of a series $\sum_k c_k$ is that magnitudes of successive terms decrease to ensure that at least a partial sum S_n yields a finite result in the limit as n tends to infinity, as described in section 2.403. Whereas $c_k = 2^k$ obviously fails this criterion, both $c_k = \dfrac{1}{k}$ and $c_k = \dfrac{(-1)^{(k+1)}}{k}$ generate series that *might* converge. Using a functional notation with arrow to specify terms in the series, we write c_k as c(k), so that $c := k \rightarrow c_k$. An integral test involves bounding a sum by an integral, over a region from 0 to ∞, and discrete variable k by continuous variable x. A decreasing function f(x) has a property that c(k) is bounded by f(k) at all k; the original sum $\sum\limits_{k=1}^{\infty} c_k$ becomes bounded by an integral $\int_0^{\infty} $ f(x) dx, with f(x) as a decreasing function as x increases. The sum $\sum\limits_{k=1}^{\infty} c_k$ is equivalent to an area that must be less than an area implied by that definite integral $\int_0^{\infty} $ f(x) dx because c_k is less than, or bounded by, f(x) at all k. One achieves convergence if this integral have a finite value, real or complex.

A power series, of form $\sum_k c_k (x - a)^k$, is useful within a circle of convergence because it can represent an analytic function f(x) and because operations addition, subtraction, multiplication, division, differentiation and integration are valid for each term. Coefficients c_k, $k = 0, 1, 2, ...$ of such a power series are unique for a given representation. Within that circle of convergence, an infinite series in the form of a power series is manipulable like a polynomial. As a generalization of a Taylor series, a series that contains a specified independent variable to both positive and negative powers, unlike only positive powers that occur in a Taylor series, is named after Laurent, and might be expressed as a doubly infinite power series in a complex number z,

$$f(z) = \sum\limits_{j=-\infty}^{\infty} c_j (z - z_0)^j$$

for which terms to non-negative powers are equivalent to a Taylor series and further when $z_0 = 0$ to a Maclaurin series, and terms to negative powers are equivalent to a Taylor series in $(z - z_0)^{(-1)}$ or $\dfrac{1}{z - z_0}$; when terms with both non-negative and negative powers exist in such a series, the general descriptor Laurent series is applicable. Terms in such a series with negative powers

constitute the *principal part*, and other terms constitute the *regular part*. A function analytic in an annular region

$$0 < r_1 \leq |z - z_0| \leq r_2$$

might be represented with such a Laurent expansion in which coefficients c_j are evaluated with this integral,

$$c_j = \frac{\displaystyle\int \frac{f(z)}{(z - z_0)^{(j+1)}}\, dz}{2\,\pi\,i}$$

with $i = \sqrt{-1}$. Integration is performed along any simple closed contour in the region of analyticity enclosing an inner boundary $|z - z_0| = r_1$. Providing that function $f(z)$ is regular in a region bounded with a closed path except for poles and isolated singularities of finite number, the value of this contour integral is a product $2\,\pi\sqrt{-1}$ with a sum of all residues at all poles and essential singularities inside that closed path. With $f(z)$ expressed as $\dfrac{p(z)}{q(z)}$ in which $q(z)$ is regular and $p(z)$ has a simple pole, or pole of order unity, at z_0, the residue is

$$c_{-1} = \frac{q(z)}{\left(\dfrac{d}{dz}\,p(z)\right)\Big|_{z = z_0}} \;;$$

if z_0 be a pole of order n, the residue is

$$c_{-1} = \frac{\left(\dfrac{\partial^{n-1}}{\partial z^{n-1}}\left((z - z_0)^{(n-1)}\,f(z)\right)\right)\Big|_{z = z_0}}{(n-1)!}.$$

If z_0 be an isolated essential singularity, the Laurent expansion yields the residue.

A *functional series*, of form

$$c_0 f_0(x) + c_1 f_1(x) + c_2 f_2(x) + \ldots$$

contains terms each of which comprises a product of a constant c_k and a formula $f_k(x)$ of some function f in which constant c_k serves as coefficient of basis function $f_k(x)$. For a power series each term $f_k(x)$ is simply x^k or $(x - a)^k$. For a Fourier series each term has a basis function of formula $\sin(k\,x)$ or $\cos(k\,x)$, or a linear combination of such terms, or equivalently an exponential formula $e^{(i k x)}$ in which $i = \sqrt{-1}$; whereas a Taylor series of $f(x)$ has coefficients determined by its derivatives at a single point, an infinite Fourier series of $f(x)$ as a sum of trigonometric, or their equivalent exponential, terms has coefficients determined by its integrals over a fixed interval. This sum of continuous and periodic formulae converges pointwise to a possibly discontinuous and non-periodic function; for a Fourier series that represents an even function, for which $f(x) = f(-x)$, all sine terms vanish, whereas for a Fourier series that conversely represents an odd function, for which $f(-x) = -f(x)$, all cosine terms vanish -- their coefficients are identically zero.

As an obvious generalization of a geometric series, expressed in symbolic form as

$$1 + x + x^2 + x^3 + ...$$

a hypergeometric series has this form:

$$1 + \frac{\alpha \, \beta \, x}{1 \, \gamma} + \frac{\alpha \, (\alpha + 1) \, \beta \, (\beta + 1) \, x^2}{1*2 \, \gamma \, (\gamma + 1)} + ...$$

Provided that γ be neither 0 nor a negative integer, this series is absolutely convergent is $|x| < 1$, divergent if $|x| > 1$ but convergent if $|x| = 1$ for $\gamma > \alpha + \beta$ or if $x = -1$ for $\gamma > \alpha + \beta + 1$. This series is commonly denoted as $_2F_1(\alpha, \beta; \gamma; x)$ in which notation $2F1$ implies two parameters of type α and one parameter of type γ, with variable x; $_2F_1(\alpha, \beta; \gamma; x) = \,_2F_1(\beta, \alpha; \gamma; x)$. Differentiation of this series with respect to its variable yields

$$\frac{\partial}{\partial x} 2F1(\alpha, \beta, \gamma, x) = \frac{\alpha \, \beta}{\gamma} \,_2F_1(\alpha + 1, \beta + 1; \gamma + 1; x)$$

and $_2F_1(\alpha, \beta; \gamma; 0) = 1$, so that

$$\left(\frac{\partial}{\partial x} 2F1(\alpha, \beta, \gamma, x) \right)\bigg|_{x=0} = \frac{\alpha \, \beta}{\gamma}$$

Although $_2F_1(-2, 1; -5; x)$ implies a series with absent terms,

$$_2F_1(-2, 1; -5; x) = 1 + \frac{2 \, x}{5} + \frac{x^2}{10} + \frac{x^6}{10} - \frac{2 \, x^7}{5} - x^8 + ...$$

Maple fails to continue the series after $\frac{x^2}{10}$.

This integral,

$$\int_0^\infty x^{(n-1)} \, e^{(-x)} \, dx = \Gamma(n)$$

defines gamma function $\Gamma(n)$ and converges for $n > 0$. Similarly, for $m > 0$ and $n > 0$, this integral,

$$\int_0^1 x^{(m-1)} \, (1-x)^{(n-1)} \, dx = B(m, n)$$

defines beta function $B(m, n)$, both of which functions are introduced for numerical arguments in section 1.117. The latter function provides a solution to this integral.

$$\int_0^1 (1-t)^{(\gamma - \alpha - \beta - 1)} \, t^{(\beta - 1)} \, dt = B(\beta, \gamma - \beta) * \,_2F_1(\alpha, \beta; \gamma; x)$$

for $\gamma - \alpha - \beta > 0$ and $\beta > 0$, hence in terms of hypergeometric functions..

improper integral

A definite integral is considered improper if one end point or both end points of integration be infinite, or if an integrand become infinite, or undefined, between end points of integration; such integrals are termed improper integrals of first and second kinds respectively. We treat an improper integral of the first kind directly with an upper or lower bound, or both bounds of an interval of integration, as follows, provided that these limits exist.

$$\int_a^\infty f(x)\,dx = \lim_{t\to\infty} \int_a^t f(x)\,dx$$

$$\int_{-\infty}^a f(x)\,dx = \lim_{t\to(-\infty)} \int_t^a f(x)\,dx$$

$$\int_{-\infty}^\infty f(x)\,dx = \lim_{t\to(-\infty)} \int_t^a f(x)\,dx + \lim_{t\to\infty} \int_a^t f(x)\,dx$$

If $f(x)$ become infinite or have a singularity at $x = b$ with $b \neq a$, an improper integral of the second kind becomes

$$\int_a^b f(x)\,dx = \lim_{h\to 0} \int_a^{b-h} f(x)\,dx\;;$$

if the limit exist, it becomes the value of the improper integral. If a singularity occur in an interior of an interval between bounds of integration, the integral is a sum of two improper integrals on subintervals above and below the singularity; with a singular point at c within $[a, b]$, the integral is accordingly

$$\int_a^b f(z)\,dz = \lim_{h\to 0} \int_a^{c-h} f(z)\,dz + \lim_{h\to 0} \int_{c+h}^b f(z)\,dz$$

If these limits exist, the corresponding improper integral converges, otherwise it diverges.

A Cauchy principal value of a definite integral of $f(x)$ over interval $[a, b]$, in which a formula is undefined at interior point $x = c$, is defined formally in the following way, in which ε is a small positive number:

$$\lim_{\varepsilon\to 0} \int_a^{c-\varepsilon} f(x)\,dx + \int_{c+\varepsilon}^b f(x)\,dx.$$

If a limit of a sum of two integrals yield a finite result as ε tends to zero, the result is termed a Cauchy principal value.

When a finite discontinuity be present within an interval of integration, a correct result is obtainable on integrating separately the two parts on either side of that discontinuity; for instance, if within an interval $[a, b]$ there exist a discontinuity at c such that applicable formulae be $f(x)$ in a subinterval $[a, c]$ and $g(x)$ in a subinterval $[c, b]$, the value of the total integral is a sum of these contributions from the two subintervals:

$$\int_a^c f(x)\,dx + \int_c^b g(x)\,dx\;.$$

A differentiable function or its formula is necessarily continuous in its domain, or a subinterval contained within that domain, but a converse condition is false, as we demonstrate in section 3.303. Just as there exist continuous algebraic functions that are not everywhere differentiable, so integration of many functions, even simple ones, is impracticable in closed form; in such cases no simple algebraic expression in terms of elementary functions and their combinations is known to have a derivative equal to a particular integrand for which integration is sought. Three

possibilities arise in such a situation: an expression in terms of a special function might be known, for which a *special function* is previously defined as the sought integral or something related to it; an integrand might yield a satisfactory approximation in series form within a region of interest -- a condition that one ought to test -- that can be integrated term by term; if those approaches fail, a final resort is numerical integration or numerical quadrature. Beyond such a situation in which an algebraic expression for an antiderivative function remains elusive, it is important to test that an integrand is well behaved, in a sense of taking a finite value for any argument of the function on an interval of interest; a plot of an integrand is enlightening in this respect. If an integrand contain a finite or infinite discontinuity, one must take care to elucidate the form of either an antiderivative or a value of a definite integral, depending whether an integral be indefinite or definite: such *improper* integrals we examine in section group 4.3.

   To evaluate a definite integral containing an integrand for which no antiderivative is known, one can either approximate that integrand by a series and integrate algebraically term by term or effect numerical integration, also called numerical quadrature. In the latter case one applies a formula involving weighted sums of function values at given points according to standard rules, such as the trapezium rule or Simpson's rule; for the latter rule the number of terms in the sum must be even. Such a rule has order n if it be exact for a polynomial of that degree: hence the trapezoidal rule has order 1 and Simpson's rule has order 3: the latter rule is hence more accurate than the former, and both are more accurate than the rectangular rule, for which the curve of the formula to be integrated is approximated by a sequence of step functions. Romberg integration is an extension of the trapezium rule in which a successive combination of estimates to produce estimates equivalent to fitting polynomials of greater order; as this method is susceptible to rounding error through multiple arithmetical operations at each stage, enhanced numerical precision in performing those operations is advisable. A quadrature formula is described as closed, as for a closed interval, if it sample the end points, or open otherwise. In gaussian quadrature, the intervals between adjacent points must be unequal and the number, $n + 1$, of evaluations of the function is fixed; this method, which is generally exact for a polynomial of order $2\,n + 1$, is amenable for quadrature in multiple dimensions.

   To indicate the accuracy of numerical integration with these methods, we consider a formula $f(x)$ as integrand for which we seek the area of a narrow region of width $h = b - a$ containing a point $x = \xi$:

for the trapezoidal rule, $n = 1$,

$$\int_a^b f(x)\,dx = \frac{(b-a)\,(f(a)+f(b))}{2} - \frac{(b-a)^3 \left(\left. \frac{d^2}{dx^2} f(x) \right|_{x=\xi} \right)}{12}$$

for Simpson's rule with three terms in a sum, $n = 2$,

$$\int_a^b f(x)\,dx = \frac{(b-a)\left(f(a) + 4\,f\!\left(\frac{b-a}{2}\right) + f(b) \right)}{3} - \frac{(b-a)^5 \left(\left. \frac{d^4}{dx^4} f(x) \right|_{x=\xi} \right)}{12}$$

for Simpson's rule with four terms in a sum, $n = 3$,

$$\int_a^b f(x)\,dx = \frac{3(b-a)\left(f(a) + 3\,f\left(\frac{b-a}{3}\right) + 3\,f\left(\frac{2(b-a)}{3}\right) + f(b)\right)}{8} - \frac{3(b-a)^5\left(\dfrac{d^4}{dx^4}f(x)\right)\Big|_{x=\xi}}{80}$$

and for five terms in a sum, $n = 4$,

$$\int_a^b f(x)\,dx = \frac{2(b-a)\left(7\,f(a) + 32\,f\left(\frac{b-a}{4}\right) + 12\,f\left(\frac{b-a}{2}\right) + 32\,f\left(\frac{3(b-a)}{4}\right) + 7\,f(b)\right)}{45}$$
$$-\frac{8(b-a)^7\left(\dfrac{d^6}{dx^6}f(x)\right)\Big|_{x=\xi}}{945}$$

in each case the specified value n denotes that the formula is exact for x^k, with $k = 0, 1, ..., n$. When the region $a \mathrel{..} b$ is not narrow, Simpson's rule is applicable in a composite form for m subintervals such that the rule is applied to each subinterval of width $h = \dfrac{b-a}{m}$ and sums apply with x incremented as $x_j = a + j\,h$ for $j = 0, 1, ..., m$; for instance,

$$\int_a^b f(x)\,dx = \frac{h\left(f(a) + 2\left(\sum_{j=1}^{\frac{m}{2}-1} f(x_{2j})\right) + 4\left(\sum_{j=1}^{\frac{m}{2}} f(x_{2j-1})\right) + f(b)\right)}{3} - \frac{(b-a)\,h^4\left(\dfrac{d^4}{dx^4}f(x)\right)\Big|_{x=\xi}}{180}$$

In each case the last term indicates the order of the error of this numerical approximation.

In contrast with differentiation for which algorithmic rules invariably yield an algebraic result as a derivative of a continuous algebraic function, no systematic approach to integration is reliable or completely general. An algorithmic approach is applicable only to a rational function, having both as numerator and denominator a polynomial, although an algorithm by Risch and Norman is embedded in *Maple* to respond to integrals of radical and trigonometric functions. For other integrands, numerical evaluation of a definite integral is practicable if no symbolic parameter appear in that integrand or if bounds of integration contain no symbolic parameters; in other cases one can obtain an idea of the behaviour of a definite integral by substituting varied numerical values of symbolic parameters. An alternative approach involves plotting an integrand to discover which region produces a large contribution to the integral, and then approximating the integrand with a series expanded about a point well within that region; if an integrand have multiple regions of large contributions separated by minima, a sum of contributions to the integral might be obtained on applying series expanded about multiple points.

Fourier series

Each unique periodic formula or function $f(x) = f(x + k\,p)$ that is partially monotonic and continuous is uniquely representable as a Fourier series with a decomposition into a spectrum of $f(x)$ according to discrete frequencies $k\,f_0$. As an instance of a functional series, an important application of integrals of trigonometric functions sine and cosine arises in construction of a Fourier series to represent, or to approximate, a discontinuous or singly valued periodic function

on assigning suitable values to coefficients in such a series; for this purpose we describe, in section group 4.5, solution of those integrals and formation of Fourier series of selected geometric forms. A Fourier series is remarkable because therein a sum of continuous and periodic functions converges pointwise to a possibly discontinuous and non-periodic function. A periodic formula or function conforms to a condition $f(x) = f(x + k p)$ in which p corresponds to the period and k is an integer, positive or negative. Because sine and cosine functions, and their exponential counterpart of form $e^{(ix)}$, are periodic functions, they are appropriate to represent a periodic function of another kind. For two periodic functions $f(x)$ and $g(x)$, their inner product (f, g) is defined as

$$(f, g) = \int_{-\frac{p}{2}}^{\frac{p}{2}} f(x) \, g(x) \, dx$$

in which the bounds of integrations define between them at least one period of these functions; the domain of integration might also be $-\infty \, .. \, \infty$ or $0 \, .. \, \infty$, or $-\pi \, .. \, \pi$, depending upon conditions. An even formula or function is one for which $f(x) = f(-x)$, such as $\cos(x)$, whereas for an odd function $f(-x) = -f(x)$, such as $\sin(x)$. These properties result from the orthogonality properties of these trigonometric functions.

$$\int_{-\pi}^{\pi} \sin(n\,x) \cos(m\,x) \, dx = 0$$

$$\int_{-\pi}^{\pi} \sin(n\,x) \sin(m\,x) \, dx = 0 \;\; \text{if } n \neq m \,, \; = \pi \text{ if } n = m$$

$$\int_{-\pi}^{\pi} \cos(n\,x) \cos(m\,x) \, dx = 0 \;\; \text{if } n \neq m, \; = \pi \text{ if } n = m$$

An integral of a product of an even function, such as cosine, and an odd function, such as sine, over a domain that contains periods of integer number, or over an infinite domain, must be zero. A further advantage of a Fourier series is its application to describe a discontinuous function; a Fourier series might be differentiated or integrated term by term, and summation of a Fourier series is practicable for sum functions.

An expansion of a function or its representation with sine and cosine, or exponential, functions serves as a Fourier series:

$$F(\theta) = \frac{a_0}{2} + \left(\sum_{k=1}^{\infty} a_k \cos(k\,\theta) \right) + \left(\sum_{k=1}^{\infty} b_k \sin(k\,\theta) \right)$$

For a periodic function $f(\theta)$ with a period of extent $2\,\pi$ rad, coefficients a_k and b_k are generated according to these integrals,

$$a_k = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \cos(k\,\theta) \, d\theta$$

$$b_k = \frac{1}{\pi} \int_0^{2\pi} f(\theta) \cos(k\,\theta) \, d\theta$$

with $k = 0, 1, 2...$, provided that these integrals exist, for which reason $f(\theta)$ must be [piecewise continuous](#) or [square-integrable](#). If the length of the interval of periodicity be L rather than 2π rad, the corresponding functions become

$$F(x) = \frac{a_0}{2} + \left(\sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi x}{L}\right) \right) + \left(\sum_{k=1}^{\infty} b_k \sin\left(\frac{k\pi x}{L}\right) \right)$$

$$a_k := \int_{-\frac{L}{2}}^{\frac{L}{2}} \frac{2\,f(x)\cos\left(\frac{2k\pi x}{L}\right)}{L}\,dx$$

$$b_k := \int_{-\frac{L}{2}}^{\frac{L}{2}} \frac{2\,f(x)\sin\left(\frac{2k\pi x}{L}\right)}{L}\,dx$$

In either case the domain of integration must be one period, whether $0 \mathinner{..} 2\pi$ or $-\frac{\pi}{2} \mathinner{..} \frac{\pi}{2}$ in radians, or $0 \mathinner{..} L$ or $-\frac{L}{2} \mathinner{..} \frac{L}{2}$ otherwise. Coefficient a_0 might be evaluated also on application of [l'Hopital's rule](#). For an odd formula or function, the corresponding Fourier series comprises only the sine terms, whereas for an even function only cosine terms; for a general formula or function that is neither even nor odd, both cosine and sine terms appear in the sums.

In exponential form, the expansion,

$$F(x) = \sum_{k=-\infty}^{\infty} c_k\,e^{\left(\frac{2Ik\pi x}{L}\right)}$$

and coefficients c_k become evaluated with these integrals.

$$c_k := \int_{-\frac{L}{2}}^{\frac{L}{2}} \frac{2\,f(x)\,e^{\left(\frac{2kI\pi x}{L}\right)}}{L}\,dx$$

The correlations between c_k and a_k or b_k are $a_k = c_{+k} + c_{-k}$ and $b_k = i\,(c_{+k} - c_{-k})$. If periodic function $f(x)$ have a [bounded](#) second derivative, its Fourier series [converges](#) absolutely.

For Fourier integrals in exponential form, the orthogonality is defined in terms of the complex conjugate of one of the two factors, with integer j and k,

$$\int_0^{2\pi} \overline{(e^{(jIx)})} \, e^{(kIx)} \, dx \; = 2\,\pi \; \text{if } j = k, \; 0 \text{ otherwise}$$

and analogously for integration over a domain of length L.

Extending the domain from one period over the entire real line implies conversion of Fourier sums into Fourier integrals, under appropriate conditions of convergence,

$$f(x) = \int_0^\infty a(s)\cos(2\,\pi\,s\,x) + b(s)\sin(2\,\pi\,s\,x)\,dx$$

in which, providing that these integrals exist,

$$a(s) \;=\; \int_{-\infty}^\infty f(x)\cos(2\,\pi\,s\,x)\,dx$$

and

$$b(s) \;=\; \int_{-\infty}^\infty f(x)\sin(2\,\pi\,s\,x)\,dx$$

Each unique function or formula $f(x)$, even if not periodic -- such as for a process occurring only once, that is partially monotonic and continuous is uniquely representable as a Fourier integral as a result of Fourier transformation, with a decomposition into a continuous spectrum of frequencies f in the infinite interval $[0, \infty]$.

If a periodic function $f(x)$ have a bounded second derivative, $\dfrac{d^2}{dx^2}f(x)$, its Fourier series converges absolutely, but the inverse statement is invalid. If $|a_j|$ and $|b_j| \le j^{(-\kappa)}$ with $0 < \kappa \le 1$, convergence is at least conditional, and $f(x)$ might have discontinuities; if $\kappa > 1$, there is absolute convergence. For a Fourier series of $f(x)$,

$$fs(x) = \frac{a_0}{2} + \left(\sum_{k=1}^\infty a_k \cos(k\,x)\right) + \left(\sum_{k=1}^\infty b_k \sin(k\,x)\right)$$

on integration, $\displaystyle\int_{x_0}^{x} fs(x)\,dx$, one obtains

$$\frac{a_0(x-x_0)}{2} + \left(\sum_{k=1}^\infty \frac{a_k \sin(k\,x)}{k}\right)\Bigg|_{x=x} - \left(\sum_{k=1}^\infty \frac{a_k \sin(k\,x)}{k}\right)\Bigg|_{x=x_0} - \left(\sum_{k=1}^\infty \frac{b_k \cos(k\,x)}{k}\right)\Bigg|_{x=x} +$$

$$\left(\sum_{k=1}^\infty \frac{b_k \cos(k\,x)}{k}\right)\Bigg|_{x=x_0}$$

which results in a divisor of each term by k so accelerating convergence. A convergent Fourier series is invariably integrable term by term with the resulting series converging uniformly to the integral of the original function; such integration term by term might be valid even if the original series above be not convergent. If $a_0 \ne 0$, the latter series is not a Fourier series, but from

$$\int_{x_0}^{x} \mathrm{f}(x)\, dx - \frac{a_0\, x}{2}$$ one still generates a Fourier series. Differentiation of a Fourier series yields a Fourier series of uncertain properties of convergence that must be tested in each case.

   A great advantage of a Fourier representation over a representation of another kind, such as a Taylor series (cf. sections 2.405 and 3.306), is its ability to represent a discontinuous function, apart from its essential property of representing a periodic function. Although a Taylor series, a Fourier sine series and a Fourier cosine series might prove valid in representing a formula within a particular interval, beyond that interval their behaviours differ markedly. Apart from the representation of electronic density in crystals as Fourier series in three dimensions of which the importance is inestimable, the specifically chemical applications of these series are few: the potential energy of internal rotation in molecules, or torsional vibrations and the spatial arrangement of nucleotides in deoxyribonucleic acid are well established, and some properties of chemical compounds with supposed additivity, such as mass density and enthalpy of formation, have been so treated. In contrast, Fourier transforms, formally based on Fourier series, play a prominent role in measurements on chemical systems, as explained in chapter 14.

   Both Fourier series and Fourier transforms arise in the solution of both ordinary and partial-differential equations with periodic boundary conditions, as discussed in chapter 7.

integration and symbolic computation

   Before programmes for symbolic computation, such as *Maple*, attained their current sophistication, a conventional approach to integration required an extensive knowledge of antiderivatives of common formulae and functions, in conjunction with standard methods to convert a given integral into a more manageable or standard form, using such techniques as partial fractions, substitution and integration by parts. If such manual methods failed to yield rapidly a desired result, one had recourse to books containing tables of integrals, of which several major compilations are well known. A problem with the latter approach is that, about year 1982, it was recognised that such tables of integrals contained errors, whether typographical or from other sources, or printed solutions that were inadequately specified, in significant proportions -- at least 7 per cent in the best tables and even exceeding 20 per cent in other tables! The corresponding tables in subsequently printed editions contain such misleading results in greatly diminished proportion largely because authors have since applied symbolic computation and because automatic rendering of results into typescript eliminates many such errors; such tables might nevertheless lack signs to indicate absolute values about arguments of logarithmic functions. Even a program for symbolic computation, such as *Maple*, is fallible, and might not only fail to provide an algebraic or symbolic antiderivative when one is known to exist, perhaps requiring explicit assumptions about parameters or variables, but also render incorrect results. To assess the latter possibility, commendable practice is to use *Maple* to differentiate a supplied antiderivative, and to simplify a difference between that result and the original integrand to verify each integration.

   In this chapter we explain how to integrate formulae involving a single independent variable with *Maple*, for which a general command is **Int** or **int**; for this purpose we motivate a notion of integration on a geometric basis, and proceed to treat indefinite integration, improper integrals and numerical integration. Integration involving multiple independent variables we consider in

chapter 5.

\>

summary of chapter 4

In this chapter we review or develop principles of definite and indefinite integration with integrals involving a single independent variable, for which animated plots prove illuminating, and demonstrate how with symbolic computation with *Maple*'s functionality such tasks of calculus become reduced to routine use of practically a single command for an integral that has an algebraic solution and a related command for numerical integration. Differentiation discussed in chapter 3 is related to integration discussed in this chapter through a fundamental theorem of calculus: if a function f(x) that is continuous on an interval [a, b] and is a derivative of F(x), such that

$$\frac{d}{dx} F(x) = f(x),$$

that is integrable, so that F(x) is an indefinite integral or antiderivative of f(x),

$$\int_a^b f(x)\, dx = F(b) - F(a)$$

Conversely, if F(x) be defined to be an integral of f(x) from a to x for all x in [a, b], f is a derivative of F at each point of that interval at which f is continuous.

As applications of integrals definite or indefinite, Fourier series we discuss in detail, enabling their use to become a routine procedure.

chapter 5   Calculus with multiple independent variables

5.0 overview and principles

The methods of differential and integral infinitesimal calculus that we explain in chapters 3 and 4 apply directly to functions of, or formulae involving, a single independent variable. In many applications and uses of such functions that we describe in those chapters, we employ extensively command plot to generate a graphic depiction of a relation between the independent and dependent variables, as a line or curve in a space of two dimensions. Many quantities in chemical experiments depend on multiple related variables; to treat these conditions, one must apply differential and integral calculus of multiple variables. For functions of two or three variables, we proceed in this chapter to provide both algebraic and geometric explanations, accompanied naturally by appropriate plots; for functions of variables numbering four or more, direct plots are impracticable, but algebraic methods analogous to those verified with three variables enable us to treat the pertinent systems effectively.

surfaces

In chemistry we work typically with formulae or functions of multiple variables, but we generate a plot of such expressions with at most two independent variables. For instance, for an equation $P = \dfrac{R\,T}{V_m}$ for an ideal gas in which appear three intensive variables pressure P, temperature T and molar volume V_m with gas constant R, we might form an adequate plot in three

<u>dimensions</u>: one spatial <u>axis</u> corresponds to each variable; the totality of all points in a plane containing axes T and V_m, which becomes the <u>domain</u> of $P = f(T, V_m)$ when we consider pressure to be the dependent variable. Because, for these physical variables, only positive values of each are possible, this domain corresponds to the first <u>quadrant</u> of that plane, and the first <u>octant</u> correspondingly represents a bulk region in which acceptable values of pressure are possible. If we express this equation in an alternative form with a fourth variable n for amount of chemical substance, as $P = \dfrac{n\,R\,T}{V}$, plotting a hypersurface in <u>hyperspace</u> that comprises four spatial dimensions is no longer practicable. For expressions containing independent variables numbering more than two, three spatial dimensions that are available are hence generally insufficient to display a concurrent variation of all variables.

   As another instance, in a plot of electronic density about some atomic nuclei, one requires four spatial variables to define such a function -- the value of density and three coordinates (x, y, z), such that a plot of this function requires four dimensions; three-dimensional plots of constant density as contours bear the same relation to such a function that a contour map of rolling countryside in two dimensions does to hills in three dimensions. A further example provokes thought: the motion, relative to a centre of mass, of an atomic nucleus in a molecule benzene requires specification of 30 vibrational coordinates and three rotational coordinates to describe its <u>trajectory</u> in a space comprising 33 formal physical dimensions, with the origin and coordinate axes fixed in the molecule.   On a basis of such considerations, to treat general functions that would require many dimensions to depict graphically, we must extend the mathematical tools that are provided in calculus -- <u>differentiation</u> and <u>integration</u> -- and <u>linear algebra</u> -- <u>matrices</u> and <u>vectors</u>. The former tools we develop in the following sections; tools of linear algebra we discuss in chapter 6, with diverse chemical applications in part II. In all these situations, *Maple* can render great service, because with its deployment one can thereby avoid much intricate manipulation associated with the details of mathematical analysis.

quadric surface

   Among figures that one can plot in three spatial dimensions, of particular interest is a <u>quadric surface</u>, which is a <u>graph</u> of an equation of second <u>degree</u> in three <u>variables</u> x, y and z.   Operations equivalent to <u>rotation</u> and <u>translation</u> on an <u>equation</u> in general form,

$$a\,x^2 + b\,y^2 + c\,z^2 + d\,x\,y + e\,y\,z + f\,z\,x + g\,x + h\,y + j\,z + k = 0$$

 in which z is implicitly a function of x and y, and with parameters $a, b, ..., k$, serve to <u>eliminate</u> <u>linear</u> <u>term</u>s and <u>product</u>s of <u>coordinate</u>s to yield one of two <u>standard forms</u>:

$$A\,x^2 + B\,y^2 + C\,z^2 + K = 0$$

 or

$$A\,x^2 + B\,y^2 + J\,z = 0 \;,$$

or the latter equation with x, y, z interchanged; in these equations majuscules $A .. K$ denote parameters when axes of <u>symmetry</u> coincide with <u>cartesian coordinate</u> axes. Quadric surfaces are analogues in three dimensions of <u>conic section</u>s -- <u>hyperbola</u>, <u>parabola</u> and <u>ellipse</u>, of which a special case of the latter is a <u>circle</u> -- in two dimensions. A <u>curve</u> that an <u>intersection</u> of a quadric <u>surface</u> makes with a <u>plane</u> <u>parallel</u> to a coordinate plane is called a <u>trace</u> or *cross section* of that

surface. Among quadric surfaces that are plotted in section 2.205, we elaborate here on the following selection.

- An ellipsoid, which is a graph of

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

is symmetric about each of three planes of coordinates x, y and z in various couples, and has intercepts ($\pm a$, 0, 0), (0, $\pm b$, 0) and (0, 0, $\pm c$) along the respective axes x, y and z. Each trace of this ellipsoid in a plane parallel to a coordinate plane is either a single point or an ellipse; an ellipsoid is accordingly a surface of revolution of an ellipse in a coordinate plane rotated about a coordinate axis in that plane. A special case arises if $a = b = c$, equal also to unity in standard form, which generates a sphere that is obviously a circle of revolution about any axis within a plane containing the centre of that circle.

- An elliptic paraboloid that represents a parabola of revolution that produces a surface in three dimensions, for which a defining equation is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2\,c\,z$$

has an ellipse as trace in a horizontal plane but parabola as trace in two perpendicular vertical planes, whereas for an *elliptical cone*, a defining equation is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = \frac{z^2}{c^2}$$

- For an elliptic hyperboloid of one sheet, a defining equation is

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1$$

whereas for an elliptic hyperboloid of two sheets, a defining equation is

$$\frac{z^2}{c^2} - \frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

When a right side of an equation defining either hyperboloid of one or two sheets is zero instead of unity or a constant, the resulting figure is a cone.

- An hyperbolic paraboloid, with defining equation

$$\frac{y^2}{b^2} - \frac{x^2}{a^2} = 2\,c\,z$$

and which has parabolic traces in two vertical coordinate planes but an hyperbolic trace in a horizontal coordinate plane, shows a well defined col: a point that is a local maximum in one direction but a local minimum in another direction, also known as a saddle point, a term originated by G. N. Watson. When an equation of a quadric surface lacks one variable x or y or z, the surface becomes an elliptic cylinder. Just as one generates a circle or hyperbola with appropriate trigonometric functions in two dimensions, one likewise generates these quadric surfaces in three dimensions.

partial, directional and total derivative

In a particular case of three variables to describe some surface, such as cartesian coordinates x, y, z, we generally take z as the dependent variable and x and y as the independent variables, so $z(x, y)$ or $z = f(x, y)$; all points in plane xy for which $f(x, y)$ is defined then become the domain of that formula $f(x, y)$. If for each point (x, y) in plane xy we plot a point $f(x, y)$ units above that plane, we generate a surface, such as those depicted in section 2.205. Each point on this surface has coordinates (x, y, z) that satisfy an equation $z = f(x, y)$, which becomes the equation of that surface. Just as a curve is a pictorial representation of a function or its formula $f(x)$ in two spatial dimensions, a surface is a pictorial representation of a function $f(x, y)$ in three spatial dimensions. Although a function $f(x, y)$ of two independent variables x, y has a geometric representation as a surface, not every surface represents a function $f(x, y)$; if and only if every vertical line, i.e. in direction z, that intersects the surface intersects it at exactly one point, that surface represents that function $f(x, y)$.

In section 3.202, we describe a derivative $\dfrac{dy}{dx}$ of a function $y = f(x)$ of a single independent variable x at a point with abscissal coordinate x_0 in terms of a limit of a quotient as $\Delta x \to 0$.

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x)}{\Delta x}$$

In this context, symbol Δx in both denominator and numerator signifies an increment of variable x from its value at x_0, and the entire numerator analogously represents a corresponding increment Δy in dependent variable y in response to that increment Δx in x, according to a functional relation in a formula $f(x)$ named y. As we recall from section 3.202, derivative $\dfrac{dy}{dx}$ represents accordingly a limit of a ratio of increments as Δx tends to zero, or

$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

We interpret $\dfrac{dy}{dx}$ here as a single quantity representing an expression obtained from y on differentiation with respect to x; we recall also from section 3.501 that that quantity implies also a ratio of differentials dy and dx of which we might make separate use. This exact derivative $\dfrac{dy}{dx}$ is represented in *Maple* sometimes as $\dfrac{dy}{dx}$, sometimes as $\dfrac{\partial}{\partial x} y$, somewhat unpredictably; the correct notation within such a derivative has 'd' in roman font and 'x' and 'y' in italic font, but such notation is not generally implemented in *Maple*. The slope or gradient of a curve pertaining to a formula $f(x)$ at a particular point with abscissa $x = x_0$ is precisely a derivative of that formula evaluated at that abscissal value, which we write as $f'(x_0)$ or $\left(\dfrac{d}{dx} f(x) \right)\Big|_{x = x_0}$.

Just as for one independent variable, the concept of a limit is crucial for a definition of a derivative in multiple dimensions. Function f with formula $f(x, y)$ and name z has limit L,

$$\lim_{x \to x[0], y \to y[0]} f(x, y) = L$$

as x approaches x_0 and y approaches y_0 if for given $\varepsilon > 0$ there exist $\delta > 0$ such that $\left| f(x, y) - L \right| < \varepsilon$ when

$$0 < \sqrt{(x - x_0)^2 + (y - y_0)^2} < \delta$$

and (x, y) is in the domain of $f(x, y)$. Function f with formula $f(x, y)$ is continuous at point (x_0, y_0) if

$$\lim_{x \to x[0], y \to y[0]} f(x, y) = f(x_0, y_0).$$

Analogous to a derivative of a function of a single independent variable, of which a geometric interpretation is a slope of a [line](#) [tangent](#) to a [curve](#) in two spatial dimensions, we define formally a first [partial derivative](#) of a formula pertaining to a function $f(x, y)$ of two independent variables x and y with respect to x at a point (a, b) as a limiting rate of change of the value of this function in the direction of x as x is incremented from value a by amount Δx, and y is kept [constant](#) at [value](#) b.

$$f_x(a, b) = \lim_{\Delta x \to 0} \frac{f(a + \Delta x, b) - f(a, b)}{\Delta x}$$

The corresponding partial derivative $f_y(a, b)$ along a line parallel to axis y at $x = a$ is

$$f_y(a, b) = \lim_{\Delta y \to 0} \frac{f(a, b + \Delta y) - f(a, b)}{\Delta y}$$

To evaluate the former partial derivative $f_x(a, b)$ with respect to x, we [differentiate](#) function $f(x, b)$ with respect to x and then evaluate this ordinary derivative at $x = a$; thus

$$f_x(a, b) = \left(\frac{\partial}{\partial x} f(x, b) \right)\bigg|_{x = a}.$$

By either $\dfrac{\partial}{\partial x} f(x, y)$ (preferably) or $f_x(x, y)$, we denote a partial derivative of function $f(x, y)$ with respect to x, implying an ordinary derivative of $f(x, y)$ with respect to x with y treated as a constant; analogously, $\dfrac{\partial}{\partial y} f(x, y)$ (preferably) or $f_y(x, y)$ denotes a partial derivative of $f(x, y)$ with respect to y. Just as, for functions of a single variable, we express a derivative as $\dfrac{dy}{dx}$ in text form, or $\dfrac{d}{dx} f(x)$, in equivalent *Maple* form, with $f'(x)$, we have a further notation to name partial derivatives. Following mathematical convention, if we designate by, and assign to, z the value $f(x, y)$ at a location (x, y), the first partial derivative of $f(x, y)$ with respect to x takes a form $\left(\dfrac{\partial}{\partial x} z \right)_y$, in which the variable in the subscript indicates that y is held constant during differentiation of the expression of z with respect to x. Accordingly, the first partial derivative with respect to y, treating x as a constant, is given by $f_y(x, y)$ or $\dfrac{\partial}{\partial y} f(x, y)$ or $\left(\dfrac{\partial}{\partial y} z \right)_x$. As the latter notation is cumbersome in *Maple*, and as the same operator [diff](#) serves to differentiate functions of

variables of any number, we prefer to use multivariate functions expressed in arrow form rather than a formula such as z. Extension of these concepts to independent variables numbering more than two is analogous: the subscripts on $\left(\dfrac{\partial}{\partial y} z\right)_x$ would then accordingly number more than one,

as in $\left(\dfrac{\partial}{\partial x_1} z\right)_{x_2, x_3, \ldots}$ et cetera. Although a derivative, or differential quotient, of a formula

involving one independent variable, such as $\dfrac{\partial}{\partial x} y$, is equally considered to be a ratio or quotient of two differential quantities dy and dx, the directly analogous situation with a multivariate formula is inapplicable. The existence of partial derivatives of a formula or function with respect to the multiple independent variables at a particular point is a necessary condition for the differentiability of the formula at that point. A formula or function of multiple variables is differentiable at a point when it has partial derivatives not only at that point but also in a neighbourhood of that point, and when these are continuous at the point itself. Formulae and functions arising in applications are generally differentiable at every point of their domains, with possible limited exceptions.

   A directional derivative is precisely a partial derivative for which the direction of concern be parallel to an axis of the system of coordinates; in another direction this derivative is most readily calculated through a rotation of axes to make the desired direction coincide with that of one axis or other.

   When we evaluate a slope of a line along a surface corresponding to a gradient of a formula arising from an application of a function to two independent variables x and y at a point $x = a$, $y = b$, we calculate

$$\left(\frac{\partial}{\partial x} f(x, y)\right)\bigg|_{x = a,\, y = b} = f_x(x, y),$$

as above. The corresponding slope $f_y(x, y)$ along a line parallel to axis y at $x = a$ is

$$f_y(x, y) = \lim_{\Delta y \to 0} \frac{f(a, b + \Delta y) - f(a, b)}{\Delta y} = \left(\frac{\partial}{\partial y} f(x, y)\right)\bigg|_{x = a}$$

We define the latter derivative also as a ratio of differential quantities as follows. If we assign name z of our dependent variable to our formula $f(x, y)$ of two independent variables x and y, we increment one independent variable x an amount Δx from particular value a, hold another independent variable y constant at b, and find a corresponding response of dependent variable z according to this expression.

$$\Delta z = [f(a + \Delta x, b) - f(a, b)]$$

On dividing by increment Δx in x, we form a ratio:

$$\frac{\Delta z}{\Delta x} = \frac{f(a + \Delta x, b) - f(a, b)}{\Delta x}$$

In the limit as $\Delta x \to 0$,

$$\frac{\partial}{\partial x} z = \lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} = \frac{f(a + \Delta x, b) - f(a, b)}{\Delta x}$$

we define a derivative $\dfrac{\partial}{\partial x} z$ that we call a <u>partial derivative</u> because, in forming this derivative of z with respect to x, we hold constant another independent variable y at a value b; such a *partial derivative* signifies a derivative, or rate of change, of a dependent variable while one independent variable is incremented infinitesimally but while any other independent variable is held constant. According to mathematical convention, a partial derivative might be displayed as

$$\left(\frac{\partial}{\partial x} z \right)\Bigg|_{y=b}$$

in which a derivative of z with respect to x is enclosed within parentheses; following the closing parenthesis, a subscripted expression consists of an equality to specify explicitly any independent variable, here only y, that is held constant, at its value b, while the particular partial derivative with respect to another independent variable x is being evaluated. Such notation is cumbersome in *Maple*: for this reason, in text mode we distinguish a total derivative $\dfrac{dz}{dx}$ from a partial derivative $\dfrac{\partial}{\partial x} z$ for which any independent variable other than x is implicitly held constant. In *Maple* an operator `diff` for differentiation, explained in chapter 3, generates directly a partial derivative, as we demonstrate with examples in succeeding sections, but, when there is only one independent variable, that partial derivative becomes a total derivative.

    The reciprocal identity states that reversal of dependent and independent variables yields a reciprocal derivative,

$$\left(\frac{\partial}{\partial x} z \right)\Bigg|_{y=b} = \frac{1}{\left(\dfrac{\partial}{\partial z} x \right)\Big|_{y=b}}$$

in which the same variables are held constant in both derivatives. For second derivatives with respect to separate independent variables, for z as a function $z(x, y)$ varying smoothly the order of differentiation is immaterial, according to Euler's reciprocal relation,

$$\frac{\partial^2}{\partial y\, \partial x} z = \frac{\partial^2}{\partial x\, \partial y} z$$

The cyclic rule is expressed as

$$\left(\frac{\partial}{\partial x} y \right)_z \left(\frac{\partial}{\partial z} x \right)_y \left(\frac{\partial}{\partial y} z \right)_x = -1$$

The equivalent of the <u>chain rule</u> for an ordinary derivative is

$$\left(\frac{\partial}{\partial x} z \right)_{w,\, v} = \left(\frac{\partial}{\partial y} z \right)_{w,\, v} \left(\frac{\partial}{\partial x} y \right)_{w,\, v}$$

in which the same variables w, v are held fixed in all three partial derivatives.

    For the evaluation of definite integrals, a rule attributed to Leibnitz might be useful:

$$\frac{\partial}{\partial y}\left(\int_a^b f(x, y)\, dx\right) = \int_a^b \frac{\partial}{\partial y} f(x, y)\, dx$$

for $y_1 < y < y_2$ when the two real formulae $f(x, y)$ and $\frac{\partial}{\partial x} f(x, y)$ are continuous in the closed interval $[a, b]$ for x and $[y_1, y_2]$ for y. This equation is valid also for improper integrals, such as for $b = \infty$, as long as $f(x, y)$ and $\frac{\partial}{\partial x} f(x, y)$ are continuous in the corresponding domain with extra conditions about the <u>uniform convergence</u> of the integral on the right side.

geometric interpretation of partial derivative and stationary points

On a surface that is a graph of a formula $z = f(x, y)$, partial derivatives $f_x(x, y)$ and $f_y(x, y)$ are hence slopes of lines tangent to certain curves, specifically those curves that lie in planes parallel to planes containing axes y and z for derivative $f_x(x, y)$, but containing axes y and z for derivative $f_y(x, y)$. Provided that that formula, $f(x, y)$, has continuous partial derivatives on a rectangle in plane xy containing a point (x_0, y_0) in its interior, these two tangent lines define a plane tangent to the surface at a point $(x_0, y_0, f(x_0, y_0))$. To find an equation of this tangent plane at that point, we recall that a typical plane non-vertical in space that passes that point has an equation of form

$$A(x - x_0) + B(y - y_0) + C(z - f(x_0, y_0)) = 0.$$

For a plane to be tangent to a surface at a point $(x, y, f(x_0, y_0))$, the values of coefficients A, B and C must conform to

$$A = f_x(x, y),\ B = f_y(x, y)\ \text{and}\ C = -1;$$

an equation for a tangent plane at that point on a surface is

$$z = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

A line passing that point and normal to that plane lies at an intersection of two planes, defined according to these relations:

$$\frac{x - x_0}{f_x(x_0, y_0)} = \frac{y - y_0}{f_y(x_0, y_0)} = f(x_0, y_0) - z = t$$

For purpose of plotting that <u>normal</u> line, we use a <u>parametric</u> form in terms of a further variable t, to which each expression in the preceding line is equal; the corresponding equations defining that line are thus

$$x = x_0 + f_x(x_0, y_0)\, t,$$
$$y = y_0 + f_y(x_0, y_0)\, t$$

and

$$z = f(x_0, y_0) - t.$$

Conditions follow according to which one can distinguish <u>stationary points</u> of a function $f(x, y)$ evaluated at (a, b), thus

$$f(x, y)\big|_{x = a,\, y = b} = f(a, b),$$

for which first partial derivatives are zero, i.e. $f_x(a, b) = 0$ and $f_y(a, b) = 0$, and with second derivatives $f_{xx}(a, b)$, $f_{yy}(a, b)$ and $f_{xy}(a, b)$ at that location:

- if $f_{xx}(a, b) < 0$ and $f_{yy}(a, b) < 0$, (a, b) is a local <u>maximum</u> or col of formula $f(x, y)$;

- if $f_{xx}(a, b) > 0$ and $f_{yy}(a, b) > 0$, (a, b) is a local <u>minimum</u> or col of formula $f(x, y)$;

- if $f_{xx}(a, b) = 0$, (a, b) is a <u>point of inflexion</u> of function $f(x, y)$ in a profile of the surface at $y = b$, or if $f_{yy}(a, b) = 0$, (a, b) is a point of inflexion of formula $f(x, y)$ in a profile of the surface at $x = a$, but further tests are required to indicate whether point (a, b) is a col for the surface; a point of inflexion is defined only for a curve in two dimensions, not for a surface in multiple dimensions;

- the value of a quantity that we might name $coltest = f_{xx}f_{yy} - f_{xy}^{2}$ with all second derivatives evaluated at stationary point (a, b) serves to confirm its nature; a negative value of $coltest$ indicates a presence of a col at that point.

Hence, for a formula of two independent variables to have a minimum at some point (a, b), its first partial derivatives are zero there but the second partial derivatives are positive; for a maximum the first partial derivatives are zero but the second partial derivatives are negative, whereas for a col the first partial derivatives are zero but the value of a formula for $coltest$ above, containing second partial derivatives, is negative. In chapter 6, we recognize that combination of quantities in $coltest$ to be characteristic of a <u>determinant</u> (section 6.101) of a particular <u>symmetric matrix</u> called a <u>hessian</u> (section 6.405), evaluated for a function with specified partial derivatives. Among such <u>extrema</u> or stationary points on this surface pertaining to formula $f(x, y)$, there might be maxima, minima or cols that one can locate on finding points at which first derivatives are zero; in these circumstances, one explores the neighbourhood of each stationary point to deduce its characteristic features.

At such an extremum -- either a minimum or maximum value of a dependent variable in multidimensional space, its location is evaluated through a solution of simultaneous equations obtained on setting first derivatives equal to zero: for an extremum of $z(x, y)$ in three dimensions, accordingly

$$\left(\frac{\partial}{\partial x}z\right)_{y} = 0 \quad \text{and} \quad \left(\frac{\partial}{\partial y}z\right)_{x} = 0$$

The nature of that extremum is discovered through evaluation of $coltest$ as specified above.

total differential

For a <u>smooth</u> function $f(x)$, named z, of one variable, x, according to 3.501 we express a differential of the dependent variable as a function of the slope of the curve of $f(x)$ at any point x_0:

$dz = \left(\frac{d}{dx}f(x)\right)_{x_0} dx$, in which derivative $\frac{d}{dx}f(x)$ evaluates to precisely that slope $\frac{\partial}{\partial x}z$ at any given point. An important consideration in chemistry, in particular for thermodynamics, relates to properties of differentials of formulae or functions of multiple variables. For a function f with a formula named $z = f(x_1, x_2)$ of two independent variables x_1, x_2, we write a <u>total differential</u> analogously as a sum of products involving partial derivatives,

$$dz = \left(\frac{\partial}{\partial x_1} z\right)_{x_2} dx_1 + \left(\frac{\partial}{\partial x_2} z\right)_{x_1} dx_2 \,,$$

with an obvious extension for a formula pertaining to multiple independent variables $z = f(x_1, x_2, x_3, \dots)$. As those two partial derivatives are, in general, functions of both independent variables, we rewrite the above expression in a form

$$dz = f(x_1, x_2)\, dx_1 + g(x_1, x_2)\, dx_2$$

in which $f(x_1, x_2)$ and $g(x_1, x_2)$ are functions of x_1 and x_2. The <u>chain rule</u> applies for partial differentiation in a form analogous to that for differentiation of a single independent variable; for a function $f(x, y, z)$ of which arguments x, y, z are functions of parameters u, v,

$$\frac{\partial}{\partial u} f(x, y, z) = \left(\frac{\partial}{\partial x} f(x, y, z)\right)\left(\frac{\partial}{\partial u} x\right) + \left(\frac{\partial}{\partial y} f(x, y, z)\right)\left(\frac{\partial}{\partial u} y\right) + \left(\frac{\partial}{\partial z} f(x, y, z)\right)\left(\frac{\partial}{\partial u} z\right)$$

$$\frac{\partial}{\partial v} f(x, y, z) = \left(\frac{\partial}{\partial x} f(x, y, z)\right)\left(\frac{\partial}{\partial v} x\right) + \left(\frac{\partial}{\partial y} f(x, y, z)\right)\left(\frac{\partial}{\partial v} y\right) + \left(\frac{\partial}{\partial z} f(x, y, z)\right)\left(\frac{\partial}{\partial v} z\right)$$

Because reversing this process does not necessarily result in function z, given an expression involving differentials of form

$$dz = p(x_1, x_2)\, dx_1 + q(x_1, q_2)\, dx_2$$

we require to discover whether z can be constructed from $p(x_1, x_2)$ and $q(x_1, x_2)$: if so, dz is termed an <u>exact differential</u> or <u>total differential</u>, otherwise, an *inexact* differential. Comparing dz in its two forms above, we deduce that a requirement is that

$$p(x_1, x_2) = \left(\frac{\partial}{\partial x_1} z\right)_{x_2} \quad \text{and} \quad q(x_1, x_2) = \left(\frac{\partial}{\partial x_2} z\right)_{x_1}$$

On partial differentiation of these equations with respect to x_2 and x_1, respectively, we find that both left sides are equal to the mixed partial derivative of z of second order; a requirement that dz be an exact differential is hence

$$\frac{\partial}{\partial x_2} p(x_1, x_2) = \frac{\partial}{\partial x_1} q(x_1, x_2)\,.$$

On solving these equations $z(x_1, x_2)$ might be found, even if a functional relation between x_1 and x_2 be unknown.

For a <u>positively</u> <u>homogeneous</u> formula of degree *n*, for which, for all $t > 0$, $f(t\,x, t\,y, t\,z) = t^n\, f(x, y, z)$, that has <u>continuous</u> first partial derivatives, this relation holds:

$$x\left(\frac{\partial}{\partial x} f(x, y, z)\right) + y\left(\frac{\partial}{\partial y} f(x, y, z)\right) + z\left(\frac{\partial}{\partial z} f(x, y, z)\right) = n\, f(x, y, z)\,.$$

.

Exact or total differentials are important because the <u>integral</u> of an <u>exact differential</u> between an initial condition (x_i, y_i), or a corresponding point on a surface of a property, and a final condition, corresponding to a point (x_f, y_f), is independent of a path of integration; for an *inexact*

differential, such an integral depends on a path. Functions of a thermodynamic state, such as energy and entropy, produce exact differentials, unlike functions such as work or heat that correspond to operations to alter a thermodynamic state. For an inexact differential, denoted δz, we might express its relation to differentials dx and dy as

$$\delta z = M(x, y)\, dx + N(x, y)\, dy$$

By finding an integrating factor one might in some cases make an inexact differential exact: for instance, if some integrating factor $\lambda(x, y)$ exist, the above relation becomes

$$dz = \lambda(x, y)\, \delta z = \lambda(x, y)\, M(x, y)\, dx + \lambda(x, y)\, N(x, y)\, dy$$

An integrating factor in a particular case is not unique, but if one exist, alternative integrating factors of infinite number also exist; in practice the finding of such an integrating factor might be difficult because there is no general method for that purpose. In thermodynamics, for a change δq of thermal energy or heat, which is inexact, an integrating factor $\dfrac{1}{T}$ makes a resulting formula $\dfrac{\delta q}{T}$ become differential dS of state function entropy, S. If in an application, such as in thermodynamics, an integral be taken around a closed loop, so that initial and final points be the same, the integral is zero if the differential involved be exact, and non-zero otherwise; in mechanics, such a zero value of integral can define a conservative system. Integrating factors arise also in the solution of differential equations, as discussed in chapter 7.

The *tangent approximation*, or *incremental approximation*, as discussed in section 3.501, is equally applicable to multiple independent variables -- but, as with one independent variable, direct differentiation and evaluation avoids errors of that approximation -- and to multiple differentiation.

partial derivatives in the complex plane

A complex function with this property of differentiability throughout some domain D is called an analytic function in D; if that domain of differentiability extend throughout the finite complex plane, the function is called entire. To assess the condition that this definition imposes on $u(x, y)$ and $v(x, y)$ from $w = f(z) = f(x + iy) = u + iv$, we express the difference quotient in terms of u and v, setting $z = x + iy$ and $\Delta z = \alpha + \beta i$, in which α and β are real numbers:

$$\frac{f(z + \Delta z) - f(z)}{\Delta z} = \frac{u(x + \alpha, y + \beta) + i\, v(x + \alpha, y + \beta) - u(x, y) - v(x, y)}{\alpha + i\beta} \qquad *$$

We first let $\Delta z \to 0$ through purely real values, so that $\beta = 0$, yielding

$$\frac{dw}{dz} = \lim_{\alpha \to 0} \frac{u(x + \alpha, y) - u(x, y)}{\alpha} + \frac{i\,(v(x + \alpha, y) - v(x, y))}{\alpha} = \left(\frac{\partial}{\partial x} u\right) + i\left(\frac{\partial}{\partial x} v\right)$$

We proceed to deduce the form of $\dfrac{dw}{dz}$ on letting $\Delta z \to 0$ through purely imaginary values, so that $\alpha = 0$; an argument of the same form yields

$$\frac{dw}{dz} = \frac{\left(\dfrac{\partial}{\partial y} u\right) + i\left(\dfrac{\partial}{\partial y} v\right)}{i}$$

If a limit of the difference quotient exist in the prepenultimate expression *, it must be unique, which requires the preceding two expressions on the right side to be the same result in alternative

forms. Equating real and imaginary parts, we then obtain two equations that must be satisfied simultaneously by the real and imaginary parts of f(z):

$$\frac{\partial}{\partial x} u = \frac{\partial}{\partial y} v \quad \text{and} \quad \frac{\partial}{\partial y} u = -\left(\frac{\partial}{\partial x} v\right).$$

A differentiable or analytic function $w = u + i\,v$ must thus satisfy these relations attributed to Cauchy and Riemann; the converse is also true -- that, if a complex function satisfy these conditions, it has a unique derivative. These conditions apply to only a function of z, i.e. $w = f(z) = f(x + i\,y)$, not of $\bar{z} = x - i\,y$. Any such function, with real and imaginary parts satisfying the latter relations, is called an analytic function of complex variable z. Any analytic function of z can represent a two-dimensional field, for instance, an electrostatic field, and it can be created by taking a well behaved function of a real variable and making it a function of $z = x + i\,y$, such as $\sin(x + i\,y)$, $\dfrac{1}{(x + i\,y)^2 + c^2}$ or $\ln(x + i\,y)$ that are all analytic functions of z at locations at which the functions do not become infinite.

A point at which $w = f(z)$ is not analytic is called a singularity of f(z); for example, for $w = \dfrac{1}{z - 1}$, this f(z) is analytic everywhere except at a point $z = 1$, which is hence a singularity. If f(z) be an analytic function, when z is purely real, the forms of f(z) and f(x) become identical; we hence deduce the form of f when expressed as $w = f(z) = u(x, y) + i\,v(x, y)$ according to a rule that an analytic function $w = u(x, y) + i\,v(x, y)$ becomes expressible in terms of z on setting $y = 0$ in the right side and then replacing x by z. For this reason $w = z\,\bar{z}$ is not an analytic function because application of that rule would yield $w = f(z) = z^2$, which is incorrect for a formula comprising an imaginary part. Analogous to the preceding rule, if f(z) $= u + i\,v$ satisfies the Cauchy-Riemann equations, the derivative $f'(z) = \dfrac{\partial}{\partial z} f$ is obtainable from the result

$$\frac{d}{dz} f(z) = \left(\frac{\partial}{\partial x} u\right) + i\left(\frac{\partial}{\partial x} v\right)$$

on formally setting $y = 0$ and replacing x by z. If second mixed derivatives $\dfrac{\partial^2}{\partial y\,\partial x} u$ and $\dfrac{\partial^2}{\partial y\,\partial x} v$ exist and be continuous, partial differentiation of the Cauchy-Riemann equations above yields

$$\left(\frac{\partial^2}{\partial x^2} u\right) + \left(\frac{\partial^2}{\partial y^2} u\right) = 0 \quad \text{and} \quad \left(\frac{\partial^2}{\partial x^2} v\right) + \left(\frac{\partial^2}{\partial y^2} v\right) = 0 .$$

These equations have a form identical to Laplace's equation, which is an important partial-differential equation of which any solution in two dimensions is called a harmonic function ; harmonic functions u and v associated with an analytic function $w = f(z) = u + i\,v$ are called conjugate harmonic functions. For example, $\sin(z) = \sin(x)\cosh(y) + \cos(x)\sinh(y)$ is an analytic function, because, with $u = \sin(x)\cosh(y)$ and $v = \cos(x)\sinh(y)$ for which $\dfrac{\partial}{\partial x} u = \dfrac{\partial}{\partial y} v$

and $\dfrac{\partial}{\partial y} u = -\left(\dfrac{\partial}{\partial x} v\right)$, both $\dfrac{\partial^2}{\partial y\,\partial x} u$ and $\dfrac{\partial^2}{\partial y\,\partial x} v$ are continuous; these functions u and v are hence conjugate harmonic functions because they are the real and imaginary, respectively, parts of the same analytic function f(z). Because a complex number $z = x + i\,y$ is expressible in polar form $z = r\,\mathbf{e}^{(i\,\theta)}$, the Cauchy-Riemann equations in polar form become

$$\frac{\partial}{\partial r} u = \frac{1}{r}\frac{\partial}{\partial \theta} v \quad \text{and} \quad \frac{1}{r}\frac{\partial}{\partial \theta} u = -\left(\frac{\partial}{\partial r} v\right).$$

multiple integrals

Because much of physical and theoretical chemistry, especially statistical thermodynamics, spectrometry and quantum theory, is concerned with functions of two or more variables, we must be able to apply methods of calculus within such areas to extract useful information about properties of a system of interest.

Whereas for a continuous function f with formula f(x) defined over an interval $[a, b]$ of x, a definite integral is defined as a Rieman sum

$$\int_a^b \text{f}(x)\,dx = \lim_{n \to \infty} \sum_{i=1}^{n} f_i \,\Delta x$$

in which x_i is a point in subdivision i of $[a, b]$, provided that this limit exist, we extend this concept to a function f with formula f(x, y) of two independent variables, defined over a rectangular region $a \le x \le b$ and $c \le y \le d$, by subdividing this region into n subregions using lines parallel to axes x and y, so to divide the total region R into small portions of area $\Delta A_{j,k} = \Delta x_j \,\Delta y_k$. In each subdivision we select a point $(x_{j,k}, y_{j,k})$ and compute a Riemann sum,

$$S_n = \sum_{j,k=1}^{n} \text{f}(x_{j,k}, y_{j,k})\,\Delta x_j\,\Delta y_k$$

We take a limit as $n \to \infty$ of this sum and with each $\Delta x_j \dashrightarrow 0$ and each $\Delta y_k \dashrightarrow 0$; if this limit exist, it constitutes a double integral of f over the region R, denoted as

$$\int_R \text{f}(x, y)\,dA = \iint_R \text{f}(x, y)\,dx\,dy = \lim_{n \to \infty} \sum_{j,k=1}^{n} \text{f}(x_{j,k}, y_{j,k})\,\Delta x_j\,\Delta y_k$$

so that geometrically this double integral represents a sum of products of the form $\text{f}(x_{j,k})\,\Delta x_j\,\Delta y_k$. A triple integral over volume V with three independent variables x, y, z is analogously defined:

$$\int_V \text{f}(x, y, z)\,dV = \iiint_V \text{f}(x, y, z)\,dx\,dy\,dz = \lim_{n \to \infty} \sum_{j,k,l=1}^{n} \text{f}(x_{j,k,l})\,\Delta x_j\,\Delta y_k\,\Delta z_l$$

The mean value <f> of a function f(x, y) of two independent variables over an region R is

$$<f> = \frac{\displaystyle\iint_R \text{f}(x, y)\,dx\,dy}{\displaystyle\iint_R dx\,dy}$$

which is equivalent to a volume divided by an area.

In a chemical context, an evaluation of definite integrals with multiple variables recurs, simply because ranges of integration variables are fixed for a particular system. A conventional way to proceed in a manual calculation involves integrating sequentially over each variable, applying appropriate boundary conditions or limits at each stage. When a function to be integrated separates into a product of functions of a single variable, as with

$$h(x, y) = f(x) \ g(y) \ .....,$$

integration is straightforward, for example,

$$\int h(x, y) \ dx \ dy = \int f(x) \ dx \int g(y) \ dy$$

providing that each latter integral is tractable somehow -- algebraically or numerically. In other cases, in which an integrand has a form $f(x, y)$ that fails so to factor, a standard method to proceed is to integrate first over one variable, either x or y, and then over a remaining variable to produce an ultimate result. Although this result is independent of choice of which integrating variable we integrate first, when proceeding by hand one might find performing integrations using one choice for an initial integration easier than an alternative approach; when working with *Maple*, one must designate a particular sequence of integration. When $h(x, y)$ or an analogous formula involving multiple variables appears in an integrand of an integral with only constants as bounds for each variable, the success of an integration might depend on the order of nested integrating variables.

For differentiation of an integral with multiple functions $f(t)$, $u(x)$ and $v(x)$ as integrand and bounds of integration, respectively,

$$\frac{\partial}{\partial x} \left(\int_{u(x)}^{v(x)} f(t) \ dt \right) = f(v) \left(\frac{d}{dx} v(x) \right) - f(u) \left(\frac{d}{dx} u(x) \right).$$

If $\dfrac{\partial}{\partial x} f(x, t)$ be continuous on a rectangle for which $a \le y \le b$ and $c \le x \le d$, and if

$\left| \dfrac{\partial}{\partial x} f(x, t) \right| \le g(t)$, for which $\displaystyle\int_a^b g(t) \ dt$ converges,

$$\frac{\partial}{\partial x} \left(\int_a^b f(x, t) \ dt \right) = \int_a^b \frac{\partial}{\partial x} f(x, t) \ dt$$

analogous to formulae in overview 4.0. The latter formulae are combined into a rule attributed to Leibnitz, for which if a partial derivative $\dfrac{\partial}{\partial x} f(x, t)$ be continuous on an area bounded by curves $y = u(x)$, $y = v(x)$, $x = \alpha$ and $x = \beta$,

$$\frac{\partial}{\partial x} \left(\int_{u(x)}^{v(x)} f(x, t) \ dt \right) = f(x, v(x)) \left(\frac{d}{dx} v(x) \right) - f(x, u(x)) \left(\frac{d}{dx} u(x) \right) + \int_{u(x)}^{v(x)} \frac{\partial}{\partial x} f(x, t) \ dt \ .$$

This rule is useful to manipulate solutions of differential equations that are represented as definite integrals.

Just as a geometric interpretation of a definite integral of a formula or function with a single

independent variable, or in general involving two variables as coordinates in space, is an _area_, a definite integral of a formula or function involving three variables as coordinates in space generates a _volume_. For three dimensions x, y and z, with $z = f(x, y)$, a product $dx\ dy$ of differentials is equivalent to an element of area dA, and $f(x, y)$ is the corresponding integrand, yielding an integral

$$\iint f(x, y)\ dy\ dx;$$

for four dimensions x, y, z, w and $w = f(x, y, z)$, a product $dx\ dy\ dz$ of differentials is equivalent to an element of volume dV and the corresponding integrand is $f(x, y, z)$:

$$V = \iiint f(x, y, z)\ dz\ dy\ dx$$

Rather than a volume, a _double integral_ can yield an area if the integrand is unity; for instance for a right triangle with base along axis x and a line of formula $y = f(x)$ passing the origin as another side, the area of the triangle between $x = 0$ and $x = a$ is

$$A = \int_0^a \int_0^{f(x)} 1\ dy\ dx ,$$

and analogously for other planar geometrical situations; this situation is evidently a special case of an area of a body of unit stature or thickness being numerically equal to the volume of the same body. Likewise, a volume can result from a constant integrand and integration for some formula $f(x, y)$ with integrating element $dx\ dy\ dz$, with corresponding results for higher dimensions.

   If the bounds of the region to be integrated to find the area be not segments of straight lines, the corresponding double integral for the area of a region in a plane might be more convenient in polar coordinates (r, θ), in which case the double integral has the form

$$A = \iint f(r \cos(\theta), r \sin(\theta))\ r\ dr\ d\theta$$

or with order of integration reversed as in

$$A = \iint f(r \cos(\theta), r \sin(\theta))\ r\ d\theta\ dr$$

whichever be more convenient.

   In cylindrical coordinates (r, θ, z), the volume is the result of this triple integral,

$$V = \iiint f(r \cos(\theta), r \sin(\theta))\ r\ dz\ dr\ d\theta$$

with integrating element $r\ dz\ dr\ d\theta$ and with integrations implemented in the most convenient order, whereas in spherical polar coordinates (r, θ, ϕ) the volume is obtained from

$$V = \iiint f(r \sin(\phi) \cos(\theta), r \sin(\phi) \sin(\theta), r \sin(\phi))\ r^2 \sin(\phi)\ dz\ dr\ d\theta$$

in which the integrating element is $r^2 \sin(\phi)\ dr\ d\theta\ d\phi$.

   According to a customary _parametric_ representation with $x(t)$, $y(t)$, $z(t)$, the _length_ of an _arc_ along a _curve_ in _space_ is

$$s = \int_a^b \sqrt{\left(\frac{d}{dt}x(t)\right)^2 + \left(\frac{d}{dt}y(t)\right)^2 + \left(\frac{d}{dt}z(t)\right)^2}\ dt$$

The area of a surface $z = f(x, y)$ is

$$A = \iint \sqrt{1 + \left(\frac{\partial}{\partial x}f(x, y)\right)^2 + \left(\frac{\partial}{\partial y}f(x, y)\right)^2}\ dy\ dx$$

with appropriate bounds or end points for each integration variable; in the latter integral, the area of interest is taken to be projected onto plane xy. Because the area of interest is equally well projected onto plane xz, in which case the area is derivable from this double integral,

$$A = \iint \sqrt{1 + \left(\frac{\partial}{\partial x}f(x, z)\right)^2 + \left(\frac{\partial}{\partial z}f(x, z)\right)^2}\ dz\ dx$$

or onto plane yz, in which case the area is derivable from this double integral,

$$A = \iint \sqrt{1 + \left(\frac{\partial}{\partial y}f(y, z)\right)^2 + \left(\frac{\partial}{\partial z}f(y, z)\right)^2}\ dz\ dy$$

The latter three formulae are merely an extension of a formula for the length of an arc along a curve in a plane considered in section 4.106.

Just as, according to section 4.105, one differentiates a definite integral involving a formula in terms of one independent variable as

$$\frac{\partial}{\partial x}\left(\int_a^b f(x, t)\ dt\right) = \int_a^b \frac{\partial}{\partial x}f(x, t)\ dt$$

which has the effect of interchanging the order of integration and differentiation, for a partial derivative $\frac{\partial}{\partial x}f(x, y)$ of $f(x, y)$ that is continuous on a rectangle for x in $[a, b]$ and y in $[c, d]$, as presented above,

$$\frac{\partial}{\partial x}\left(\int_c^d f(x, y)\ dy\right) = \int_c^d \frac{\partial}{\partial x}f(x, y)\ dy$$

The latter result is a particular case in which one or both bounds of integration might depend on x as in $c(x)$ and $d(x)$; in that case, additional terms arise as follows, as presented in Leibnitz's rule above.

$$\frac{\partial}{\partial x}\left(\int_{c(x)}^{d(x)} f(x, y)\ dy\right) = \int_{c(x)}^{d(x)} \frac{\partial}{\partial x}f(x, y)\ dy + f(x, d(x))\left(\frac{d}{dx}d(x)\right) - f(x, c(x))\left(\frac{d}{dx}c(x)\right)$$

There exist a few theorems to convert single integrals to multiple integrals, and vice versa, that might serve to simplify a particular problem and that might be implemented with *Maple*. Green's theorem converts a line integral over a closed curve into an area. Stokes's theorem relates a surface integral to a line integral, whereas Gauss's theorem relates a triple integral extended over a solid to a surface integral taken over the boundary of this solid; we explain the latter two theorems in section group 6.4 on calculus with vectors.

The fundamental theorem of infinitesimal calculus, expressed as $\displaystyle\int_a^b \frac{d}{dx} f(x)\, dx = f(b) - f(a)$,

signifies that the integral, over interval $a \ldots b$, of the derivative of a formula of a single variable as integrand is evaluated as a difference of the values of that formula at the bounds of that interval. As a version of that fundamental theorem in two dimensions, Green's theorem expresses the double integral of a particular derivative of a formula involving two independent variables; in that manner Green's theorem provides a means to convert a line integral along a curve constituting a closed contour to an integral of area within that region. If smooth, simple, closed curve **C** in plane *xy* be continuous at least piecewise and if region **R** consist of **C** and its interior, an integral of two continuous functions m(x,y) and n(x,y) with also continuous first partial derivatives throughout an open region containing **R** as an integral, in two parts, along that curve

$$\int m(x, y)\, dx + \int n(x, y)\, dy$$

becomes a double integral

$$\iint \left\{ \left(\frac{\partial}{\partial x} n(x, y) \right) - \left(\frac{\partial}{\partial y} m(x, y) \right) \right\}\, dx\, dy$$

as illustrated in the following example. In a space of multiple dimensions, multiple paths might exist between two independent points. A line integral or path integral involves definite integration of a differential according to a particular curve or path, which is performed with separate integration over independent variables expressed in terms of each other. For instance, for $F(x, y)$ an integral of $dF(x, y)$ over a path C with subsidiary functions $M(x, y(x))$ and $N(x(y), y)$, yields a sum of integrals of only one variable,

$$\int_C dF = \int_{x_1}^{x_2} M(x, y(x))\, dx + \int_{y_1}^{y_2} N(x(y), y)\, dy$$

in which a specification of the path C implies not only initial (x_1, y_1) and terminal (x_2, y_2) points but also the functional relations $y(x)$ in the integral with $M(x, y(x))$ and $x(y)$ in the integral with $N(x(y), y)$.

With a line integral involving an exact differential is associated an important theorem: if a differential for integration is exact, the value of a line integral depends on only initial and terminal points or conditions, independent of a path between these points; a value of this integral is then equal to a difference of its values at final and initial points. For example, to evaluate a curvilinear

or line integral $\int_P y^2\, dx + x^2\, dy$ along a closed path P with piecewise continuous and smooth segments from the origin at $x = 0$ horizontally with $y = 0$ to $x = 2$, then from $y = 0$ vertically with $x = 2$ to $y = 2$, and returning to the origin diagonally along a segment of a line with $y = x$, we apply the result of Green's theorem above to yield

$$\int_P y^2\, dx + x^2\, dy = \iint 2\,x - 2\,y\, dx\, dy$$

to which we apply bounds to the ranges of integration,

$$= \int_0^2 \int_0^x 2\,x - 2\,y\, dy\, dx$$

for which we evaluate the inner integral at its bounds,

$$= 2 \int_0^2 \left(\left(x\,y - \frac{y^2}{2} \right)\Big|_{y=x} - \left(x\,y - \frac{y^2}{2} \right)\Big|_{y=0} \right) dx$$

which leaves the outer integral, to be evaluated as follows.

$$= \int_0^2 x^2\, dx = \frac{x^3}{3}\Big|_{x=2} - \frac{x^3}{3}\Big|_{x=0} = \frac{8}{3}$$

Hence the line integral $\int_P y^2\, dx + x^2\, dy$ along the specified path evaluates to $\frac{8}{3}$.

Both Simpson's rule and gaussian quadrature are applicable in multiple dimensions. For an integral

$$I = \int_c^d \int_a^b \mathrm{f}(x, y)\, dx\, dy$$

in two dimensions, we divide the range of x into n intervals each of width h and the range of y into m intervals each of width k, so that $n\,h = b - a$ and $m\,k = d - c$. Applying Simpson's rule, we form

$$I = \frac{h\,k}{9} \sum_{j=1}^{m} \left(\sum_{i=1}^{n} w_{i,\,n}\, w_{j,\,m}\, \mathrm{f}(a + i\,h, c + j\,k) \right)$$

in which values of both $w_{i,\,n}$ and $w_{j,\,m}$ conform to a pattern 1, 4, 2, 4, ..., 4, 2, 4, 1 with $n + 1$ and $m + 1$ members, respectively. For gaussian quadrature with n points in direction x and m points in direction y, the integral is evaluated as

$$I = (b - a)\,(d - c) \left(\sum_{j=1}^{m} \left(\sum_{i=1}^{n} w_{i,\,n}\, w_{j,\,m}\, \mathrm{f}(a + k_{i,\,n}\,(b - a), c + k_{j,\,m}\,(d - c)) \right) \right)$$

in which values of $k_{i,\,n}$ and $k_{j,\,m}$ are abscissae of points along axes x and y respectively, and $w_{i,\,n}$ and $w_{j,\,m}$ are weights of those points.

All these operations extend in an obvious way to functions of three or more independent variables. In a chemical situation we encounter multiple integrals over numerous coordinates. For

example, in a problem to calculate an electronic energy of benzene, involving 42 electrons per molecule, we must integrate an energy function over 126 spatial and up to 84 spin coordinates, although the actual number depends upon a spin state of interest; this integral thus might involve integration over 210 variables. For most real applications we forego a graphical depiction and work with only mathematical tools at our disposal.

series in multiple dimensions

Both Taylor series and Fourier series have their counterparts in three or more dimensions. For a Taylor series of $f(x, y)$ as a formula involving two variables expanded about $(x = a, y = b)$, the result retaining terms to second order is

$$f(x, y) = f(a, b) + \left(\frac{\partial}{\partial x} f(a, b)\right)(x - a) + \left(\frac{\partial}{\partial y} f(a, b)\right)(y - b)$$

$$+ \frac{\left(\frac{\partial^2}{\partial x^2} f(a, b)\right)(x - a)^2}{2} + \left(\frac{\partial^2}{\partial y\, \partial x} f(a, b)\right)(x - a)(y - b) + \frac{\left(\frac{\partial^2}{\partial b^2} f(a, b)\right)(y - b)^2}{2} + \ldots$$

$$= f(a, b) + (f')_x (a, b)(x - a) + (f')_y (a, b)(y - b) + \ldots$$

$$+ \frac{1}{j!} ((x - a)(f')_x (a, b) + (y - b)(f')_y (a, b))^{(j)} + \ldots$$

in which $(f')_x{}^{(j)}$ implies a derivative of order j with respect to x, with the corresponding terms for y and the mixed derivatives, in which terms have meaning obvious on comparison with a Taylor series in a single variable in section 3.306. Extension to multiple independent variables is effected in an obvious manner.

In section 2.414 we introduce Fourier series and in section group 4.5 discuss their applications as integrals of functions of a single variable. Recall that orthogonal functions of variable x in a set,

$$\left\{ \cos\left(\frac{2\,n\,\pi\,x}{L}\right),\ \sin\left(\frac{2\,n\,\pi\,x}{L}\right),\ 0 \leq n \right\}$$

including a constant function, serve to expand a function with period L. For a function of two independent variables x and y, we analogously construct a double Fourier series of orthogonal functions

$$\left\{ \cos\left(\frac{2\,n\,\pi\,x}{L_1}\right)\cos\left(\frac{2\,m\,\pi\,y}{L_2}\right),\ \sin\left(\frac{2\,n\,\pi\,x}{L_1}\right)\sin\left(\frac{2\,m\,\pi\,y}{L_2}\right), \right.$$

$$\left. \cos\left(\frac{2\,n\,\pi\,x}{L_1}\right)\sin\left(\frac{2\,m\,\pi\,y}{L_2}\right),\ \sin\left(\frac{2\,n\,\pi\,x}{L_1}\right)\cos\left(\frac{2\,m\,\pi\,y}{L_2}\right) \right\},$$

in which $m, n = 1,2,3, \ldots$, together with functions

$$\sin\left(\frac{2\,n\,\pi\,x}{L_1}\right),\ \sin\left(\frac{2\,n\,\pi\,y}{L_2}\right),\ \cos\left(\frac{2\,n\,\pi\,x}{L_1}\right),\ \cos\left(\frac{2\,n\,\pi\,y}{L_2}\right),$$

in which $n = 1, 2, 3, \ldots$, and a constant function. A linear region over which integration is taken becomes rectangular of area $L_1\,L_2$. An alternative formulation in terms of exponential functions rather than circular trigonometric functions is practicable, just as for a single variable.

optimization

   Optimization implies an evaluation of optimal values of a formula or function, such as maximal or minimal; if conditions on optimization be set, these become constraints. A method to reduce a problem of constrained optimization to an unconstrained problem, whereby one avoids substitution of constraint relations into the function, involves an addition of a sequence of products of real numbers λ_j and constraint formulae such that a point x_0 that minimizes a formula $f(x)$ subject to n constraints $g_1(x) = 0$, $g_2(x) = 0$, **...**, $g_n(x)$ becomes a stationary point of a lagrangian $L(x, \lambda)$,

$$L(x, \lambda) = f(x) + \left(\sum_{j=1}^{n} \lambda_j \, g_j(x) \right)$$

in which coefficients λ_j are called lagrangian multipliers. This method is valid if the gradients of the constraints be linearly independent at x_0.

   In this chapter we thus describe how to perform operations in differential and integral calculus on functions or their formulae of multiple independent variables, i.e. multivariate formulae or functions, with direct applications in thermodynamics among other chemical topics. As a basis for that explanation we recall how to define a formula or function of several variables, and then proceed to differentiation and integration. *Maple*'s package Student[MultivariateCalculus] contains 18 commands and operators that complement or supplement material in section groups 5.1, 5.3 and 5.4.

[−]   **summary of chapter 5**

   Our concern in this chapter is to develop an infinitesimal calculus of formula expressing functions of more than one independent variable, with some illustrative applications. We explain important concepts associated with stationary or critical points on a surface or hypersurface, located through use of derivatives and testing for cols, to prepare for chemical applications, particularly those involving chemical reactions, in which a pertinent hypersurface might represent internuclear potential energy. Both differentiation and integration of functions of multiple variables provide important tools to develop and to understand principles of chemical thermodynamics; the idea of a function of a thermodynamic state becomes established through consideration of exact differentials. We elaborate a link with Fourier series in section group 4.5 to demonstrate how one can construct expansion functions in sets as an initial step towards deriving Fourier series of functions of multiple variables: in this context, a concept of an outer product of sets is a valuable aid. Optimization of a formula involving multiple variables with constraints involves derivatives and solving simultaneous equations to locate maxima or minima satisfying both the formula and the constraint. All these powerful commands and operations to treat multiple independent variables require only a few *Maple* commands additional to those already introduced essentially for a single independent variable.

chapter 6   Linear algebra

[−]   **6.0 overview and principles**

   We present here a concise description of mathematical constructs pertaining to linear algebra,

their properties and principal operations, and a summary of important definitions. All these topics are discussed at length in succeeding sections grouped according to topic within linear algebra: a reader who is previously unacquainted with linear algebra might find this terse description indigestible; such a reader might proceed directly to examine ensuing material in this chapter one section group at a time, and return subsequently to this overview. Rigorously linear relations and systems are rare -- non-linear relations and their combinations abound. Treatment of linear relations and systems is, if not invariably easy, at least subject to a systematic and highly developed scheme of algorithms and their implementation; treatment of non-linear relations and systems is almost invariably difficult, despite enormous effort devoted to that field. Linear algebra, known formerly also as linear analysis, not only as a subject of study but also in application owes its importance not merely to those truly linear and rare relations and systems but especially to the ingenuity of mathematicians, scientists and engineers in discovering methods to find conditions under which non-linear systems are treatable with linear regions and approximations. A study of linear algebra hence rewards a practitioner beyond all proportion to the prevalence of linear systems.

   Two approaches to the study of linear algebra include linear algebraic equations that become expressed in a matrix form and an initial discussion of a vector space; as the latter is more abstract than the former, we adopt the former approach.

   From *Maple* release 9, a package **Student[LinearAlgebra]** contains many commands designed to assist one to understand concepts of linear algebra, through interactive operation with Maplets, graphic depictions and calculations; from *Maple* release 10, a package **Student[VectorCalculus]** contains analogously further commands that illuminate aspects of vector calculus. The LinearAlgebra Computation Example Worksheet illustrates selected commands in the former package.

   A sequence, list or set comprises numeric or symbolic items collected in a linear order, or in one dimension, although for a set the order is immaterial; in chapter 1 we introduce their properties as data structures that can contain both numeric and algebraic quantities, but these collections possess intrinsically no particular mathematical properties. To hold data, *Maple* provides other structures with dimension possibly greater than one, including a table, as introduced in section 1.120 and an array, in section 1.121. A matrix and a vector are special cases of arrays for which various arithmetic operations are defined. A matrix arises commonly in a mathematical description of a chemical or physical or engineering problem, and is typically applicable when data are presented in a tabular form.

   For a set of elements $a_{1,1}, a_{1,2}, ..., a_{m,n}$ numbering m n and which might be numerical or, as here, symbolic but typically representing numbers, a rectangular array $\mathbf{A} = (a_{i,j})$, with $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$, arranged in m rows and n columns as

$$
\begin{array}{cccc}
a_{1,1} & a_{1,2} & ... & a_{1,n} \\
a_{2,1} & a_{2,2} & ... & a_{2,n} \\
. & . & . & . \\
a_{m,1} & a_{m,2} & ... & a_{m,n}
\end{array}
$$

constitutes a *matrix* $m \times n$; if $m = n$, $\mathbf{A}$ is n-square. As such a rectangular array, a *matrix* is amenable to established mathematical operations; a square matrix has a determinant that evaluates

to a single expression or value, i.e. a scalar quantity. Two further mathematical objects vector and tensor possess well developed mathematical properties, although each originated in a physical context; with their derivatives and integrals, these quantities have important applications in chemical and physical calculations. A solution of an eigenvalue problem requires familiarity with at least three of these data structures. An ordered n-tuple $\underline{v} = (v_1, v_2, ..., v_n)$ of elements, also called components, constitutes an *n-vector*, formally implying an n-space or space of *n* dimensions; in general mathematical terms, a matrix *m* x 1 might be likewise called a column vector, whereas a matrix 1 x *n* is called a row vector; *Maple* objects Matrix and Vector are distinct entities. Calculus with vectors involves differential operators with properties resembling those of a vector. A spreadsheet provides a rectangular array, according to rows and columns, of elements, called cells, that not only facilitate viewing of abundant data but also simulate operations on lists, vectors and matrices; *Maple*'s spreadsheet enables operations on fully symbolic content of cells.

   The following paragraphs on topics of linear algebra contain condensed summaries of many important definitions and aspects of this major branch of mathematics. The material that follows in these paragraphs might appear somewhat formidable at a first reading; for this reason we encourage a reader to browse through this material to acquire an overview of the chapter and its constituent section groups, and then to refer to it in conjunction with calculations and practical applications of these principles in the various sections of this chapter, as appropriate. Upon completion of this chapter, a reader should be able to understand and to appreciate the terms and operations explained succinctly here.

matrix and determinant

   To acquire a notion of the nature of these linear algebraic structures and their expressions, we suppose that two simultaneous linear equations involve variables *x* and *y* corresponding to chemical or physical quantities that arise from the measurements of some chemical property, such as the concentrations of the solutes of a liquid solution that absorbs light at two wave lengths in the visible region; at each wave length there is a contribution to the total absorption from each solute, present at unknown concentration. On a basis of those measurements we might seek to evaluate a concentration of each separate solute even though their broad absorption lines overlap, so that absorption by both solutes contributes to the total absorbance at each wave length. Absorption coefficients, such as *a*, *b*, *d* and *e* that might pertain to each component at each wave length, might be known independently through separate calibration with each single solute; with such information we solve simultaneously these two equations to evaluate these concentrations if we measure the total absorbances *c* and *f* at those two wave lengths for the same solution containing both solutes. In these two algebraic equations,

$$a\,x + b\,y = c\,,$$
$$d\,x + e\,y = f\,,$$

symbols *a*, *b*, *c*, *d*, *e*, *f* take numerical values in practical conditions; mathematical solutions of these two equations apply to the same chemical system under the same experimental conditions, thus simultaneously, and both variables *x* and *y* and coefficients *a*, *b*, *c*, *d* appear only to an implicit first power, thus linearly. We express these two simultaneous linear equations in an alternative form as comprising three arrays, each demarcated by brackets [ ],

$$\begin{bmatrix} a & b \\ d & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} c \\ f \end{bmatrix}$$

in which coefficients a, b, c, d of variables x and y occupy the same relative positions as in the original equations, and c and f likewise; the orientation of x with respect to y differs between this -- vertical -- and the former -- horizontal alignment. According to this notation, we multiply leftmost element a in the top row of the first quantity demarcated by brackets by the uppermost element x in the second quantity within brackets and add to that product $a\,x$ a separate product of b with y to produce uppermost element c of the third quantity within brackets across the equality sign; an analogous sum of product d with x and of product e with y yields f. As an ordered arrangement of symbols, such a collection of letters or names of quantities between brackets in each separate structure above is an array; as each letter or name therein we expect to denote a number or variable in a particular application, such an array might contain numeric entries, or symbolic and numeric entries in an appropriate combination. Because we associate with such an array a possibility of involvement in well defined mathematical operations, such as the multiplication of the various quantities between two arrays as practised above, each array has significance beyond being an ordered arrangement of symbols in a space of two dimensions: for this reason each such array constitutes a matrix, a term introduced by Cayley, that implies certain mathematical properties. Such a matrix we treat as a single intrinsic entity, such as denoting a particular matrix by an informative name, rather than as a cluster of component parts. The matrix containing only the coefficients of the variables in the two algebraic equations is called a *coefficient matrix*, as

$$\begin{bmatrix} a & b \\ d & e \end{bmatrix};$$

when we augment that coefficient matrix with a further column of which each entry or element is a quantity in the right side of a linear equation, or correspondingly what is expressed as an array on the right side of the above equality, we generate an augmented matrix, as

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}.$$

We can solve the simultaneous linear equations merely through systematic operations on this augmented matrix.

   An alternative view of a matrix is that it involves -- enables -- a linear transformation: for instance, we apply $\begin{bmatrix} a & b \\ d & e \end{bmatrix}$ to transform $\begin{bmatrix} x \\ y \end{bmatrix}$ in the domain of the transformation into $\begin{bmatrix} c \\ f \end{bmatrix}$ in the codomain or range of that transformation, or we implement a linear mapping for the same purpose; the latter quantities $\begin{bmatrix} x \\ y \end{bmatrix}$ and $\begin{bmatrix} c \\ f \end{bmatrix}$ we might regard as column matrices or vectors or vector spaces.

   To facilitate our exploitation of a mechanism to manipulate such a quantity as a matrix, we rewrite two equations in an alternative form:

$$A_{1,1} X_1 + A_{1,2} X_2 = C_1,$$
$$A_{2,1} X_1 + A_{2,2} X_2 = C_2,$$

Instead of distinct names for variables x and y, we use an indexed or subscripted name, in this case

just X in form either X_1 or X_2, so that these equivalences $X_1 = x$, $X_2 = y$ show a correspondence with a preceding pair of linear equations. Likewise, instead of distinct names for coefficients such as a, b, d, e, we employ according to a more economical notation a doubly indexed or subscripted name, in this case $A_{j,\,k}$ in which the first subscript j indicates either the order of the equation in a vertical list, such as that displayed above, or the row in the matrix to which those equations become translated, as displayed below; a second subscript k corresponds to a variable of which a particular coefficient is a multiplicand, such as $A_{1,\,1} = a$ as coefficient of X_1 in equation 1, $A_{1,\,2} = b$ as coefficient of X_2 in equation 1, et cetera, or a column of a resulting matrix. We treat analogously quantities on the right side of equations above, so that c in the first equation becomes C_1, f in the second equation becomes C_2.

$$\begin{bmatrix} A_{1,\,1} & A_{1,\,2} \\ A_{2,\,1} & A_{2,\,2} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

The correspondence between the latter equation involving matrices and the preceding two simultaneous linear equations defines multiplication between matrices on the left side of the equality. We hence write the latter equation in a compact form

$$\mathbf{A} \cdot \mathbf{X} = \mathbf{C}$$

in which $\mathbf{A}$ denotes a matrix with four elements $A_{1,\,1}$, $A_{1,\,2}$, $A_{2,\,1}$, $A_{2,\,2}$; here is an explicit correspondence between this square matrix in two notations,

$$\mathbf{A} = \begin{bmatrix} A_{1,\,1} & A_{1,\,2} \\ A_{2,\,1} & A_{2,\,2} \end{bmatrix} = \begin{bmatrix} a & b \\ d & e \end{bmatrix}$$

and two column matrices in their analogous notations.

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} c \\ f \end{bmatrix}$$

Operator $\cdot$ denotes multiplication between matrices. These matrix forms in $\mathbf{A} \cdot \mathbf{X} = \mathbf{C}$ across an equality sign imply only linear equations, whether involving conventional algebraic quantities or derivatives of only first order or other quantities, but a particular matrix has a significance apart from such an environment. Not all sets of equations have solutions, if the equations be incompatible such as with $x + y = 3$ and $x + y = 4$, or solutions might number uncountably. A set of equations with terms on the right side, such as c and f in the original equations, or such as C_1 and C_2 in the matrix above, equal to zero is called homogeneous. Equations in a set that are expressible in a form $\mathbf{A} \cdot \mathbf{X} = \mathbf{C}$ comprise a *linear system,* of which matrix $\mathbf{A}$ is called the coefficient matrix; the matrix formed on adjoining $\mathbf{C}$ to the right of $\mathbf{A}$ is called an augmented matrix, as explicitly follows.

$$\begin{bmatrix} A_{1,\,1} & A_{1,\,2} & C_1 \\ A_{2,\,1} & A_{2,\,2} & C_2 \end{bmatrix}$$

A simple equation such as

$$2 + 3 = 5$$

is a particular arithmetical expression that indicates on the left of the equality a sum of two

numbers and on the right another number. When we replace the numbers by symbols,
$$x + y = z$$
we create an algebraic expression of general character in which symbols x, y, z might represent
either the particular numbers in the preceding displayed expression or any other numbers in an
appropriate combination to maintain the equality. The latter expression is an abstraction of the
former.

A matrix has its origin in a transformation of coordinates: in two dimensions a particular point
with coordinates (x, y) is transformed to another location with coordinates (x', y'). For instance, for
a transformation of coordinates according to an anticlockwise rotation of axes by an angle θ about
a common origin, the new coordinates become
$$x' = x \cos(\theta) - y \sin(\theta)$$
$$y' = x \sin(\theta) + y \cos(\theta)$$
which we express in matrix notation as
$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix};$$
each matrix might then be represented with a single symbol, as $\mathbf{r'} = \mathbf{M} \cdot \mathbf{r}$. Two transformations
of coordinates in sequence then become representable as a product of two matrices, one for each
transformation; if the first transformation be represented with a matrix $\mathbf{M}$ and the second
transformation be represented with a matrix $\mathbf{N}$, the total effect of both transformations is
represented as $\mathbf{r'} = \mathbf{N} \cdot \mathbf{M} \cdot \mathbf{r}$, in which successive operations are implemented in an order from
right to left so that matrix $\mathbf{N}$ of the second transformation premultiplies matrix $\mathbf{M}$ of the first
transformation. The result of calculating $\mathbf{M} \cdot \mathbf{r}$ is a matrix; the succeeding product of $\mathbf{N}$ with that
matrix yields $\mathbf{r'}$.

The derivative of matrix $\mathbf{A}$ is formed on differentiation of each element of that matrix; the
integral of $\mathbf{A}$, either definite or indefinite is formed on integrating each element of that matrix.

An elementary matrix of order n results from these elementary operations on an identity matrix
-- a square matrix with unity along the principal diagonal and zero elsewhere:
$$\begin{bmatrix} 1 & 0 & 0 & 0 & .. \\ 0 & 1 & 0 & 0 & .. \\ 0 & 0 & 1 & 0 & .. \\ 0 & 0 & 0 & 1 & .. \\ .. & .. & .. & .. & .. \end{bmatrix}$$

- interchanging any two rows of a unit matrix or identity matrix;

- multiplying a single row of an identity matrix by a scalar other than zero;

- replacing a particular row of an identity matrix by a sum of that row and another row multiplied
  by a scalar, which has the effect of inserting a scalar as an element off the principal diagonal.

A non-singular matrix, possessing an inverse, is expressible as a product of elementary matrices.
An elementary matrix has invariably an inverse. Operation with such a non-singular matrix on
another matrix effects a linear transformation that is applicable in solving linear simultaneous
equations, for inversion of a matrix and for other purposes. The determinant of a unit or identity
matrix is unity.

A rectangular matrix $\mathbf{A}$ having m rows and n columns might have two particular diagonals, one leading from $a_{1,1}$ to $a_{m,n}$ and the other leading from $a_{m,1}$ to $a_{1,n}$; of these two, the former is typically called the principal or main diagonal. The superdiagonal is the diagonal above the latter principal diagonal, and the subdiagonal is that diagonal below that principal diagonal. A banded matrix has zero elements along a few diagonals and zero elements elsewhere; a sparse matrix has mostly zero elements, but not necessarily in any particular order. A symmetric matrix is identical with its transpose, $\mathbf{A} = \mathbf{A}^t$ whereas for an antisymmetric or skew-symmetric matrix $\mathbf{A} = -\mathbf{A}^t$. The spectrum of a matrix constitutes its eigenvalues.

The permutation of the rows of a unit matrix in some order yields a permutation matrix, such as the following.

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Such a matrix is orthogonal, such that each such matrix multiplied by its inverse in either order equals an identity matrix.

A determinant is a formally important scalar property of a square matrix of order n; that matrix hence constitutes an array comprising n^2 elements, and its determinant represents a defined alternating sum of all possible products, numbering $n!$, of elements, one from each column and each row of that matrix; each term in the sum has a positive or negative sign depending whether the number of permutation inversions is even or odd. The order of a determinant is defined as the order of the square matrix from which it arises. The algebraic sum is called the expansion or value of the determinant; each product in that expansion with its associated sign is called a term in the expansion of that determinant. For instance for matrix $\mathbf{A}$ as formed above,

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

a determinant, hence of order two, is expanded for its evaluation as

$$|\mathbf{A}| = A_{1,1}A_{2,2} - A_{1,2}A_{2,1}$$

comprising two terms. The determinant function has thus as domain a set of square matrices; the range of this function depends upon the nature of elements of a matrix that serves as argument: for a matrix with purely numeric elements, its determinant yields a number, whereas, for a matrix with other than numeric elements, an algebraic or other expression is a result. Elements $a_{1,1}, a_{2,2}, ..., a_{n,n}$ form the principal diagonal, and elements $a_{1,n}, a_{2,n-1}, ..., a_{n,1}$ form the *secondary diagonal*; a product of the elements along the principal diagonal yields the *principal member*. A determinant has these properties:

- if each element of a particular row, or a particular column, be multiplied by a scalar quantity c, the value of the determinant becomes multiplied by c; conversely, multiplying the entire determinant by a scalar quantity c is equivalent to multiplying all elements in any one column or any one row;

- a factor that is found in all elements of a particular row or a particular column can be factored out;

- a determinant has zero value
  - if all elements of a particular row, or a particular column, be zero, or
  - if two rows, or two columns, contain identical corresponding elements, or
  - if two rows, or two columns, contain proportional corresponding elements;
- if two rows, or two columns, be interchanged, the sign of the value of the determinant becomes reversed;
- transposing the determinant such that rows become columns leaves the value of the determinant unaltered;
- when a factor is removed from each element of one row, or column, to yield a new determinant, the value of that determinant multiplied by the factor removed is the same as the value of the original determinant;
- when a determinant is multiplied by a constant or scalar quantity, the latter quantity can be absorbed into the determinant by multiplying therewith by all elements of one row, or one column;
- augmenting a determinant by adding at the top a row 1 c_0 c_1 c_2 ... c_n and adding 1 0 0 0 ... 0 as a new first column leaves the value of the determinant unaltered;
- a product of two determinants is equal to the determinant of the product of the two matrices that are the sources of the two determinants;
- the value of a determinant remains constant
  - if all rows and columns are interchanged,
  - if rows become written as columns, and columns as rows,
  - if to each element of one row be added a scalar quantity multiplied by the corresponding element of another row, and analogously for elements of columns.

For conforming square matrices of order n and non-zero scalar c,

$$| c\,\mathbf{A} | = c^n | \mathbf{A} |,$$
$$| \mathbf{A} . \mathbf{B} | = | \mathbf{A} | | \mathbf{B} |,$$

in which $| c\,\mathbf{A} |$ implies a multiplication of each element of the entire matrix by scalar c. Notations for a determinant of matrix $\mathbf{A}$ include det($\mathbf{A}$) and $|\mathbf{A}|$. A square matrix of which the determinant evaluates to zero is called singular and has no inverse; a matrix other than square is also singular but might have a pseudo-inverse. Although a determinant plays a central role in the theory of linear algebra and matrices, it serves generally no useful purpose in practical computation involving a matrix containing as elements real numbers because of prospectively severe loss of numerical precision when an alternating sum of products of elements be expressed as decimal numbers. For equations in an homogeneous set, the non-trivial solutions are uncountable if the determinant of the coefficients be zero or if the only solution be that the matrix of variables constitutes a zero matrix.

A permutation inversion describes a couple of elements that become out of order when described by their indices; for instance, for four elements a_1, a_2, a_3, a_4 permutation a_1 a_2 a_3 a_4 has all elements in order of increasing index, but permutation a_2 a_4 a_1 a_3 contains permutation inversions a_2 a_1, a_4 a_2 and a_4 a_3.

For a square matrix $\mathbf{A}$, the minor $M_{i,j}$ of element $A_{i,j}$ is the determinant of the matrix that remains after deleting row i and column j from $\mathbf{A}$; the cofactor $C_{i,j}$ of element $A_{i,j}$ is the determinant of the matrix that remains after deleting row i and column j from $\mathbf{A}$ multiplied by $(-1)^{(i+j)}$: the minor and cofactor hence differ only in sign: $C_{i,j} = \pm M_{i,j}$. The determinant of a square matrix becomes a sum of products of elements of any row or column with their cofactors. As an alternative definition, a determinant is a sum of product of elements with permutations of indices or subscripts

$$| \mathbf{A} | = \Sigma \, (\pm A_{1,j_1} A_{2,j_2} \, ... \, A_{n,j_n})$$

in which j_i is an index or subscript of numbers in the set $\{1, 2, ..., n\}$, and sign $+$ or $-$ is selected for each term depending whether the permutation is even or odd -- i.e. whether an even or odd number of interchanges is required to yield a particular permutation from the order 1 2 3 4 ... n.

According to a geometric interpretation of a determinant with real elements, the elements across each row of a matrix $n \times n$ become coordinates of a point in a space of n dimensions: for a matrix for which $n = 1$, which is just a number or scalar quantity, its determinant is interpreted as the signed length of a vector from the origin to this point along the single axis; for a matrix 2×2 as $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$, the determinant is the signed area of the parallelogram with four vertices that comprise the origin (0,0), two points (a,b) and (c,d) separately and their sum combinations in ($a + c$, $b + d$); for a matrix 3×3, the determinant is analogously the signed volume of the parallelepiped that includes the origin, each row as defining a vertex, and the sums of the rows in three-dimensional space defined by the matrix; the concept is extensible to space of n dimensions.

The wronskian of three functions u, v and w in formulae u(x), v(x) and w(x) of the same independent variable x is a determinant of this matrix containing the formulae and their derivatives, also called a fundamental matrix,

$$\begin{bmatrix} u(x) & v(x) & w(x) \\ \dfrac{d}{dx}u(x) & \dfrac{d}{dx}v(x) & \dfrac{d}{dx}w(x) \\ \dfrac{d^2}{dx^2}u(x) & \dfrac{d^2}{dx^2}v(x) & \dfrac{d^2}{dx^2}w(x) \end{bmatrix}$$

or analogously for other functions numbering k that imply evaluation of derivatives up to order $k - 1$. If the functions u, v, w be linearly dependent, the columns of the wronskian are likewise linearly dependent; as differentiation is a linear operation, the wronskian vanishes. The wronskian can hence serve to show that differentiable functions in a set are linearly independent on an interval by determinng that it does not vanish identically. If these formulae u(x), v(x) and w(x) be analytic and have two continuous derivatives on an open interval and this determinant evaluate to zero for functions and derivatives evaluated at any value of x on that interval, these functions

are linearly dependent. If the wronskian be not equal to zero for at least one point x on an interval, the functions are linearly independent on that interval.

A unit matrix, which might be rectangular or other than square, has elements unity along its principal diagonal, such as this one with three columns and two rows, or dimensions 3 x 2,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

whereas an identity matrix, commonly denoted **I** by mathematicians but **E** by chemists, is a square identity matrix, such as this unit matrix of dimensions 2 x 2,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

that might be denoted **I**[2] to indicate its order. Because I has a defined meaning as $\sqrt{-1}$ in *Maple*, it might prove convenient to use **E** to denote a unit matrix in various contexts. One might add, subtract or multiply two matrices of conformable dimensions, but division of one matrix by another matrix is undefined. A matrix comprising a single column, such as **X** being

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} x \\ y \end{bmatrix}$$

or **C** being

$$\mathbf{C} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} c \\ f \end{bmatrix}$$

might also constitute a vector, specifically a column vector, of a particular kind, which has mathematical properties and physical applications; a row matrix can analogously constitute a row vector. A vector might represent in turn a tensor of rank one (polar vector) or two (axial vector), whereas a matrix might represent a tensor of rank two; tensors of greater rank exist, as described below. The rank of a matrix is describable as the order of the largest determinant, of value not zero, that might be formed from elements of a given matrix. As our concern in this chapter, these quantities have properties and applications far beyond their apparently innocuous origin within compact notation to denote variables and coefficients in linear equations.

In its general form, a *matrix* comprises elements in a rectangular array with m columns and n rows, m and n being positive integers, that has well defined mathematical properties; in *Maple* numbering of rows and columns must begin at unity. If $m = n$, a matrix is square and has order n. A rectangular matrix m by n is characterized by its rank, which can be no larger than a minimum of m and n; rank signifies the maximum number of linearly independent equations that a particular matrix can represent. The dimension of the row space and column space of a matrix **A** with m columns and n rows is equal to the rank of that matrix, and the null space or kernel of **A**, denoted null(A), is thesubspace of dimension equal to that rank that comprises solutions of the homogeneous linear system **A** $x = 0$; the nullity, denoted nullity(A), of that matrix is the dimension of its null space, so that rank(A) + nullity(A) = n or the order of a square matrix. An element of a matrix can be a real or imaginary or complex number, or an algebraic quantity that denotes such a number, or even another matrix; in our exploration in this chapter we employ only a number or variable or algebraic expression as such an element. A matrix is considered to be an operator of a particular type in various circumstances, such as when it operates on an eigenvector

to yield a [product](#) of that eigenvector with a [scalar](#) [eigenvalue](#), or when it functions as a rotation matrix to rotate a point, line or vector.

Each element of a zero matrix is [zero](#), but that zero matrix is distinct from scalar zero. For an identity or unit matrix, only elements along the principal diagonal of a unit matrix are not zero, but unity. A scalar matrix is a unit matrix multiplied by a scalar quantity. Addition or multiplication of three conformable matrices is [associative](#), but multiplication of two distinct conformable matrices $\mathbf{A}$ and $\mathbf{X}$ is [commutative](#), such that $\mathbf{A} \cdot \mathbf{X} = \mathbf{X} \cdot \mathbf{A}$, only when one matrix is a zero matrix, a unit matrix, a scalar matrix or the other matrix raised to a power. [Transposition](#) of matrix $\mathbf{A}$ to generate [transpose](#) matrix $\mathbf{A}^T$ involves making each row of elements of $\mathbf{A}$ into a column of elements of $\mathbf{A}^T$. For matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

its transpose $\mathbf{A}^T$ has a form

$$\mathbf{A}^T = \begin{bmatrix} A_{1,1} & A_{2,1} \\ A_{1,2} & A_{2,2} \end{bmatrix}$$

The determinant of the transpose of a square matrix, obtained by interchanging rows and columns, is equal to the determinant of the same matrix,

$$|\mathbf{A}^T| = A_{1,1} A_{2,2} - A_{1,2} A_{2,1}$$

but interchanging two columns, or two rows, of a matrix reverses the sign of the determinant. If two rows, or two columns, of a matrix be identical, or alter only by a common factor, the value of the determinant is zero. The *rank* of a given matrix is accordingly the largest integer r such that at least one determinant of order r, for a submatrix of that order formed from that matrix by deleting rows and columns, differs from zero.

If matrices $\mathbf{A}$ and $\mathbf{B}$ have each order n over the field of real or complex numbers, these properties of their determinants hold:

- $\det(\mathbf{A}) = \det(\mathbf{A}^T)$, in which $\mathbf{A}^T$ is a [transpose](#) of $\mathbf{A}$;

- $\det(\mathbf{A} \cdot \mathbf{B}) = \det(\mathbf{A}) \det(\mathbf{B})$;

- $\det(\mathbf{A}) = \det(\mathbf{A}^*)$ in which $\mathbf{A}^*$ is [complex conjugate](#) of $\mathbf{A}$;

- if $\mathbf{B}$ result from $\mathbf{A}$ through interchange a pair of rows or columns, $\det(\mathbf{B}) = -\det(\mathbf{A})$;

- if $\mathbf{B}$ result from $\mathbf{A}$ through multiplication of elements of a row or column by a scalar quantity k, $\det(\mathbf{B}) = k \det(\mathbf{A})$;

- if two rows, or two columns, of $\mathbf{A}$ be identical or a row, or column, comprise zeros, $\det(\mathbf{A}) = 0$.

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are equal only if each element of one matrix is identically equal to the corresponding element in the other matrix: $a_{i,j} = b_{i,j}$ for all i, j. Multiplication of a matrix $\mathbf{A} = (a_{ij})$ by a constant or scalar quantity c yields another matrix $\mathbf{B}$ of the same dimensions of which the elements are $b_{i,j} = c\, a_{i,j}$. A [sum](#), not a [direct sum](#), of two matrices $\mathbf{A}$ and $\mathbf{B}$ of conformable dimensions yields a matrix $\mathbf{C}$ of which each element is a sum of elements of the combining matrices,

$$c_{i,j} = a_{i,j} + b_{i,j}$$

for which purpose the number of rows of **A** and **B** must be equal and the number of columns of **A** and **B** must be equal: such matrices conform for operation addition. A product, not a direct product, of two matrices **A** and **B** of conformable dimensions yields a matrix **C** of which each element $c_{i,j}$ is a sum of products of elements of the combining matrices,

$$c_{i,j} = \sum_k a_{i,k} b_{k,j}.$$

for which purpose the number of columns of matrix **A** must equal the number of rows of matrix **B**: such matrices conform for operation multiplication. If products **A B** and **B A** of two matrices **A** and **B** be equal, **A** and **B** commute, but this condition holds in only special cases. A product of a square matrix **A** and its reciprocal $\mathbf{A}^{(-1)}$ generates an identity matrix: $\mathbf{A}\,\mathbf{A}^{(-1)} = \mathbf{A}^{(-1)}\mathbf{A} = \mathbf{I}$ or **E**.

The trace or spur of a square matrix of order n is a sum of elements along the principal diagonal:

$$trace(\mathbf{A}) = \sum_{j=1}^{n} a_{j,j}$$

this trace is invariant under cyclic permutation of matrices in a product. For two matrices **A** and **B** conforming for the particular operations, their traces possess these properties:

- $\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$;

- $\mathrm{tr}(\mathbf{A}\,\mathbf{B}) = \mathrm{tr}(\mathbf{B}\,\mathbf{A})$; $\mathrm{tr}(\mathbf{A}\,\mathbf{B}\,\mathbf{C}\,\mathbf{D}) = \mathrm{tr}(\mathbf{B}\,\mathbf{C}\,\mathbf{D}\,\mathbf{A}) = \mathrm{tr}(\mathbf{C}\,\mathbf{D}\,\mathbf{A}\,\mathbf{B}) = \mathrm{tr}(\mathbf{D}\,\mathbf{A}\,\mathbf{B}\,\mathbf{C})$, a *cyclic* property;

- $\mathrm{tr}(c\,\mathbf{A}) = c\,\mathrm{tr}(\mathbf{A})$;

- $\mathrm{tr}(\mathbf{A}^{T}) = \mathrm{tr}(\mathbf{A})$;

the second property reflects the statement about cyclic permutation above. The trace of a matrix is also the sum of its eigenvalues.

A square matrix has a reciprocal provided that its determinant is not zero: a singular matrix has a zero determinant. For a matrix **A** to be invertible -- so as to generate an inverse, which is unique, a criterion is that its determinant be not zero; both the matrix **A** and its inverse $\mathbf{A}^{(-1)}$ have the same order that is equal to the rank. For **A** an invertible matrix and n a positive integer, $\mathbf{A}^{(-n)} = (\mathbf{A}^{(-1)})^{n} = (\mathbf{A}^{n})^{(-1)}$. The determinant of a non-singular matrix **A** is equal to the reciprocal of the determinant of its inverse matrix $\mathbf{A}^{(-1)}$: $|\mathbf{A}| = 1/|\mathbf{A}^{(-1)}|$. A square matrix has at most one inverse.

A direct sum of two matrices, which are not necessarily conformable, represented with a symbol with + and O superimposed, yields a matrix with each addend as a block along the principal diagonal.

$$\mathbf{A} = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$$

$$\mathbf{B} \;=\; \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}$$

$$\mathbf{A} + \mathrm{O}\ \mathbf{B} \;=\; \begin{bmatrix} a_{1,1} & a_{1,2} & 0 & 0 & 0 \\ a_{2,1} & a_{2,2} & 0 & 0 & 0 \\ 0 & 0 & b_{1,1} & b_{1,2} & b_{1,3} \\ 0 & 0 & b_{2,1} & b_{2,2} & b_{2,3} \\ 0 & 0 & b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}$$

A <u>direct product</u> or <u>Kronecker product</u> of two matrices, which likewise are not necessarily conformable and represented with a symbol with x and O superimposed, yields a matrix of which each element is a product of one element from each multiplicand matrix; for a matrix of dimensions m x n in a direct product with another matrix of dimensions p x q, the resulting matrix hence has dimensions mp x nq; such a direct product is associative but not commutative. For two

matrices $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\begin{bmatrix} e & f & g \\ h & i & j \\ k & l & m \end{bmatrix}$, their direct product is

$$\begin{bmatrix} ae & af & ag & be & bf & bg \\ ah & ai & aj & bh & bi & bj \\ ak & al & am & bk & bl & bm \\ ce & cf & cg & de & df & dg \\ ch & ci & cj & dh & di & dj \\ ck & cl & cm & dk & dl & dm \end{bmatrix}.$$

　　With conforming matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ and scalar a and b,the laws that govern matrices of conformable dimensions and that are, for certain properties, not singular, are

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} \;=\; \mathbf{A} + (\mathbf{B} + \mathbf{C}) \qquad \text{\underline{associative} addition}$$
$$(\mathbf{A} . \mathbf{B}) . \mathbf{C} \;=\; \mathbf{A} . (\mathbf{B} . \mathbf{C}) \qquad \text{associative multiplication}$$
$$\mathbf{A} + \mathbf{0} \;=\; \mathbf{0} + \mathbf{A} \;=\; \mathbf{A} \qquad \mathbf{0} \text{ is the zero matrix}$$
$$a\,(\mathbf{A} + \mathbf{B}) \;=\; a\,\mathbf{A} + a\,\mathbf{B} \quad \text{left \underline{distributive} property of addition}$$
$$(a + b)\,\mathbf{A} \;=\; a\,\mathbf{A} + b\,\mathbf{A} \quad \text{right distributive property of addition}$$
$$(a\,b)\,\mathbf{A} \;=\; a\,(b\,\mathbf{A}) \qquad \text{associative multiplication by scalars}$$
$$\mathbf{A} . (\mathbf{B} + \mathbf{C}) \;=\; \mathbf{A} . \mathbf{B} + \mathbf{A} . \mathbf{C} \qquad \text{distributive}$$
$$(\mathbf{A} + \mathbf{B}) . \mathbf{C} \;=\; \mathbf{A} . \mathbf{C} + \mathbf{B} . \mathbf{C} \qquad \text{distributive}$$
$$\mathbf{A} + \mathbf{B} \;=\; \mathbf{B} + \mathbf{A} \qquad \text{\underline{commutative} addition}$$
$$\mathbf{A} - \mathbf{B} \;=\; \mathbf{A} + (-1)\,\mathbf{B} \qquad \text{subtraction}$$
$$\mathbf{A} . \mathbf{B} \;\diamond\; \mathbf{B} . \mathbf{A} \qquad \text{non-commutative multiplication}$$
$$\mathbf{A} . \mathbf{1} \;=\; \mathbf{1} . \mathbf{A} \;=\; \mathbf{A} \quad \text{commutative multiplication with identity matrix } \mathbf{1}$$
$$a\,(\mathbf{A} . \mathbf{B}) \;=\; (a\,\mathbf{A}) . \mathbf{B} \;=\; \mathbf{A} . (a\,\mathbf{B}) \quad \text{distributive multiplication by scalars}$$
$$\mathbf{A}^{r} . \mathbf{A}^{s} \;=\; \mathbf{A}^{(r+s)} \qquad \text{for } r, s \text{ non-negative integers}$$
$$(\mathbf{A}^{r})^{s} \;=\; \mathbf{A}^{(rs)} \qquad \text{for } r, s \text{ non-negative integers}$$
$$\mathbf{A}^{0} \;=\; \mathbf{1} \qquad \text{with identity matrix } \mathbf{1}$$

$$\mathbf{A}^n = \mathbf{A} \cdot \mathbf{A} \cdot \mathbf{A} \dots \mathbf{A}$$ exponentiation of a matrix for an integer power implies repeated multiplication

$$(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$$ sum of transposes

$$(\mathbf{A}^T)^T = \mathbf{A}$$ transpose of transpose

$$(\mathbf{A} \cdot \mathbf{B})^T = \mathbf{B}^T \cdot \mathbf{A}^T$$ distribution of transpose

$$(\mathbf{A}^{(-1)})^{(-1)} = \mathbf{A}$$ inverse of inverse

$$(\mathbf{A} \cdot \mathbf{B})^{(-1)} = \mathbf{B}^{(-1)} \cdot \mathbf{A}^{(-1)}$$ distribution of inverse

$$(\mathbf{A}^{(-1)})^T = (\mathbf{A}^T)^{(-1)}$$ inverse and transpose

Unlike scalar quantities, the product of two conforming non-zero matrices might yield a zero matrix. For determinants of transpose and inverse matrices,

$$|\mathbf{A}^T| = |\mathbf{A}|,$$

$$|\mathbf{A}^{(-1)}| = \frac{1}{|A|}.$$

A real matrix is orthogonal if its inverse equals its transpose,

$$\mathbf{A}^{(-1)} = \mathbf{A}^T$$

and its determinant is ± 1. For such a real symmetric square matrix $\mathbf{A}$ of order n there exists a real orthogonal matrix $\mathbf{B}$ such that $\mathbf{B}^{(-1)} \mathbf{A} \mathbf{B}$ or $\mathbf{B}^T \mathbf{A} \mathbf{B} = $ a diagonal matrix.

For a particular square matrix $\mathbf{A}$ or $\mathbf{B}$, conforming and in general complex but not singular, we define *associated matrices*, namely a complex conjugate *matrix* denoted $\mathbf{A}*$, a transpose *matrix* $\mathbf{A}^T$ and an hermitian conjugate or adjoint *matrix* $\mathbf{A}^{*T}$ and analogously for $\mathbf{B}$, that imply the following properties:

$$(\mathbf{A} + \mathbf{B})* = \overline{\mathbf{A}* + \mathbf{B}*}$$ complex conjugate of a sum;

$$(z\,\mathbf{A})* = \overline{z}\,\mathbf{A}*$$ complex conjugate of a scalar multiple;

$$(\mathbf{A} \cdot \mathbf{B})* = \mathbf{B}* \cdot \mathbf{A}*$$ complex conjugate of a product;

$$(\mathbf{A}*)* = \mathbf{A}$$ composition of complex conjugate operation;

if $\mathbf{A} = \mathbf{A}*$, $\mathbf{A}$ is *real*;

if $\mathbf{A} = -\,\mathbf{A}*$, $\mathbf{A}$ is *imaginary*;

if $\mathbf{A} \cdot \mathbf{A}* = \mathbf{A}* \cdot \mathbf{A}$, $\mathbf{A}$ is normal;

if $\mathbf{A} = \mathbf{A}^T$, $\mathbf{A}$ is *symmetric*;

if $\mathbf{A}$ and $\mathbf{B}$ be each *symmetric*, the product is symmetric if $\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$;

if $\mathbf{A} = -\,\mathbf{A}^T$, $\mathbf{A}$ is *antisymmetric* or *skew* (also called *skew symmetric*);

if square matrix $\mathbf{A} = \mathbf{A}^{*T}$, $\mathbf{A}$ is *hermitian conjugate* or *adjoint*;

if square matrix $\mathbf{A} = -\,\mathbf{A}^{*T}$, $\mathbf{A}$ is *antihermitian*;

if $\mathbf{A}^2 = \mathbf{A}$, $\mathbf{A}$ is idempotent;

if $\mathbf{A} = \mathbf{A}^{(-1)}$, $\mathbf{A}$ is *self-reciprocal*;

if $\mathbf{A}^{(-1)} = \mathbf{A}^T$, matrix $\mathbf{A}$ and its transpose $\mathbf{A}^T$ are orthogonal, and their product $\mathbf{A}\,\mathbf{A}^T$ is a *unit matrix*, $\mathbf{I}$ or $\mathbf{E}$;

$$\text{if } \mathbf{A}^{(-1)} = \mathbf{A}^{*T}, \quad \mathbf{A} \text{ is } \underline{\text{unitary}};$$

$$\text{for a } \underline{\text{permutation matrix}} \mathbf{A}, \quad \mathbf{A}^{(-1)} = \mathbf{A}^{T}.$$

A complex conjugate matrix $\mathbf{B}$ of $\mathbf{A}^*$ is formed from matrix $\mathbf{A}$ by taking the complex conjugate of each element in the latter matrix. The transpose of $\mathbf{B}$ or $\mathbf{A}^*$ becomes the complex conjugate transposed matrix, or adjoint matrix, of $\mathbf{A}$. The product $\mathbf{A}^{T}. \mathbf{A}$ or $\mathbf{A}. \mathbf{A}^{T}$ of a general matrix $\mathbf{A}$ and its transpose $\mathbf{A}^{T}$ is a symmetric matrix. The determinant of a unitary matrix has value ± 1. Hermitian and unitary matrices play the same roles for matrices with complex elements as symmetric and orthogonal matrices play for matrices with real elements. An inverse matrix arises notably in the solution of simultaneous linear equations in sets and in deriving a concept of a group that is the basis of symmetry theory; because an inverse matrix of a matrix with real numbers as elements is greatly susceptible to rounding error, and because its determinant is involved in the production of an inverse matrix, efficient calculations avoid direct use of an inverse matrix when practicable. The determinant of an orthogonal matrix $\mathbf{A}$ is equal to the determinant of its transpose $\mathbf{A}^{T}$; the determinant of the product $\mathbf{A}\mathbf{A}^{T}$ is equal to the square of the determinant of matrix $\mathbf{A}$; because that product is equal to a unit matrix, each determinant must evaluate to ± 1. Such an orthogonal matrix plays an important role in transformations of coordinates that serve to characterize the symmetry properties of molecules.

For the $\underline{\text{derivative}}$ or $\underline{\text{differential}}$ $\underline{\text{quotient}}$ of a matrix, each element is differentiated individually to form a matrix of the derivatives of the elements provided that these elements are differentiable: $\dfrac{\partial}{\partial t} \mathbf{A}(t) = (\dfrac{\partial}{\partial t} a_{j,k})$. The elements are likewise integrated individually:

$$\int_a^b \mathbf{A}(t)\, dt = (\int_z^b a_{j,k}\, dt).$$

A quantity $\mathbf{e}^A$ containing square matrix $\mathbf{A}$ is called an $\underline{\text{exponential matrix}}$ or matrix exponential, defined as $\mathbf{e}^A = \sum_{j=0}^{\infty} \dfrac{A^j}{j!}$; this quantity with scalar k is expanded as

$$\mathbf{e}^{(kA)} = \sum_{j=0}^{\infty} \frac{k^j A^j}{j!} = \mathbf{I} + \frac{kA}{1!} + \frac{k^2 A^2}{2!} + \ldots;$$

in which A^j implies multiplication of a matrix by itself j times, and for $j = 0$ yields a unit matrix $\mathbf{I}$; this exponential matrix is hence a matrix of the same order as $\mathbf{A}$. As both infinite series converge for every $\mathbf{A}$ and k, the matrix exponential is defined for all square matrices. For a square matrix of order n, an alternative expansion that is computationally simpler is

$$\mathbf{e}^{(kA)} = f_0 \mathbf{I} + f_1 k \mathbf{A} + \ldots + f_{n-2} k^{(n-2)} \mathbf{A}^{(n-2)} + f_{n-1} k^{(n-1)} \mathbf{A}^{(n-1)}$$

in which f_j are functions of k that are determined for each $\mathbf{A}$. The matrix exponential has these properties:

- if matrices $\mathbf{A}$ and $\mathbf{B}$ commute such that $\mathbf{A}. \mathbf{B} = \mathbf{B}. \mathbf{A}$, $\mathbf{e}^{(A+B)} = \mathbf{e}^A \mathbf{e}^B$;

- for any **A**, the matrix e^A is invertible, and has an inverse $\left(e^A\right)^{(-1)} = e^{(-A)}$;

- for **M** any invertible matrix conforming with **A**, $e^{\left(M^{(-1)} A M\right)} = M^{(-1)} e^A M$, and

- for eigenvectors $\underline{v}$ for which $\mathbf{A}\,\underline{v} = \lambda\,\underline{v}$, $e^A\,\underline{v} = e^\lambda\,\underline{v}$, thus connecting the eigenvalues λ and eigenvectors $\underline{v}$ of **A** with those of e^A.

An exponential matrix has application in the solution of ordinary differential equations in systems.

   Two square matrices **A** and **B** are classified as similar if some invertible matrix **C** over the field of real or complex numbers transforms one into another, such as in

$$\mathbf{B} = \mathbf{C}^{(-1)}.\,\mathbf{A}.\,\mathbf{C},$$

or equivalently

$$\mathbf{A} = \mathbf{C}.\,\mathbf{B}.\,\mathbf{C}^{(-1)};$$

operation with **C** thus generates a similarity transformation of **A** to **B**, or the reverse; matrices **A** and **B** are then congruent. Similar matrices have the same eigenvalues. Two matrices **A** and **B** are similar if there exist an invertible transformation of the above type between them; **A** and **B** then represent the same linear transformation with respect to bases related by **C**. A square matrix **A** is diagonalizable if there exist a matrix **C** such that

$$\mathbf{D} = \mathbf{C}^{(-1)}.\,\mathbf{A}.\,\mathbf{C},$$

with **D** a diagonal matrix. A square matrix **A** is orthogonally diagonalizable if there exist a matrix **C** such that

$$\mathbf{D} = \mathbf{C}^{T}.\,\mathbf{A}.\,\mathbf{C},$$

with **D** a diagonal matrix, because $\mathbf{C}^{T} = \mathbf{C}^{(-1)}$ for an orthogonal matrix **C**. The inverse of a diagonal matrix is also diagonal; each element is the reciprocal of the corresponding element in the original matrix. If matrix **A** be symmetric, the result of $\mathbf{C}^{T}.\,\mathbf{A}.\,\mathbf{C}$ is also symmetric for any conforming **C**.

If **C** be a column matrix $\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$ of variables and square matrix **A** be symmetric, $\mathbf{C}^{T}.\,\mathbf{A}.\,\mathbf{C}$ constitutes a quadratic form that plays an important role in problems of optimization and in representing a tensor.

   As illustrated above, a permutation matrix **P** has precisely one element unity in each row and in each column, all other entries being zero. Matrix **A** of order n is decomposable, also called reducible, if there exist permutation matrix **P** such that

$$\mathbf{P}.\,\mathbf{A}.\,\mathbf{P}^{(-1)} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ O_{n-k,k} & A_{2,2} \end{bmatrix}$$

in which $\mathbf{A}_{1,1}$ is a square *matrix* of order k, $\mathbf{A}_{2,2}$ is a square *matrix* of order $n - k$, and $O_{n-k,k}$ is a zero or null matrix $(n - k) \times k$, with $1 \le k < n$. If no such **P** exist, **A** is *indecomposable*. The methods of Gauss, Doolittle, Crout and Cholesky are particular algorithms to compute the

decomposition of a matrix $\mathbf{A}$ to various products $\mathbf{L} \cdot \mathbf{D} \cdot \mathbf{U}$ of unique matrices such that $\mathbf{L}$ has elements zero above the principal diagonal, $\mathbf{D}$ is a diagonal matrix with non-zero elements only on its principal diagonal, $\mathbf{U}$ has elements zero below the principal diagonal, and matrices $\mathbf{L}$ and $\mathbf{U}$ might also have unity along their diagonals; then $\mathbf{A} = \mathbf{L} \cdot \mathbf{D} \cdot \mathbf{U}$. If $\mathbf{A}$ be non-singular, a permutation of rows and columns brings $\mathbf{A}$ to a form that satisfies these conditions.

A rectangular matrix $\mathbf{A}$ might have a pseudo-inverse matrix $\mathbf{A}^{+}$; if $\mathbf{A}$ be a square matrix and invertible, $\mathbf{A}^{+} = \mathbf{A}^{(-1)}$. If a product $\mathbf{A}^{T} \cdot \mathbf{A}$ be invertible,

$$\mathbf{A}^{+} = (\mathbf{A}^{T} \cdot \mathbf{A})^{(-1)} \mathbf{A}^{T},$$

and similarly if $\mathbf{A} \cdot \mathbf{A}^{T}$ be invertible,

$$\mathbf{A}^{+} = \mathbf{A}^{T} \cdot (\mathbf{A} \cdot \mathbf{A}^{T})^{(-1)}.$$

An application of pseudo-inverse matrices arises in solution of linear equations because a vector $\underline{v} = \mathbf{A}^{+} \underline{b}$ is that solution of equations in a system $\mathbf{A}^{T} \cdot \mathbf{A} \, \underline{v} = \mathbf{A}^{T} \underline{b}$ that has a minimal norm. These properties are associated with a pseudo-inverse matrix:

- for $\mathbf{A}$ having dimensions $p \times q$, $\mathbf{A}^{+}$ has dimensions $q \times p$;

- the rank of $\mathbf{A}^{+}$ equals the rank of $\mathbf{A}$;

- $(\mathbf{A}^{+})^{+} = \mathbf{A}$;

- $\mathbf{A} \cdot \mathbf{A}^{+} \cdot \mathbf{A} = \mathbf{A}$;

- $\mathbf{A}^{+} \cdot \mathbf{A} \cdot \mathbf{A}^{+} = \mathbf{A}^{+}$;

- $(\mathbf{A} \cdot \mathbf{A}^{+})^{T} = \mathbf{A} \cdot \mathbf{A}^{+}$ and $(\mathbf{A}^{+} \cdot \mathbf{A})^{T} = \mathbf{A}^{+} \cdot \mathbf{A}$, as $\mathbf{A} \cdot \mathbf{A}^{+}$ and $\mathbf{A}^{+} \cdot \mathbf{A}$ are symmetric.

The most efficient method to form a pseudo-inverse matrix involves a singular-value decomposition.

For a matrix $\mathbf{A}$ containing symbolic elements in terms of variable t as parameter, which we express as $\mathbf{A}(t)$, we define a limit matrix as the matrix that results from the transition $t \to t_0$ for each element such that $\lim_{t \to t_0} \mathbf{A}(t) = \lim_{t \to t_0} a_{j,k}(t)$.

For the derivative or differential quotient of a matrix the elements are differentiated individually, $\frac{d}{dt} \mathbf{A}(t) = (\frac{\partial}{\partial t} a_{j,k})$, and for an integral the elements are integrated individually, $\int_{a}^{b} \mathbf{A}(t) \, dt = (\int_{a}^{b} a_{j,k} \, dt)$.

A matrix of matrices is called a *hypermatrix* of which an element of an inner matrix is labeled $a_{j,k_{m,n}}$.

Maple provides several operations on quantities of type Matrix in package LinearAlgebra, or of type matrix in package linalg; the latter package is obsolescent, becoming superseded by package **LinearAlgebra** and is retained only for applications that have been based on it: it should not

be generally used for new calculations. For convenience we assign a name to represent a matrix that in text appears in bold font, such as **A**, except when limitations of *Maple*'s notation preclude such usage in compound symbols within a worksheet. All properties in the preceding summary are readily tested and verified on application of *Maple* commands explained in this chapter.

vector

As the world that we inhabit has three spatial dimensions, fundamental physical laws and their applications in chemistry are expressed as mathematical relations that involve these three dimensions, or at most four dimensions when one ncludes time to produce a *four-vector* in a relativistic treatment; vector analysis is a branch of mathematics devised to express and to implement these relations, but is not limited to three or four dimensions. We generally denote a vector in text with symbol $\underline{v}$, i.e. an underscored minuscule italic letter; alternative traditional designations have an arrow placed above that symbol, or, in perhaps increasingly obsolescent usage, an arrow beneath that symbol, or more commonly a tilde "~" beneath that symbol; the latter is not realisable in a *Maple* worksheet. In a cartesian space of three dimensions, vector $\underline{v}$ is definable abstractly as three numbers (v_x, v_y, v_z), called components, in an ordered set, such that multiplication with a number or scalar quantity α yields a product

$$\alpha\,\underline{v} = (\alpha\,v_x,\ \alpha\,v_y,\ \alpha\,v_z),$$

and such that addition of two vectors $\underline{u}$ and $\underline{v}$ yields

$$\underline{u} + \underline{v} = (u_x + v_x,\ u_y + v_y,\ u_z + v_z)$$

According to an algebraic mathematical point of view, a vector is merely a column matrix or a row matrix, whereas, according to a geometric point of view, a vector with real components is a difference between two points in an euclidean space as a displacement from one point to another; such a vector preserves no information about either the source or destination point but might have only a length, direction and a directional sense. A vector might thus be represented as a directed segment of a line in space; if the origin be chosen as one point, the other point defining a directed segment of a line has coordinates (v_x, v_y, v_z) in a cartesian system. The magnitude $|\underline{v}|$ of vector $\underline{v}$ is defined as its length, calculated as

$$|\underline{v}| = (\underline{v}\cdot\underline{v})^{\left(\frac{1}{2}\right)} = \sqrt{v_x^{\,2} + v_y^{\,2} + v_z^{\,2}}$$

The scalar or dot or inner product $\underline{u}\cdot\underline{v}$ of two polar vectors $\underline{u}$ and $\underline{v}$ is a scalar quantity or number defined with

$$\underline{u}\cdot\underline{v} = \underline{v}\cdot\underline{u} = |\underline{u}|\,|\underline{v}|\cos(\theta)$$

in which θ denotes the angle between the two vectors; if $|\underline{u}| \neq 0$ and $|\underline{v}| \neq 0$ and $\cos(\theta) = 0$ such that $\theta = \dfrac{\pi}{2}$ rad, $\underline{u}\cdot\underline{v} = 0$: the two vectors are mutually orthogonal or perpendicular. This scalar product is expressed in terms of the components of the vectors as

$$\underline{u}\cdot\underline{v} = u_x\,v_x + u_y\,v_y + u_z\,v_z$$

If any component of these two vectors be complex, the scalar product must be calculated through the latter formula rather than that above containing $\cos(\theta)$ because that angular quantity has no ordinary geometric meaning when one or more components of $\underline{u}$ and $\underline{v}$ are not real. The component of $\underline{v}$ in the direction $\underline{u}$ is the projection of $\underline{v}$ on $\underline{u}$, such that

$$\text{component of } \underline{v} \text{ on } \underline{u} \ = \ v_u = \ \underline{v} \cdot \underline{u} \, / \, |\underline{u}|$$
$$\text{component of } \underline{u} \text{ on } \underline{v} \ = \ u_v = \ \underline{v} \cdot \underline{u} \, / \, |\underline{v}|$$
$$\text{or } \underline{v} \cdot \underline{u} \ = \ v_u \, |\underline{u}| \ = \ u_v \, |\underline{v}| \ .$$

For a vector $\underline{v}(t)$ that is a function of scalar variable t, its derivative is expressed in terms of its components as $\mathrm{d}\,\underline{v}(t)/\mathrm{d}t = (\frac{\partial}{\partial t} v_x, \frac{\partial}{\partial t} v_y, \frac{\partial}{\partial t} v_z)$ and the corresponding differential is $\mathrm{d}\,\underline{v}(t) = (\mathrm{d}v_x, \mathrm{d}v_y, \mathrm{d}v_z)$. For a scalar function $\alpha(t)$ of t, $\mathrm{d}(\alpha\,\underline{v}) = \alpha\,\mathrm{d}\underline{v} + \underline{v}\,\mathrm{d}\alpha$. For $\underline{u}(t)$ and $\underline{v}(t)$ both functions of scalar variable t, $\mathrm{d}(\underline{u} \cdot \underline{v}) = \underline{u}\,\mathrm{d}\underline{v} + \underline{v}\,\mathrm{d}\underline{u}$.

   A physical or chemical point of view of such a vectorial quantity is more complicated. Many a quantity, such as temperature, volume, mass, concentration and frequency, of concern to a chemist we describe as having type scalar, which signifies that it has magnitude and generally also units appropriate for that quantity; for a particular measurement, its specification is independent of any system of coordinates and requires just one number, with its units. For each point in a region of space, a scalar function or field has a value expressed as a scalar quantity. A quantity of another type, such as an electric dipolar moment of a molecule or of a macroscopic sample of substance, or the velocity of a molecule with respect to a system of coordinates, has associated with it both a magnitude and a direction with respect to some either internal or external axes of a system of coordinates; such a physical object has type vector, which according to a geometric definition involving a directed segment of a line generally implies *magnitude, direction* and sense, in addition to appropriate units, but not necessarily position. A vector is classifiable according to three types: a position vector or bound vector has a fixed point, typically the origin, as its base; a *line vector* can slide along its line of action without rotation, such as a mechanical force acting on a moving body; a *free vector* or *general vector* is unrestricted as to its location but defined completely according to its magnitude, direction and sense, representable graphically as any parallel line of equal length in a given set: this property of a free vector with an unrestricted location makes it translationally invariant. With an affine space that has no preferred origin are particularly associated free or general vectors. Two directed line segments are equal if they have the same length, direction and sense.   A vector in two spatial dimensions has properties similar to those of a complex number; such a complex number implies two independent pieces of information -- its real and complex parts, whereas a vector in n dimensions implies as much information as n components. A vector function or vector field is a vectorial quantity associated with each point in a region of space.

   A linear vector space comprises vectors, or functions, in a set and standard operations addition and scalar multiplication. For a conventional euclidean space of two dimensions, the vector space is just plane $x\,y$ that has two standard vectors traditionally denoted $\underline{i}$ and $\underline{j}$; vector $\underline{i}$ lies along axis x and $\underline{j}$ along axis y. Any point in plane xy is describable as a linear combination or superposition of those two standard vectors with appropriate coefficients: those vectors span that space. Only two vectors are required to span that space in two dimensions -- further vectors are redundant for that purpose, because these two vectors are linearly independent and thus inexpressible one as a multiple of the other. Two vectors $\underline{u}$ and $\underline{v}$ in a two-dimensional space are linearly independent if the only solution of linear equation $c_1\,\underline{u} + c_2\,\underline{v} = 0$ be $c_1 = c_2 = 0$, which implies geometrically that

these vectors do not lie along the same direction but define a plane; these vectors u and v are otherwise linearly dependent. A description of a plane has having two dimensions implies that only two linearly independent vectors are required to express any other vector in that plane. Any two non-collinear vectors can likewise span the vector space in plane xy; such sets thus number uncountably, but any is reducible to a linear combination of unit vectors i and j. An euclidean space of n dimensions requires, in general, n vectors as the minimum number, which must be linearly independent, to span that space. Linearly independent vectors in a set that span a space become a basis for that particular space, and their number defines the dimension of that space. Any vector in that space is expressible as a unique superposition of those basis vectors.

According to a conventional link between geometry and algebra, a point in a plane that requires for its location the specifications of abscissal and ordinate values of its cartesian coordinates, relative to coordinate axes in a fixed set, as an ordered list becomes a vector with the same ordered values as its components, and analogously for a point in a space of three or more dimensions. In this way, a vector escapes from the limits imposed by three spatial dimensions of common experience. A point becomes thereby represented as a set of numbers, and another geometric object as an equation.

A vector is further classified as being *polar* or *axial*, depending on its behaviour under inversion of coordinates through the origin of coordinates in a cartesian system; for this operation of symmetry according to such a system, all coordinates become negated: for a point with cartesian coordinates (x, y, z), inversion effects

$$x \to -x, y \to -y, z \to -z$$

On reflexion of a point in a plane containing the origin, coordinates along either axis within that plane are unaffected, but coordinates along an axis perpendicular to that plane have that sign altered: for instance, for reflexion in plane xy,

$$x \to x, y \to y, z \to -z$$

An axial or pseudovector arises as a vectorial product of two polar or true vectors. On inversion through the origin of a system of cartesian coordinates, a polar vector p with its base at that origin is transformed into its inverse vector such that signs of its components become reversed -- i.e. its sense is reversed, whereas an axial vector a that results from a vectorial product of two such polar vectors is unaffected by such an operation:

$$\text{on inversion, } p \to -p, a \to a$$

On reflexion in a plane through the origin of a system of cartesian coordinates, a polar vector p within that plane is not so inverted, but a polar vector perpendicular to that plane has its sense reversed. An axial vector formed as a vectorial product of two polar vectors within a plane of reflexion has its sense reversed, but an axial vector formed as a vectorial product of a vector within that plane and another vector perpendicular to that plane is thus unaffected by such an operation: treating that axial vector as a result of an appropriate vectorial product, we hence conclude that

$$\text{on reflexion within a parallel plane, } p \to p, a \to -a \text{ ;}$$
$$\text{on reflexion perpendicular to a plane, } p \to -p, a \to a \text{ .}$$

For a polar or axial vector parallel to but outside a plane of reflexion, this operation causes a translation across this plane, but does not alter the effect on the sense as specified above. These

properties showing the effect of an operation of symmetry clearly distinguish axial and polar vectors. An axial vector hence transforms like a polar vector under a proper rotation, but has its sense reversed under an improper rotation that arises as a result of first inversion and then a proper rotation; an improper rotation thus occurs upon inversion of coordinates.

The direction cosines $[\,l,\,m,\,n\,]$ of a vector are the cosines of the angles between that vector and cartesian axes Ox, Oy and Oz respectively. For a vector $\underline{v} = v_x\,\underline{i} + v_y\,\underline{j} + v_z\,\underline{k}$, in which $\underline{i}$, $\underline{j}$ and $\underline{k}$ are unit vectors along positive x, y and z axes or Ox, Oy and Oz respectively, $l = \dfrac{v_x}{|v|}$, $m = \dfrac{v_y}{|v|}$ and $n = \dfrac{v_z}{|v|}$, in which $|v| = \sqrt{v_x^{\,2} + v_y^{\,2} + v_z^{\,2}}$ is the length of vector $\underline{v}$. Angle θ in range $[0,\,\pi]$ between two non-zero vectors $\underline{u} = u_x\,\underline{i} + u_y\,\underline{j} + u_z\,\underline{k}$ and $\underline{v} = v_x\,\underline{i} + v_y\,\underline{j} + v_z\,\underline{k}$ that are co-terminal -- having tails at a common point -- is defined through its cosine from the scalar product $\underline{u} \cdot \underline{v} = |u|\,|v|\,\cos(\theta)$ as

$$\cos(\theta) = l_1\,l_2 + m_1\,m_2 + n_1\,n_2 \; = \; \frac{u_x\,v_x + u_y\,v_y + u_z\,v_z}{|u|\,|v|}$$

in which l_1, m_1, n_1 are the direction cosines of $\underline{u}$ with its length $|u|$ and l_2, m_2, n_2 are the direction cosines of $\underline{v}$ with its length $|v|$.

An equation for a plane through a point $(x_1,\,y_1,\,z_1)$ perpendicular to a vector $\underline{v} = v_x\,\underline{i} + v_y\,\underline{j} + v_z\,\underline{k}$ is

$$v_x\,(x - x_1) + v_y\,(y - y_1) + v_z\,(z - z_1) = 0.$$

If all vectors of a space except the null vector be normalized, such that their magnitude or length is unity but they retain their direction and sense, any one of these vectors can be transformed into any other by means of a rotation, possibly complex for complex vectors. A matrix describing such a rotation is unitary.

Both linear momentum of a molecule and electric field are instances of polar vectors, whereas angular momentum and magnetic field are two instances of axial vectors. Angular momentum measures the extent to which a linear momentum is directed about a particular point, called the origin, and is thus a moment of momentum; as angular momentum depends upon a chosen origin, when working with angular momentum one must take care to specify the origin and not to combine angular momenta about separate origins. Like electric field, magnetic field is a vector field: with each point in space of three dimensions is associated a position vector that might vary temporally; a magnetic field is a physical entity produced by moving electric charges, hence constituting an electric current, that exert force on other moving charges. The direction of this field is the equilibrium direction of the needle of a compass placed in that field. The intrinsic angular momentum of an elementary particle such as an electron or proton produces a magnetic field and acts on that field as if it were an electric current; this interpretation is a basis of explanation of a ferromagnet or other solid material acting as a permanent magnet.

A *vector* in one spatial dimension differs from a *scalar* because its *sense* is retained. A vector representing a chemical or physical quantity involves commonly three spatial dimensions: its components in these directions correspond to elements of an array with one array dimension, or a matrix with only a single row or column. In a context of a space having three dimensions, a vector

comprises three entities in an ordered set that, under rotation of coordinates, transform as coordinates of a fixed point, i.e. as (x, y, z). One must bear in mind a distinction between spatial dimension, here numbering three, and array dimension, here one; the latter resembles a significance of dimension in Basic or Fortran programming language. In a general mathematical sense a vector comprises an ordered column or ordered row of n symbols of arbitrary dimensions that also number n. A vector in *Maple* that implies n components is generated according to a basis with n basis vectors; a collection of all such vectors forms a vectorial space of dimension n: such a space is defined in terms of mathematical requisites, details of which are of no concern here. A vector in three spatial dimensions might be represented graphically as an arrow, of which the length is proportional to the magnitude of the vector, its orientation indicates the direction and the arrowhead the sense, or in a particular coordinate system by a triplet of real numbers; the vector is not that triplet of numbers, but has a representation as three numbers that varies in a systematic way as the coordinate system alters. A polar vector in space of three dimensions has its origin in a quaternion -- cf section group 12a.8, but lacking a real component, and might be represented as a matrix comprising a single row or column; an axial vector might be represented as a tensor of second rank that has a form of an antisymmetric or skew-symmetric square matrix of order three. The transpose of a column vector is a row vector, and vice versa. A quaternion resembles a vector in a real vector space with four dimensions; one might also consider a quaternion to be a hypercomplex number, in terms of three separate imaginary components and one real component, or as a couple comprising a vector in three dimensions and a scalar. Although there be similarities of a quaternion, which has common applications in computer animation, and a polar vector in three spatial dimensions, there are also important distinctions.

A polar or proper vector has all three attributes -- magnitude, direction and sense; an axial vector, also known as *pseudovector*, lacks *sense*, in a way that a polar vector possesses that attribute, and requires instead a direction of rotation to express its sense attribute. A polar vector can represent a translation or a mechanical force; a position of an atomic nucleus in space relative to a system of axes, a translation, a mechanical force, an electric field and electric dipolar moment are all instances of a polar vector that arises in a chemical context. The density of magnetic flux associated with a magnetic field and angular momentum are examples of an axial vector in that their context implies a sense -- clockwise or counterclockwise -- of rotation, or a screw sense, such as a current flowing in a loop, or a mass rotating in a closed circular path; a sense of this character differs from that indicated with an arrowhead that might serve to indicate the sense of a polar vector depicted as an arrow. A sense of an axial vector is thus associated with physical attributes of a system, and provides an additional defining feature to a vector that represents a dynamic property. Geometrically, a polar vector is represented with a displacement or a directed segment of a line that defines its magnitude; with an axial vector one associates an area, analogous to a length to depict a polar vector. An axial vector in three dimensions is an antisymmetric tensor of rank two, for which reason its components might be denoted with two indices rather than one for a polar vector; an axial vector or pseudovector is meaningful only in a space of three dimensions, whereas a polar vector can be meaningful in a space with dimensions of arbitrary number. Whereas in three spatial dimensions a scalar function might depend on three spatial coordinates but have a single value at a given point (x, y, z), a vectorial function has both a value and a

direction at such a point and thus requires three components to describe that function. An axial vector arises from a vector product of two polar vectors in a space of three dimensions; an example is angular momentum of a mass moving on a circular path, which results from a product of polar vectors linear momentum and radius of curvature, for which reversing the direction of rotation reverses the sign of the angular momentum. Whereas the strength of an electric field is a polar vectorial quantity that can result from static electric charges with a particular spatial distribution, as well as from moving electric charges or from an electromagnetic field, the density of magnetic flux is an axial vector that results from a vector product of one vector -- current density -- and another vector that specifies a spatial relation between a point of measurement and an element of current density.

A vector treated in *Maple* is implicitly a polar vector, and is a mathematical quantity rather than a physical quantity when any such distinction might arise; a vectorial product of two vectors in three spatial dimensions in *Maple* properly yields an axial vector, with a sense determined according to a right-hand rule. *Maple* fails to distinguish, one should bear in mind, between an axial and a polar vector; the best way to generate an axial or pseudovector in *Maple* is as a vectorial or cross product between two polar or true vectors.

With scalar a and b and zero vector $\underline{0}$, vectors $\underline{u}$, $\underline{v}$ and $\underline{w}$ in three spatial dimensions have these properties:

$$\underline{u} + \underline{v} = \underline{v} + \underline{u} \qquad \text{commutative addition}$$
$$(\underline{u} + \underline{v}) + \underline{w} = \underline{u} + (\underline{v} + \underline{w}) \qquad \text{distributive addition}$$
$$\underline{u} + \underline{0} = \underline{0} + \underline{u} = \underline{u} \qquad \text{action of zero or null vector}$$
$$\underline{u} + (-\underline{u}) = \underline{0} \qquad \text{vector and negative vector}$$
$$a(b\underline{u}) = (a b)\underline{u} \qquad \text{associative scalar multiplication}$$
$$a(\underline{u} + \underline{v}) = a\underline{u} + a\underline{v} \quad \text{distributive scalar multiplication}$$
$$(a + b)\underline{u} = a\underline{u} + b\underline{u} \quad \text{distributive scalar multiplication}$$

Vectors conforming to these rules form a linear vector space. A commutative addition of two vectors yields another vector. A difference of one vector with itself, which corresponds to a sum of one vector with another vector of the same length and direction but reverse sense, yields a zero vector. Vectors in a set in a two-dimensional space added to yield a zero sum form a closed polygon.

A vector originated as a physical quantity; the concept of a vector became a mathematical abstraction, but it has numerous realizations in physical science. In the most familiar applications of vectors, various physical quantities are represented with polar vectors in an euclidean space of three dimensions: an euclidean space has an underlying euclidean geometry. In other applications dimensions might number more than three. For instance, according to a conventional treatment of vibrational motions of atomic nuclei within a polyatomic molecule, methane has nine characteristic molecular vibrational modes; each of these one can visualize in three dimensions of euclidean space: thus a trajectory of a given nucleus is defined formally in terms of a space having nine dimensions, because an arbitrary motion of a nucleus is expressible as a linear combination of nine vibrational modes with symbolic coefficients. For comparison with a scalar field that is a region of space in which with each point a scalar function, such as temperature, is associated, for a vectorial quantity, such as an electric field of which the strength is specified at each point, those points and associated vectors constitute a vector field: a vector field is thus a

function according to which a vector is assigned to each point in its domain; that domain might be a plane, a curve in space or a region in two or more dimensions. A scalar field has no vectorial properties but is not necessarily a tensor of zero rank; a vector field might be invariant to rotations, but remains distinct from a scalar field. A scalar product of a vector and a vector field is undefined.

Vector analysis conventionally signifies the geometry and calculus of such vectors in a space of any dimension. For one, two or three dimensions, we represent geometrically a polar vector as a directed segment of a line, with sense determined according to movement from an initial point to an end point of that segment; the magnitude or conventional norm or absolute value of a vector is the length of that segment. For a null vector, the origin and end point are the same: this vector has thus zero length or magnitude and no definite direction. A *vector* in one spatial dimension remains distinct from a *scalar* because its *sense* is retained: if a vector lie along the terrestrial equator, its direction is well defined; the sense of that vector specifies whether it points east or west from its base or origin. The initial point of a position vector -- the tail of an arrow that might represent a vector -- coincides generally with the origin of a system of orthogonal coordinates; an alternative choice might be appropriate for a local or position vector, such as a force, that acts at a particular point.

```
> #  Execute this command.
  plots[arrow]([2,3], title="vector in two dimensions",
  axes=frame,
      scaling=constrained, titlefont=[TIMES,BOLD,14], colour=red
  );
>
```

For a vectorial space, multiplication of a vector by a scalar quantity, or addition of two vectors, generates another vector with components no more numerous than the original vector, or vectors, thus another vector in the same vectorial space. Addition of vectors is commutative and associative. A vectorial space of order n, represented as $\underline{\mathbf{R}}^{n}$, implies a vector in that space to have no more than n components; such a vectorial space comprises elements in a set, each element of which is a vector, and for a common physical space $n = 2$ for a plane in space or $n = 3$ corresponding to conventional dimensions of physical space. Addition of two elements of $\underline{\mathbf{R}}^{n}$ generates another element of that space; likewise multiplication of an element by a scalar quantity generates another element of that space. That space $\underline{\mathbf{R}}^{n}$ is closed under addition of vectors and scalar multiplication. If a vectorial space $\underline{\mathbf{R}}^{n}$ contain vectors $\underline{v1}$, $\underline{v2}$, $\underline{v3}$, ..., $\underline{vn}$, a further vector $\underline{v}$ is expressible as a linear combination of those vectors if there exist scalar quantities -- numbers -- c_1, c_2, c_3, ..., c_n such that

$$\underline{v} = c_1 \underline{v1} + c_2 \underline{v2} + c_3 \underline{v3} + ... + c_n \underline{vn} .$$

Vectors $\underline{v1}$, $\underline{v2}$, $\underline{v3}$, ... $\underline{vn}$ span a vectorial space $\underline{\mathbf{R}}^{n}$ if every vector in that space be expressible as a linear combination of these vectors; if vectors $\underline{v1}$, $\underline{v2}$, ..., $\underline{vm}$ be vectors in a vectorial space $\underline{\mathbf{R}}^{n}$ that span a subspace U, those vectors generate that subspace. Vectors in a set { $\underline{v1}$, $\underline{v2}$, $\underline{v3}$, ..., $\underline{vn}$} are linearly dependent if there exist scalar quantities c_1, c_2, c_3, ..., c_n, not all zero, such that a sum

$$c_1 \underline{v1} + c_2 \underline{v2} + c_3 \underline{v3} + ... + c_n \underline{vn} = 0,$$

yields zero, whereas those vectors in a set $\{\ \underline{v1},\ \underline{v2},\ \underline{v3},\ ...,\ \underline{vn}\ \}$ are linearly independent if that above sum is satisfied only when $c_1 = 0$, $c_2 = 0$, $c_3 = 0$, ... $c_n = 0$. This condition of linear independence is important because it is a necessary and sufficient condition for a solution of an equation

$$\underline{v}\ =\ c_1\ \underline{v1}\ +\ c_2\ \underline{v2}\ +c_3\ \underline{v3}\ +\ ...\ +c_n\ \underline{vn}$$

to be unique for all $\underline{v}$ that depend on $\underline{v1}$, $\underline{v2}$, $\underline{v3}$, ..., $\underline{vn}$; hence $\underline{v1}$, $\underline{v2}$, $\underline{v3}$, ..., $\underline{vn}$ are linearly independent if and only if every vector $\underline{v}$ be expressible in one and only one way as a linear combination of $\underline{v1}$, $\underline{v2}$, $\underline{v3}$, ..., $\underline{vn}$, or not at all. With $\{\underline{v1},\ \underline{v2},\ \underline{v3},\ ...,\ \underline{vn}\}$ as base vectors $\{\underline{e1},\ \underline{e2},\ \underline{e3}$,.., $\underline{en}\}$ or a basis, this formula provides an algebraic definition of a vector.

A matrix operates on a vector to transform it into another vector. A rotational matrix alters not the magnitude of a vector, only its orientation. According to a composition law, for all conforming matrices **A** and **B** and column vector $\underline{v}$, **A** . (**B** . $\underline{v}$) = (**A** . **B**) . $\underline{v}$. Quantity L is a linear operator if it conform to this property involving scalars a_1 and a_2 and vectors $\underline{v}_1$ and $\underline{v}_2$:

$$L\,(\,a_1\,\underline{v}_1\ +\ a_2\,\underline{v}_2\,)\ =\ \ a_1\,L\,\underline{v}_1 +\ a_2\,L\,\underline{v}_2$$

A matrix can represent an operator and obey this linear property; a matrix can thus represent a linear operator.

A linear transformation with a non-singular matrix **A** is a mapping $\tau\ \underline{\mathbf{R}}^{\,n}$ --> $\underline{\mathbf{R}}^{\,n}$, with operator τ, that preserves the linearity of a vectorial space in that

- a line is mapped into a line,

- a segment of a line is mapped into a segment of a line,

- parallel lines are mapped into parallel lines, and

- lines through the origin are mapped into lines through the origin.

A rotation, dilation, contraction or reflexion is a non-singular transformation because a corresponding matrix has an inverse. An inverse of a dilation is obviously a contraction. These operations are important in regard to molecular symmetry and vibrational modes.

As an orthogonal matrix **A**, for which $\mathbf{A}^{(-1)} = \mathbf{A}^{\,T}$, is non-singular, an orthogonal transformation with such a matrix preserves linearity, but also preserves the 2-norm or length, angle and distance of a vector.

For a translation, a transformation slides a vector or point in a direction and through a distance defined with a vector $\underline{t}$; such a translation preserves a line, angle and distance. An affine transformation involves a linear transformation of a type such as one of the four named above, followed by a translation, or $\tau\,\underline{\mathbf{R}}^{\,n}$ --> $\underline{\mathbf{R}}^{\,n}$ followed with $T\,(\underline{u1})\ =\ \mathbf{A}\,\underline{u1} +\underline{t}$. Neither a translation nor an affine transformation is a linear transformation.

For a linear transformation τ that maps a vector $\underline{u1}$ into another vector $\underline{u2}$, or $\tau : \underline{u1} \to \underline{u2}$, the null space of τ is a subspace of $\underline{u1}$ and the range of τ is a subspace of $\underline{u2}$; the range of a transformation τ is the set of vectors in $\underline{u2}$ that are the images of vectors in $\underline{u1}$. Two square matrices **A** and **B** produce the same linear transformation if and only if they be similar.

These transformations are important in relation to processes affecting a solid body, such as a crystal. When such a body is subjected to application of a load or stress, deformation occurs, to an

extent depending on elasticity and plasticity of that body. A linear shear is a transformation in which all points in one line or in one plane remain fixed but all other points or lines translate parallel to the fixed line or plane through a distance proportional to their distance from that fixed line or plane; for instance, a linear shear applied to a rectangle parallel to one edge yields a parallelogram. Scaling occurs when a diagonal matrix has unequal elements along the principal diagonal, which distorts a body upon its application thereto.

Multiplication of scalar quantities is an unambiguous operation well defined, whereas for vectors the situation is complicated -- there exist scalar, vectorial and direct products. For two vectors $u1$ and $u2$, neither of which is zero, $u2$ is expressible in terms of its vector components along $u1$ or perpendicular to $u1$ just as it be expressible in terms of components along the unit or base vectors. A matrix

$$\mathbf{P} \,=\, (\,\underline{u1}\,.\,\underline{u1}^{\,T}\,)\,/\,(\,\underline{u1}^{\,T}\,.\,\underline{u1}\,)$$

that contains a dyadic in the numerator and a scalar in the denominator serves to project any vector onto vector $u1$ in the same vectorial space, of which further description appears below. The projection of a vector onto a subspace involves a pseudo-inverse matrix that has as columns linearly independent vectors that generate that subspace. A linear transformation and concepts of null space and range are important in an analysis of a system of linear equations.

A binary multiplicative combination of two polar vectors $\underline{R}$ and $\underline{S}$ to form a scalar quantity is termed a scalar product, dot product or inner product that is a scalar quantity, according to a formula

$$\underline{R}\,.\,\underline{S} \,=\, |\,\underline{R}\,|\,|\,\underline{S}\,|\,\cos(\theta) \,=\, \underline{S}\,.\,\underline{R}$$

in which $|\underline{R}|$ and $|\underline{S}|$ denote magnitudes of combining vectors and θ is the angle between them; this scalar product is commutative. The value of this scalar product is hence a measure of the coalignment of two vectors, and is independent of the system of coordinates. The square root of a scalar product of a vector with itself $\underline{R}\,.\,\underline{R}$, for which $\theta = 0$ and $\cos(\theta) = 1$, yields the length of that vector, or the vector norm. That an outcome of such a scalar product of these three scalar quantities -- $|\underline{R}|$, $|\underline{S}|$ and $\cos(\theta)$ -- vindicates one name for combination of vectors in this manner; this scalar product thus yields a *true scalar*, the simplest invariant that one might form from two vectors. If $\theta = \dfrac{\pi}{2}$ rad, $\cos(\theta) = 0$ and $\underline{R}$ and $\underline{S}$ are described as orthogonal -- their scalar product is zero, $\underline{R}\,.\,\underline{S} \,=\, 0$; in three dimensions, an equivalent description is that these two vectors are oriented at right angles to one another, or perpendicular, whereas for dimensions of greater number a general term *orthogonal* is preferable. A scalar or inner product of two vectors $f(x)$ and $g(x)$ in a vector space that contains functions continuous over a finite closed interval $[a, b]$, with respect to weight function $w(x)$, is defined as integral $\displaystyle\int_a^b f(x)\,g(x)\,w(x)\,dx$. If that integral evaluate to zero, the vectors are orthogonal to each other. The norm or length of vector $f(x)$ becomes $\mathrm{norm}(f(x)) = \sqrt{\displaystyle\int_a^b f(x)^2\,w(x)\,dx}$. Dividing a vector by its norm yields a normalized vector; vectors in a set that are both normalized and orthogonal comprise an orthonormal set, and

a scalar or inner product of any two such vectors is either <u>zero</u>, if the vectors differ, or <u>unity</u>, if the vectors be alike, so being describable with <u>Kronecker's delta</u> function $\delta_{f,\,g}$. This scalar product of two vectors obeys also the distributive law,

$$\underline{R} \cdot (\underline{S} + \underline{T}) = \underline{R} \cdot \underline{S} + \underline{R} \cdot \underline{T}$$

The relation $\underline{R} \cdot \underline{S} = \underline{R} \cdot \underline{T}$ implies not that $\underline{S} = \underline{T}$ but that $\underline{R} \cdot (\underline{S} - \underline{T}) = 0$, so that $\underline{R}$ is perpendicular to the difference vector $\underline{S} - \underline{T}$. In manual calculations of a scalar product, the left side of the formula reduces to a weighted sum of scalar products involving all <u>basis vectors</u> $\underline{i}$, $\underline{j}$ and $\underline{k}$ two at a time: as angle θ between two unit (polar) vectors is either zero, if they be the same, or $\dfrac{\pi}{2}$ rad, if they be distinct, we have $\underline{i} \cdot \underline{i} = \underline{j} \cdot \underline{j} = \underline{k} \cdot \underline{k} = 1$ for $\theta = 0$, and $\underline{i} \cdot \underline{j} = \underline{j} \cdot \underline{k} = \underline{k} \cdot \underline{i} = 0$

for $\theta = \dfrac{\pi}{2}$; this scalar product of two unit vectors is thus equal to the <u>direction cosine</u> relating the two directions. The projection of a polar vector onto a coordinate axis, so defining its cartesian components, is a special case of a scalar product; for example $\underline{R} \cdot \underline{i}$ yields a projection of vector $\underline{R}$ along cartesian coordinate axis x. A scalar product conforms to associative and distributive properties. If a product of two basis vectors be either zero or unity, thus representable with <u>Kronecker's</u> δ function, these basis vectors form an <u>orthonormal set</u>. Formation of n orthonormal vectors from n linearly independent vectors is known as <u>Gram-Schmidt</u> *orthogonalization*. Division by a vector is impracticable. This application of a dot product becomes a basis of euclidean geometry; an alternative definition of an inner product, of which a dot or scalar product as defined above is a special case, enables one to work with more general, or non-euclidean, geometries, which have applications in special and <u>general relativity</u>. This scalar product expresses many physical properties, such as work done in moving a body that is a scalar product of force and displacement.

   A scalar product of a general polar or proper vector, with initial point at the origin, and an axial vector or pseudovector is a *pseudoscalar* -- a quantity that is invariant under translation or rotation of coordinate axes, but that reverses its sign when the direction of each axis in a cartesian system is reversed -- <u>inversion</u>. A scalar product of two vectors of the same type, either polar or axial, thus yields a scalar quantity, whereas a corresponding combination of a polar vector and axial vector yields a pseudoscalar quantity.

   In a space of three dimensions containing two polar vectors $\underline{R}$ and $\underline{S}$ in a plane at angle $\theta \neq 0$ to each other -- hence not parallel, a vectorial product, also called a <u>cross product</u> or *matrix product* or *skew product*, and designated $\underline{R} \times \underline{S}$, generates a further vector normal or perpendicular to that plane; that vector, of type axial vector or pseudovector, has accordingly magnitude $|\underline{R}|\,|\underline{S}|\sin\theta$, and direction and sense depending on order of multiplication or on left or right nature, or chirality, of the coordinate system; this product is anticommutative in that $\underline{R} \times \underline{S} = - \underline{S} \times \underline{R}$, but these vectors conform to the distributive law, $\underline{R} \times (\underline{S} + \underline{T}) = \underline{R} \times \underline{S} + \underline{R} \times \underline{T}$. Thus $\underline{R} \times \underline{S} = |\underline{R}|\,|\underline{S}|\sin\theta\,\underline{n}$, in which $\underline{n}$ is a unit vector normal to a plane defined by $\underline{R}$ and $\underline{S}$. An angle between $\underline{R}$ and $\underline{S}$ is definable in either clockwise or counterclockwise senses. This product is expressible also as

$$\underline{R} \times \underline{S} = \underline{i}\,(R_y S_z - R_z S_y) + \underline{j}\,(R_z S_x - R_x S_z) + \underline{k}\,(R_x S_y - R_y S_x)$$

and in determinantal form as

$$(\underline{R} \times \underline{S}) = \begin{Vmatrix} i_{-} & j_{-} & k_{-} \\ r_x & r_y & r_z \\ s_x & s_y & s_z \end{Vmatrix}$$

A vectorial product of a vector with itself vanishes, $\underline{R} \times \underline{R} = 0$ because $\sin(0) = 0$; hence for unit vectors, $\underline{i} \times \underline{i} = \underline{j} \times \underline{j} = \underline{k} \times \underline{k} = 0$ for $\theta = 0$, and $\underline{i} \times \underline{j} = \underline{k}$, $\underline{j} \times \underline{k} = \underline{i}$, $\underline{k} \times \underline{i} = \underline{j}$, $\underline{j} \times \underline{i} = -\underline{k}$, $\underline{k} \times \underline{j} = -\underline{i}$ and $\underline{i} \times \underline{k} = -\underline{j}$. Whereas a sum of two polar vectors not collinear corresponds to a diagonal of a parallelogram with these two vectors as adjacent sides, a vectorial product of two polar vectors corresponds to the area of that parallelogram; such an area is represented by its normal. If we form a matrix 2 x 2 by incorporating two vectors each with two components as columns therein, the determinant of that matrix is equal to the area of that parallelogram, which in turn equals the length of a pseudovector resulting from the vector product. Analogously for three dimensions, if we form a matrix 3 x 3 by incorporating three non-parallel vectors each with three components as columns therein, the determinant of that matrix is equal to the volume of a parallelepiped: each face of that solid body is a parallelogram and each vector defines four parallel edges; the volume of this body is also equal to the magnitude of a scalar product of one vector with the cross product of the other two vectors. Whereas a vectorial product of two vectors of the same type, either polar or axial, yields an axial vector, a vectorial product of two vectors of distinct types yields a polar vector. Two vectors commonly represented as vector products are angular momentum of a particle, which arises as a cross product of its linear momentum and the radius vector from the origin to the particle, and torque, which forms as a cross product of force and a vector representing a lever arm. A multiplicative combination $\underline{R} \cdot (\underline{S} \times \underline{T})$ is a triple scalar product, evaluated as this determinant of the components of the three vectors,

$$\underline{R} \cdot (\underline{S} \times \underline{T}) = \begin{Vmatrix} r_x & r_y & r_z \\ s_x & s_y & s_z \\ t_x & t_y & t_z \end{Vmatrix}$$

that accordingly yields a scalar result with a geometric interpretation as the volume of a parallelepiped defined with those three vectors emanating from one vertex, whereas $\underline{R} \times (\underline{S} \times \underline{T})$ is a vector triple product. that accordingly yields a vectorial result; parentheses are required to specify the order of binary combination. These two identities apply to such vector triple products:

$$\underline{R} \times (\underline{S} \times \underline{T}) = (\underline{R} \cdot \underline{T})\,\underline{S} - (\underline{R} \cdot \underline{S})\,\underline{T}$$
$$(\underline{R} \times \underline{S}) \times \underline{T} = (\underline{T} \cdot \underline{R})\,\underline{S} - (\underline{T} \cdot \underline{S})\,\underline{R}$$

The definition of a vectorial or cross product applies to only three spatial dimensions.

Although pseudovector $\underline{R} \times \underline{S}$ is orthogonal to both $\underline{R}$ and $\underline{S}$, and therefore perpendicular to a plane containing $\underline{R}$ and $\underline{S}$, two possible directions for this pseudovector are compatible with these requirements: by convention, its sense is chosen so that $\underline{R}$, $\underline{S}$ and $\underline{R} \times \underline{S}$ -- in that order -- follow a right-hand rule. To see how this rule works, hold the right fist clenched in front of the body: initially direct the first finger, representing $\underline{R}$, to the left, and the second finger, representing $\underline{S}$, toward the body; the thumb extended upward then indicates the direction of vectorial product $\underline{R} \times \underline{S}$. If to orient $\underline{R}$ and $\underline{S}$ in directions of the extended thumb and first finger, respectively, be easier, resultant $\underline{R} \times \underline{S}$ is directed downwards.

A vectorial product of vectors in a space of three dimensions enables one to define basis

vectors for another space that is termed a <u>dual</u> or reciprocal space: this space is clearly a mathematical construct, which one can not envisage using ordinary senses. Whereas components of a vector in direct space might have dimensions of length, speed et cetera, those in a dual space have corresponding dimensions of inverse length, inverse speed ...; for this reason the term reciprocal space is used. A common use of a dual space in a chemical context occurs in construction of a model to interpret data from experiments involving diffraction of xrays from a crystalline sample. For the most common space group of chemical compounds, *monoclinic*, and also for another space group *triclinic*, the unit cells lack mutually perpendicular axes; in such cases basis vectors are distinct from unit vectors in a cartesian system of coordinates.

   If basis vectors in direct space be a_1, a_2, a_3, basis vectors b_1, b_2, b_3 for its dual space are defined as follows:

$$b_i \cdot a_i = 1 ; \quad b_i \cdot a_j = 0 \ \ (i \neq j)$$

Thus, for example b_1 is orthogonal to both a_2 and a_3 and to each vector that lies in a plane containing a_2 and a_3: as b_1 has a direction parallel to $a_2 \times a_3$ this requirement is expressible as

$$b_1 = e \, (a_2 \times a_3);$$

in which e is determined so that $b_1 \cdot a_1 = 1$. For the remaining dual basis vectors, the analogous expressions are

$$b_2 = f(a_1 \times a_3) ; \quad b_3 = g\,(a_1 \times a_2) \ .$$

Basis vectors a_i and b_i can thus serve to define a location of a point in direct and reciprocal space. For a system of orthogonal cartesian coordinates, the basis vectors for the dual space are the same as those for the direct space, and are hence just orthogonal unit vectors.

   For a crystalline solid, we work with basis vectors c_i that lie along edges of a unit cell and that are in the same directions as respective vectors a_i: thus, for a unit cell of class neither monoclinic nor triclinic, with edges described with vectors

$$c_1 = 3\,\underline{i}, \ \ c_2 = \underline{i} + 2\,\underline{j}, \ c_3 = \underline{i} + \underline{j} + \underline{k},$$

basis vectors a_i become $\underline{i}$, $\underline{j}$ and $\underline{k}$ that are unit vectors in directions of c_i. A general lattice point in the direct lattice is therefore defined with a vector

$$n_1\,c_1 + n_2\,c_2 + n_3\,c_3,$$

in which n_1, n_2, n_3 are integers, whereas a general point in the lattice is defined by $a_1\,x + a_2\,y + a_3\,z$. We first evaluate unit vectors a_i; we then obtain basis vectors for a dual space on solving the defining equations above. Dual vectors B_i, associated with c_i, analogously define a unit cell in a reciprocal space, and take forms analogous to those involving b_i, a_i:

$$B_1 = p\,(c_2 \times c_3), \quad B_2 = q\,(c_1 \times c_3) , \quad B_3 = r\,(c_1 \times c_2)$$

thus facilitating evaluation of B_i.

   A <u>direct</u> or *outer product* of two vectors each in three spatial dimensions, one $\underline{u}$ with components x_1, x_2 and x_3 and another $\underline{v}$ with components y_1, y_2 and y_3, is defined in terms of a following expression that treats basis vectors e_i, with $i = 1, 2, 3$, and general components x_i or y_i as algebraic entities:

$$\underline{u}\,\underline{v} = (\, x_1\,e_1 + x_2\,e_2 + x_3\,e_3\,)(\, y_1\,e_1 + y_2\,e_2 + y_3\,e_3\,)$$

$$= \; x_1 \, y_1 \, e_1 \, e_1 + x_2 \, y_2 \, e_2 \, e_2 + y_3 \, y_3 \, e_3 \, e_3 + x_1 \, y_2 \, e_1 \, e_2 + x_2 \, y_1 \, e_2 \, e_1$$

$$+ \, x_1 \, y_3 \, e_1 \, e_3 + x_3 \, y_1 \, e_3 \, e_1 + x_2 \, y_3 \, e_2 \, e_3 + x_3 \, y_2 \, e_3 \, e_2$$

Each pair of vectors of form $e_i \, e_j$ is termed a <u>dyad</u>; expression $\underline{u} \, \underline{v}$ is termed a <u>dyadic</u>, which is a sum of two or more dyads each with components $x_i \, y_j$ as scalar coefficient and which is also a special tensor of second rank that resembles a matrix of order 3. Any dyadic is representable with nine components $a_{i,j}$ resulting from expansion of a dyadic in a form

$$\underline{u} \, \underline{v} = \sum a_{i,j} \, e_i \, e_j, \qquad\qquad \text{in which } a_{i,j} = x_i \, y_j \, ;$$

its components associated with each dyad bear the same relation to components x_i of a vector $\underline{x}$ defined in terms of basis vectors e_i. These components are generally expressed in the form of a matrix so that methods of matrix algebra are applicable to the handling of dyadics. In the preceding discussion of vectors, we introduce a concept of a representative of a vector as a column vector formed from its components, but, as also observed, its transposed representative as a row vector is required in evaluating a scalar product of two vectors or the norm of a single vector. Unit basis vectors, such as are associated with cartesian axes in a system in three spatial dimensions, are represented in the same way with column vectors, with elements of which one is unity and another two are zero. The sum of two dyadics is another dyadic of which each component is a sum of the corresponding components of the addend dyadics. A product of a dyadic with a scalar quantity yields another dyadic of which each component is a product of the original component with the scalar quantity. A scalar product of a dyadic with a vector generates another vector, whereas a scalar product of one dyadic with another, which is in general <u>non-commutative</u>, yields another dyadic. A product comprising a dyadic between two vectors yields a scalar quantity as result. A vectorial product of a dyadic with a vector yields another dyadic.

Premultiplying a column vector $\underline{b}$ with a row vector $\underline{a}$ generates a number n, or a symbolic quantity that, on evaluation, yields a number -- a scalar quantity -- according to a prescription to form a scalar or dot product $\underline{a} \cdot \underline{b}$:

$$[a_1 \quad a_2 \quad a_3] \cdot \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = n$$

In contrast, <u>pre-multiplication</u> of a row vector with a column vector, indicated generally with no mark or punctuation between symbols for vectors as in $\underline{a} \, \underline{b}$, generates a *dyadic* with a representative in the form of a square matrix. To illustrate this effect we take as a column vector the representative of unit vectors $\underline{i}$, $\underline{j}$ and $\underline{k}$, directed along axes x, y and z, respectively, and produce three dyads $\underline{ii}$, $\underline{ij}$ and $\underline{ik}$:

$$\underline{ii} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} [1 \quad 0 \quad 0] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\underline{ij} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\underline{ik} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

When a matrix represents a dyadic, that matrix can represent an operator in a particular system of coordinates, or a matrix can effect a change from one coordinate system to another, of which an orthogonal transformation is a special case. As an operator L is linear if, as presented above,

$$L(a_1 \underline{v}_1 + a_2 \underline{v}_2) = a_1 L \underline{v}_1 + a_2 L \underline{v}_2$$

in which appear scalar coefficients a_1, a_2 and vectors (or functions) $\underline{v}_1$, $\underline{v}_2$, the properties of a matrix imply that it conforms to this definition; a matrix can thus serve as a linear operator, even though a matrix is not intrinsically an operator. A similarity transformation relies on this property, and an orthogonal matrix remains orthogonal after an orthogonal transformation of the coordinates.

Another operation involving a vector and a matrix -- or a matrix representative of a tensor of second rank as a dyadic -- allows resolution of a vector into orthogonal components. For two vectors $\underline{a}$ and $\underline{b}$, neither zero, we seek to express $\underline{b}$ in terms of its components along $\underline{a}$ and perpendicular to $\underline{a}$; each such component constitutes a projection of one vector on the other. A projection of $\underline{b}$ along $\underline{a}$ is given either by this ratio of scalar products of vectors as a factor of $\underline{a}$,

$$\text{projection of } \underline{b} \text{ on } \underline{a} = ((\underline{b} \cdot \underline{a})/(\underline{a}^T \cdot \underline{a})) \underline{a}$$

or by a product of this ratio of matrices,

$$(\underline{a} \cdot \underline{a}^T)/(\underline{a}^T \cdot \underline{a})$$

as a factor of $\underline{b}$,

$$((\underline{a} \cdot \underline{a}^T)/(\underline{a}^T \cdot \underline{a})) \underline{b},$$

in which superscript T denotes a transpose of a matrix. In the latter case the numerator is a dyadic, or tensor, as described above, and the denominator is a scalar corresponding to a scalar product of a transpose of a vector with the original vector. This projection is a more general case than simply projecting onto a unit vector, explained above, but is equivalent because for a unit vector the denominator is just unity. A projection or component of $\underline{b}$ perpendicular to $\underline{a}$ is a difference of $\underline{b}$ with its projection along $\underline{a}$.

The length of a vector, or the distance between two points defined by given position vectors, must be defined in a space of three dimensions. For vector $n\underline{r}$ to be n times as long as $\underline{r}$ is readily measurable for distances along a given straight line, but a comparison of the lengths of vectors of disparate direction requires definition of a space that has defined lengths and angles, so possessing a metric, which implies a rule for comparison of the lengths of any two vectors and hence of assignment of a length to every vector by comparison with a unit vector. For three vectors $\underline{r}$, $\underline{s}$ and $\underline{t}$ in euclidean geometry, the scalar or dot product is symmetric, $\underline{r} \cdot \underline{s} = \underline{s} \cdot \underline{r}$; the distributive law is satisfied, $\underline{r} \cdot (\underline{s} + \underline{t}) = \underline{r} \cdot \underline{s} + \underline{r} \cdot \underline{t}$; the scalar product is proportional to the length of each vector, $\underline{r} \cdot (c\underline{s}) = c(\underline{r} \cdot \underline{s})$, and the square of the length is positive but zero for only a zero vector, $\underline{r} \cdot \underline{r} > 0$, for

all $\underline{r} > 0$, so $(\underline{r} \cdot \underline{r})^{\left(\frac{1}{2}\right)} \geq 0$. If two vectors be expressed in terms of [basis](#) vectors, $\underline{r} = \sum_{j=1}^{3} r_j \underline{e}_j$ and $\underline{s}$

$= \sum_{j=1}^{3} s_j \underline{e}_j$, their product becomes expressible as $\underline{r} \cdot \underline{s} = \sum_{k=1}^{3} \left(\sum_{j=1}^{3} r_j s_k \right) \underline{e}_j \cdot \underline{e}_k$; this scalar product

is hence a [bilinear](#) form in the vector components and the coefficients are scalar products of the basis vectors.

Here follow some vector identities in which $\underline{a}$, $\underline{b}$, $\underline{c}$ and $\underline{d}$ are vectors and $\underline{del}$ is the differential

operator $\dfrac{\partial}{\partial x} \underline{i} + \dfrac{\partial}{\partial y} \underline{j} + \dfrac{\partial}{\partial z} \underline{k}$ in which $\underline{i}, \underline{j}, \underline{k}$ are unit vectors along orthogonal axes, f, g are scalar

functions, $.$ denotes a scalar product and $\times$ denotes a vector product.

$$\underline{a} \cdot (\underline{b} \times \underline{c}) = \underline{b} \cdot (\underline{c} \times \underline{a}) = \underline{c} \cdot (\underline{a} \times \underline{b})$$
$$\underline{a} \times (\underline{b} \times \underline{c}) = \underline{b} (\underline{a} \cdot \underline{c}) - \underline{c} (\underline{a} \cdot \underline{b})$$
$$(\underline{a} \times \underline{b}) \cdot (\underline{c} \times \underline{d}) = (\underline{a} \cdot \underline{c})(\underline{b} \cdot \underline{d}) - (\underline{a} \cdot \underline{d})(\underline{b} \cdot \underline{c})$$
$$\underline{del} (f g) = f \underline{del} g + g \underline{del} f$$
$$\underline{del} \cdot (f \underline{a}) = \underline{a} \cdot (\underline{del} f) + f (\underline{del} \cdot \underline{a})$$
$$\underline{del} \times (f \underline{a}) = (\underline{del} f) \times \underline{a} + f (\underline{del} \times \underline{a})$$
$$\underline{del} (\underline{a} \cdot \underline{b}) = \underline{b} \times (\underline{del} \times \underline{a}) + \underline{a} \times (\underline{del} \times \underline{b}) + (\underline{b} \cdot \underline{del}) \underline{a} + (\underline{a} \cdot \underline{del}) \underline{b}$$
$$\underline{del} \cdot (\underline{a} \times \underline{b}) = \underline{b} \cdot (\underline{del} \times \underline{a}) - \underline{a} \cdot (\underline{del} \times \underline{b})$$
$$\underline{del} \times (\underline{a} \times \underline{b}) = \underline{a} (\underline{del} \cdot \underline{b}) + (\underline{b} \cdot \underline{del}) \underline{a} - \underline{b} (\underline{del} \cdot \underline{a}) - (\underline{a} \cdot \underline{del}) \underline{b}$$
$$\underline{del} \times (\underline{del} \times \underline{a}) = \underline{del} (\underline{del} \cdot \underline{a}) - \underline{del}^2 \underline{a}$$

According to Dirac's notation, a column vector $\underline{v}$, with components v_j, might be called a [ket](#) and represented as $|v>$; a particular component v_j is generated according to $<j|v>$. Transposition of a ket yields a [bra](#), denoted $<v|$. These quantities arise in quantum mechanics.

eigenvector and eigenvalue

Besides other mathematical operators such as difference operator $\Delta = f(x + h) - f(x)$ and differential operator D, both of which are discussed in relation to differentiation in chapter 3, sum operator $\Sigma = x_1 + x_2 + x_3 + ...$ and product operator $\Pi = x_1 x_2 x_3 ...$, both of which are introduced in chapter 1, a square matrix **A** can act as an operator; it so acts on a vector $\underline{v}$ of the same number of components as its number of columns or its column dimension, $\mathbf{A} \cdot \underline{v} = \underline{u}$, to generate another vector $\underline{u}$ of the same dimension, called an [image](#) vector. The result of that operation yields a vector with the direction either the same as or different from that of $\underline{v}$; if the direction be the same, whether or not with the same sense, that image vector is also a characteristic vector or [eigenvector](#), and the ratio of the lengths of eigenvector and original vector is called a characteristic value or eigenvalue.

An important [problem](#) involving a [square matrix](#) relates to a condition whereby for a certain [vector](#) *representative*, called an *eigenvector* or *proper vector* or *principal vector* or *characteristic vector*, a product of that column vector premultiplied with a matrix yields the same column vector multiplied by a constant, a [scalar](#) quantity; for matrix **A** of order n, scalar λ and vector $\underline{x}$ with n components, we express this condition symbolically as

$$\mathbf{A} \underline{x} = \lambda \underline{x}$$

For a square matrix **A** of order n, its characteristic matrix is $\lambda\mathbf{I} - \mathbf{A}$ in which **I** is an identity matrix also of order n; the characteristic polynomial of **A** is the determinant of $\lambda\mathbf{I} - \mathbf{A}$. The eigenvalues of **A** are the n complex roots of that characteristic polynomial; n numbers in this set, each counted with its proper multiplicity, are denoted $\lambda(A)$, and any particular such number is $\lambda_i(A)$. If all eigenvalues be real, a conventional ordering is $\lambda_1(A) \geq \lambda_2(A) \geq \ ... \geq \lambda_n(A)$, hence in descending order, but for various purposes the reverse order might be preferable. For real λ, the effect of **A** is to stretch -- for $\lambda > 1$, to shrink -- for $0 < \lambda < 1$, to stretch or to shrink and to invert -- for $\lambda < 0$, or to annihilate -- for $\lambda = 0$, vector $\underline{x}$. Such an eigenvalue equation implies that
$$(\mathbf{A} - \lambda\,\mathbf{I}\,)\,\underline{x} \ = \ \underline{0}\,,$$
in which $\underline{0}$ denotes a column vector with n components each of value zero. From this product of two quantities that equals zero, a trivial solution is that $\underline{x}$ is a vector with each component of value zero, which lacks physical interest; a non-trivial solution requires a determinant of a matrix resulting from subtraction of $\lambda\,\mathbf{I}$ from **A** to be zero:
$$\det(\,\mathbf{A} - \lambda\,\mathbf{I}\,) \ = \ |\,\mathbf{A} - \lambda\,\mathbf{I}\,| \ = \ 0$$
This condition yields a polynomial in λ of which the roots become the eigenvalues of matrix **A**. A symmetric square matrix has real eigenvalues, but an unsymmetric matrix might have complex eigenvalues. An invertible matrix has no zero eigenvalue, and a real matrix with no zero eigenvalue is invertible. If matrix **A** of order n have n linearly independent eigenvalues, it can be made diagonal, and vice versa. The number of times that a particular eigenvalue of a matrix occurs is called its algebraic multiplicity. Components x_i of each eigenvector $\underline{x}$ are determined only within a multiplicative scalar quantity; to impose normalization, $\sum_{i=1}^{n} x_i^{\,2} = 1$ suffices to fix the values of x_i. If a square matrix **A** have no eigenvalue equal to zero, that matrix is invertible; a zero vector can not be an eigenvector, but an eigenvalue can be zero.

The signature of an hermitian matrix is the surplus of positive over negative coefficients in any real diagonal matrix similar to the given one; this number equals the excess of positive over negative eigenvalues.

Each stochastic matrix has a unit eigenvalue.

A real symmetric matrix has these properties:

- the eigenvalues of a real symmetric matrix are real;

- the eigenvectors of a real symmetric matrix can invariably be chosen to be real;

- a real symmetric matrix is diagonalizable, such that it has eigenvalues that can form a similar matrix;

- eigenvectors of a real symmetric matrix corresponding to distinct eigenvalues are orthogonal, and

- each real symmetric matrix possesses a complete orthonormal set of eigenvectors; a set of vectors is a complete orthonormal set for a square matrix of order n if the set be orthonormal, if each vector be an eigenvector of that matrix, and if the set contain exactly n vectors.

Applications exist for which a generalized problem of eigenvalues requires solution: instead of an identity matrix **I** there appears another matrix **B** such that
$$\mathbf{A}\,\underline{x} = \lambda\,\mathbf{B}\,\underline{x} \qquad \text{and} \qquad |\,\mathbf{A} - \lambda\,\mathbf{B}\,| = 0$$
To solve either equation, which arises in various physical and chemical applications, one might expand the latter determinant to yield a <u>characteristic polynomial</u> $p(\lambda)$ of order n in λ; the solution of an equation of that polynomial set equal to zero, $p(\lambda) = 0$, produces up to n possible values of λ termed *eigenvalues* or *proper values* or *characteristic values* of matrix **A**. If distinct eigenvalues number less than the degree of that polynomial, degeneracy exists. For each value of λ, solution of linear equations $\mathbf{A}\,\underline{x} = \lambda\,\underline{x}$ or $\mathbf{A}\,\underline{x} = \lambda\,\mathbf{B}\,\underline{x}$ yields an associated *eigenvector*. If the eigenvalues of a square matrix be distinct, the corresponding eigenvectors are <u>linearly independent</u>. For a non-trivial solution, multiplication of $\underline{x}$ by an arbitrary constant N is also a solution of the eigenvalue problem; the direction of eigenvector $\underline{x}$ is hence well defined but not its magnitude; according to convention one might choose N to give a *normalized* eigenvector that has unit magnitude. For a square invertible matrix **A** with eigenvalue λ, a corresponding eigenvector $\underline{x}$ and positive integer n, $\mathbf{A}^{\,n}$ has eigenvalue λ^n and eigenvector $\underline{x}$, and $\dfrac{1}{\lambda}$ is an eigenvalue of $\mathbf{A}^{(-1)}$ with corresponding eigenvector $\underline{x}$. The characteristic polynomials of <u>similar</u> matrices are identical, and hence likewise their eigenvalues.

We consider further an eigenvalue problem with **B** taken to be identity matrix **I**; we first pre-multiply matrix **A** by the inverse of a matrix **V**, to be determined,
$$\mathbf{V}^{(-1)}\,\mathbf{A}\,\underline{x} = \lambda\,\mathbf{V}^{(-1)}\,\underline{x}$$
and insert a unit matrix in a form $\mathbf{I} = \mathbf{V}\,\mathbf{V}^{(-1)}$ to obtain
$$\mathbf{V}^{(-1)}\,\mathbf{A}\,\mathbf{V}\,\mathbf{V}^{(-1)}\,\underline{x} = \lambda\,\mathbf{V}^{(-1)}\,\underline{x}$$
If we let $\underline{y} = \mathbf{V}^{(-1)}\,\underline{x}$, $\underline{y}$ becomes an eigenvector of $\mathbf{V}^{(-1)}\,\mathbf{A}\,\mathbf{V}$ because
$$\mathbf{V}^{(-1)}\,\mathbf{A}\,\mathbf{V}\,\underline{y} = \lambda\,\underline{y} = \lambda\,\mathbf{I}\,\underline{y}$$
The final step to construct **V** so that $\mathbf{V}^{(-1)}\,\mathbf{A}\,\mathbf{V}$ is a diagonal matrix is termed *matrix diagonalization*, which is also an instance of a <u>similarity transformation</u>. An important result of this process is that a sum of eigenvalues of square matrix **A** is equal to the <u>trace</u> of matrix **A**, which is just a sum of elements of **A** along its principal diagonal. A square matrix **A** of order n is diagonalizable if **A** have n linearly independent eigenvectors and hence n distinct eigenvalues; there exist then an invertible matrix **P** and a diagonal matrix **D** such that $\mathbf{P}^{(-1)}\,\mathbf{A}\,\mathbf{P} = \mathbf{D}$ with the columns of **P** being n linearly independent eigenvectors of **A** and the diagonal elements of **D** being the corresponding eigenvalues of **A** in the same order.

For a case of a square matrix **A** of order 3 with these elements
$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$
and square matrix **C** that must have also order 3 of which the columns are the eigenvectors of **A**,

$$\begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,1} & c_{3,2} & c_{3,3} \end{bmatrix}$$

to form a diagonal matrix $\Lambda = \mathbf{C}^T . \mathbf{A} . \mathbf{C}$ with elements

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix},$$

column n of product $\mathbf{A} . \mathbf{C}$ must be

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \begin{bmatrix} c_{n,1} \\ c_{n,2} \\ c_{n,3} \end{bmatrix} = \begin{bmatrix} \lambda_n c_{n,1} \\ \lambda_n c_{n,2} \\ \lambda_n c_{n,3} \end{bmatrix}.$$

Product $\mathbf{A} . \mathbf{C}$ must then be equal to

$$\begin{bmatrix} \lambda_1 c_{1,1} & \lambda_2 c_{1,2} & \lambda_3 c_{1,3} \\ \lambda_1 c_{2,1} & \lambda_2 c_{2,2} & \lambda_3 c_{2,3} \\ \lambda_1 c_{3,1} & \lambda_2 c_{3,2} & \lambda_3 c_{3,3} \end{bmatrix} = \begin{bmatrix} c_{1,1} & c_{1,2} & c_{1,3} \\ c_{2,1} & c_{2,2} & c_{2,3} \\ c_{3,1} & c_{3,2} & c_{3,3} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \mathbf{C} \Lambda$$

The matrix to make $\mathbf{A}$ diagonal is thus $\mathbf{C}$ and the diagonal matrix that results is Λ.

Every square matrix is *similar* to an almost diagonal matrix, or precisely, a matrix in Jordan canonical form, hence having non-zero elements on only the main diagonal and the first diagonal above that principal diagonal called the superdiagonal; an element on that superdiagonal might be either zero or unity. A diagonal matrix is a matrix in Jordan canonical form for which all elements on the superdiagonal equal zero. For a general square matrix $\mathbf{A}$ of order n and $\mathbf{I}$ an identity matrix of the same order, we define a generalized problem of eigenvalues as

$$(\mathbf{A} - \lambda \mathbf{I})^k \underline{x} = \mathbf{0} ,$$

which applies to every square matrix $\mathbf{A}$; when a Jordan canonical form is diagonal, only case $k = 1$ matters, whereas, for a non-diagonal Jordan form, further exponents k must be used, with $k \leq n$ for n as the order of the largest Jordan block or submatrix that contains off-diagonal elements.

A singular value of a matrix is a positive square root $\sqrt{\lambda}$ of eigenvalue λ of a product of a transpose of a matrix with the original matrix, $\mathbf{A}^T \mathbf{A}$; the eigenvalues of a symmetric matrix of form $\mathbf{A}^T \mathbf{A}$ are invariably non-negative.

In a chemical context, modeling $3n - 6$ vibrations of an angular molecule containing n atomic centres requires matrix $\mathbf{A}$, and $\mathbf{B}$ if appropriate, to be specified in terms of data associated with motions involving displacements of lengths of chemical bonds and deformations of angles between chemical bonds. Components of a given vector $\underline{x}$ then correspond to internal coordinates that define a collective vibrational motion of atomic centres in a *normal mode* such that all atomic nuclei move in phase with frequency λ. Among other applications of a vector is a determination of electronic amplitude functions for atoms and molecules according to wave mechanics.

calculus with vectors

For a scalar function $f(x, y, z)$ in three spatial dimensions, a variation in f when point (x, y, z) is altered to point $(x+dx, y+dy, z+dz)$ is expressed as

$$df = \frac{\partial}{\partial x}f\,dx + \frac{\partial}{\partial y}f\,dy + \frac{\partial}{\partial z}f\,dz$$

which, with $d\underline{r} = (dx, dy, dz)$, is expressible as $df = \underline{Del}\ f.\ d\underline{r}$ with differential <u>vectorial</u> operator $\underline{Del} = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$. With

$$\underline{Del} = \underline{i}\frac{\partial}{\partial x} + \underline{j}\frac{\partial}{\partial y} + \underline{k}\frac{\partial}{\partial z}$$

in terms of unit vectors $\underline{i}, \underline{j}, \underline{k}$, vector

$$\underline{Del}\ f = \underline{i}\frac{\partial}{\partial x}f + \underline{j}\frac{\partial}{\partial y}f + \underline{k}\frac{\partial}{\partial z}f$$

becomes the <u>gradient</u> of f(x, y, z), expressible also as grad f, which constitutes a <u>vectorial field</u>; the effect of the gradient operator acting on a scalar field f is hence to yield a vectorial field $\underline{F}(x,y,z$) such that this vectorial field assigns a vector to each point in space. Like any vector, a vectorial function specifies a magnitude, direction and sense of a quantity at each point in space, and has components parallel to each coordinate axis in the particular system of coordinates.

For a real scalar variable t in interval $[t_1, t_2]$ to which we assign a vector $\underline{R}$, for that vectorial function of variable t over that interval, we express the components of $\underline{R}(t)$ as

$$\underline{R}(t) = \underline{i}\,R_x(t) + \underline{j}\,R_y(t) + \underline{k}\,R_z(t).$$

Differentiation of that vectorial function with respect to variable t yields

$$\frac{\partial}{\partial t}\underline{R}(t) = \underline{i}\frac{d}{dt}R_x(t) + \underline{j}\frac{d}{dt}R_y(t) + \underline{k}\frac{d}{dt}R_z(t);$$

Because a ratio $\frac{d}{dt}R_x(t) : \frac{d}{dt}R_y(t) : \frac{d}{dt}R_z(t)$ differs in general from a ratio $R_x : R_y : R_z$, vector $\frac{\partial}{\partial t}\underline{R}($ t) has a direction different from $\underline{R}$. If f(t), $\underline{R}(t)$ and $\underline{S}(t)$ be differentiable functions of t,

- $\frac{\partial}{\partial t}(\underline{R}(t) + \underline{S}(t)) = \frac{\partial}{\partial t}\underline{R}(t) + \frac{\partial}{\partial t}\underline{S}(t)$;

- $\frac{\partial}{\partial t}(f(t)\underline{R}(t)) = f(t)\frac{\partial}{\partial t}\underline{R}(t) + \underline{R}(t)\frac{d}{dt}f(t)$;

- $\frac{\partial}{\partial t}(\underline{R}(t).\underline{S}(t)) = \underline{R}(t).\frac{\partial}{\partial t}\underline{S}(t) + \underline{S}(t).\frac{\partial}{\partial t}\underline{R}(t)$, and

- $\frac{\partial}{\partial t}(\underline{R}(t)\mathbf{x}\,\underline{S}(t)) = \underline{R}(t)\mathbf{x}\frac{\partial}{\partial t}\underline{S}(t) + \underline{S}(t)\mathbf{x}\frac{\partial}{\partial t}\underline{R}(t)$.

A derivative of $\underline{R}(t)$ that has a constant magnitude but varying direction is a vector perpendicular to $\underline{R}(t)$. A second derivative of $\underline{R}(t)$ is

$$\frac{d^2}{dt^2}\mathrm{R}(t) = \underline{i}\frac{d^2}{dt^2}R_x(t) + \underline{j}\frac{d^2}{dt^2}R_y(t) + \underline{k}\frac{d^2}{dt^2}R_z(t),$$

and higher derivatives analogously. For a vectorial function that depends on both spatial and temporal variables,

$$\underline{R}(x, y, z, t) \ = \ \underline{i} \, R_x(x, y, z, t) \ + \ \underline{j} \, R_y(x, y, z, t) \ + \ \underline{k} \, R_z(x, y, z, t) \, .$$

If each point $P(x, y, z)$ in a region R of space have associated with it a scalar quantity $f(x, y, z)$, which is a scalar function, and a scalar field exists in region R, for that scalar function $f(x, y, z)$, the change in f that results from a point (x, y, z) becoming a point $(x + dx, y + dy, z + dz)$ is expressed as

$$d \, f(x, y, z) = \left(\frac{\partial}{\partial x} f(x, y, z) \right) dx + \left(\frac{\partial}{\partial y} f(x, y, z) \right) dy + \left(\frac{\partial}{\partial z} f(x, y, z) \right) dz$$

or

$$d \, f(x, y, z) = \underline{\text{Del}} \, f(x, y, z) \cdot d\underline{r} \ = \ \frac{\partial}{\partial x} f(x, y, z) \cdot dx \ + \ \frac{\partial}{\partial y} f(x, y, z) \cdot dy \ + \ \frac{\partial}{\partial z} f(x, y, z) \cdot dz$$

in which $d\underline{r} = (dx, dy, dz)$; for applications involving scalar or vectorial quantities, we define again a differential vectorial operator $\underline{Del}$ in three spatial dimensions with cartesian coordinates:

$$\underline{Del} = \underline{i} \, \frac{d}{dx} + \underline{j} \, \frac{d}{dy} + \underline{k} \, \frac{d}{dz}$$

This operator has no practical use or value by itself, but, when it operates on a scalar function $f(x, y, z)$, it yields a vectorial sum of gradients of f in the directions of unit basis vectors for the same system of cartesian coordinates; according to convention, this vectorial sum is called $\underline{grad \, f}$:

$$\underline{\text{grad}} \, f \ = \ \underline{\text{Del}} \, f \ = \ \underline{i} \, \frac{\partial}{\partial x} f \ + \ \underline{j} \, \frac{\partial}{\partial y} f \ + \ \underline{k} \, \frac{\partial}{\partial z} f$$

Thus $\underline{grad \, f}$, but neither $\underline{grad}$ nor f separately, is a vector: its components at a point are rates of change of scalar function f with distance along directions of coordinate axes at that point; its magnitude at that point is the maximum rate of change of that function with distance; its direction is that of a maximum rate of change of function f, and its sense is toward increasing values of function f. These conditions characterize a vectorial field, according to which a vector becomes assigned to each point in space. As a physical or geometrical interpretation of $\underline{grad} \, \phi$, we consider surfaces in a family over which ϕ has constant values, $f(x, y, z) = c$; for displacement $d\underline{r}_s$ on such a surface, $\underline{grad} \, f(x, y, z) \cdot d\underline{r}_s = 0$ because $f(x, y, z)$ is constant on such a surface. As that displacement $d\underline{r}_s$ is parallel to that surface, provided that $f(x, y, z) \neq 0$, vector $\underline{grad} \, f(x, y, z)$ must be perpendicular to that surface at a point at which $\underline{grad} \, f(x, y, z)$ is evaluated. Moreover,

$$d \, f(x, y, z) = \underline{\text{Del}} \, f(x, y, z) \cdot d\underline{r} \ = \ | \, \underline{\text{Del}} \, f(x, y, z) \, | \ | \, d\underline{r} \, | \cos(\theta)$$

in which θ is the angle subtended between the normal to the surface of constant $f(x, y, z)$ and displacement $d\underline{r}$. A maximum change in $f(x, y, z)$ per unit displacement occurs for $\theta = 0$, and thus in a direction that is normal to the surface of constant $f(x, y, z)$. Magnitude $| \, \underline{\text{Del}} \, f(x, y, z) \, |$ is equal to the normal derivative, $\dfrac{\partial}{\partial n} f(x, y, z)$, in which dn is a displacement normal to the surface.

Although we consider above only cartesian coordinates, this gradient operator in cylindrical and spherical coordinates has great importance; instances of its application appear in section 6.402

Vector $\underline{grad} \, f$ contains information necessary to calculate a rate of variation of $f(x, y, z)$ in any direction. In a direction having direction cosines $\cos(\alpha)$, $\cos(\beta)$, $\cos(\chi)$, the directional derivative with respect to an element ds of distance such that $ds = \sqrt{dx^2 + dy^2 + dz^2}$ is

$$\frac{\partial}{\partial s}f = \left(\frac{\partial}{\partial x}f\right)\cos(\alpha) + \left(\frac{\partial}{\partial y}f\right)\cos(\beta) + \left(\frac{\partial}{\partial z}f\right)\cos(\chi).$$

In a direction of unit vector $\underline{s}$, $\dfrac{\partial}{\partial s}f = \underline{s} \cdot \underline{grad\ f}$, which is the component of $\underline{grad\ f}$ in direction $\underline{s}$; hence

$$\frac{\partial}{\partial s}f = |\underline{grad\ f}|\cos(\phi)$$

in which ϕ is the angle between $\underline{grad\ f}$ and unit direction vector $\underline{s}$.

If a *vectorial field* $\underline{F}(x, y, z)$ be formed as a *gradient* of a *scalar* function, expressed as
$$\underline{F}(x, y, z) = \underline{Del}\ f(x, y, z)$$
that vectorial field $\underline{F}$ is <u>conservative</u>, and f(x, y, z) is called a <u>potential function</u> for $\underline{F}$. Fields such as gravitational in <u>mechanics</u> and coulombic or electrostatic in electricity, for which, in either case, force varies as an inverse square of distance between appropriate points, are conservative. Here f(x, y, z) is a scalar potential function, but a vectorial potential function $\underline{V}(x, y, z)$ can also exist if
$$\underline{F}(x, y, z) = \underline{Del} \times \underline{V}(x, y, z) \equiv curl\ \underline{V}(x,\ y,\ z),$$
involving a <u>vectorial product</u>, discussed below. A vector field for which $\underline{Del} \cdot \underline{R}(x, y, z) = 0$ is called <u>solenoidal</u>.

Vectorial operator $\underline{Del}$ operates also on a vector, to form both a <u>scalar</u>, or dot, product and a <u>vectorial</u>, or cross, product. For a vectorial field of formula $\underline{F}$ that we express in three dimensions as
$$\underline{F}(x, y, z) = Q(x, y, z)\,\underline{i} + R(x, y, z)\,\underline{j} + S(x, y, z)\,\underline{k}$$
so that <u>expressions</u> Q, R and S, each with dependences on x, y, z, have <u>partial derivatives</u> with respect to <u>coordinates</u> x, y and z, we define first a scalar product of $\underline{Del}$ operating on $\underline{F}$ as the <u>divergence</u> of $\underline{F}$; the corresponding term in typical mathematical or physical parlance is just $\underline{div}$.

$$div\ F \equiv \underline{Del} \cdot \underline{F} = \frac{\partial}{\partial x}Q + \frac{\partial}{\partial y}R + \frac{\partial}{\partial z}S$$

A scalar or dot product of two vectors yields a result that is a *scalar* expression; if a quantity operating to the left of a vectorial quantity be differential operator $\underline{Del}$, consistent with convention, a scalar product might result. The divergence operator that acts on a vectorial field hence produces a scalar field. A physical interpretation of this divergence is that div $\cdot$ ($\rho\,\underline{v}$) might represent the net flow, per unit volume and per unit time, of a compressible fluid of density ρ and velocity $\underline{v}$ from a volume element $d\tau$, which implies a decreased density of fluid inside that volume. A calculation of this outgoing flux presents an application of divergence, amplified as follows: if a vector $\underline{V}$ represent at each point in space the direction and magnitude of flow of a fluid of density ρ moving with velocity $\underline{v}$, the product $\rho\,\underline{v} = \underline{V}$, which is called the flux density, represents the total flow of fluid per unit cross section and per unit time. The outgoing flux or total loss of fluid per unit time from an elemental <u>parallelepiped</u> of volume $d\tau$ is a product $\underline{Del} \cdot \underline{V}$ $d\tau$, so that $\underline{Del} \cdot \underline{V}$ is the outgoing flux per unit volume. Vector $\underline{F}$ such that dif $F = 0$ everywhere is described as a solenoidal vector.

For a vectorial field $\underline{F}(x, y, z)$ and a path along curve c extending from point P_1 to point P_2, a

line integral is the integral of the component of $\underline{F}(x, y, z)$ along that path, which becomes

$$\int_c F(x, y, z) \, . \, dl$$

that, in cartesian components with $\underline{F}(x, y, z) = \underline{i}\, F_x + \underline{j}\, F_y + \underline{k}\, F_z$ and $d\underline{l} = \underline{i}\, dx + \underline{j}\, dy + \underline{k}\, dz$, becomes

$$\int_c F(x, y, z) \, . \, dl = \int F_x \, dx + \int F_y \, dy + \int F_z \, dz$$

For that vectorial field to be a gradient of a scalar function $f(x, y, z)$ such that $\underline{F}(x, y, z) = \underline{Del}\, f(x, y, z)$, the path integral depends on only the bounds of the path and is independent of the path between those bounds:

$$\int_c F(x, y, z) \, . \, dl = f(P_2) - f(P_1) \, .$$

For that reason, a path integral along a closed path for which P_2 is the same as P_1 is zero.

For a vectorial field $\underline{F}(x, y, z)$ and a surface S divided into many infinitesimal surface elements each of area ds_j, $1 \leq j \leq n$, with each surface element ds_j as a vector of magnitude ds_j and direction perpendicular to the surface at a particular point, a surface integral of vectorial field $\underline{F}(x, y, z)$ over surface S becomes

$$\int_S F(x, y, z) \, . \, ds$$

implying a double integral, that in cartesian components with $\underline{F}(x, y, z) = \underline{i}\, F_x + \underline{j}\, F_y + \underline{k}\, F_z$ and $d\underline{s} = \underline{i}\, ds_x + \underline{j}\, ds_y + \underline{k}\, ds_z$ becomes

$$\int_S F(x, y, z) \, . \, ds = \int F_x \, ds_x + \int F_y \, ds_y + \int F_z \, ds_z$$

in which each integral on the right side is likewise a double integral. For closed surface S, that

surface integral is denoted $o\int_S F(x, y, z) \, . \, ds$ with $d\underline{s}$ is directed outward from that surface; o

should appear superimposed on the integral sign in $\int d$ (ignore d), but that expression is impracticable in *Maple* text here.

For a vectorial field $\underline{F}(x, y, z)$ and a volume V in three spatial dimensions with dv its element of volume, the volume integral of $\underline{F}(x, y, z)$ over V is

$$\int_V F(x, y, z) \, dv = \underline{i} \int_V F_x \, dv + \underline{j} \int_V F_y \, dv + \underline{k} \int_V F_z \, dv$$

in which all integrals are implicitly triple integrals, each evaluated on performing three ordinary integrations involving only scalar formulae.

According to the divergence theorem or Gauss's theorem, we convert a volume integral over τ with integrand $\underline{Del} \, . \, \underline{V}$ and integrating element $d\, \tau$ to a surface integral of $\underline{V} \, . \, d\underline{S}$ over surface S,

with τ being a total volume enclosed by S; the volume integral involves values of flux density $\underline{V}$ throughout a volume τ enclosed by S, whereas the surface integral involves values of $\underline{V}$ only on that surface S. For the divergence of a sum of two vectors,

$$\underline{Del} \cdot (\, \underline{R}(x, y, z) + \underline{S}(x, y, z)\,) \; = \; \underline{Del} \cdot \underline{R}(x, y, z) \; + \; \underline{Del} \cdot \underline{S}(x, y, z\,)$$

and the divergence of a product of a scalar function and a vectorial function,

$$\underline{Del} \cdot (\, f(x, y, z)\, \underline{R}(x, y, z)) \; = \; f(x, y, z)\, \underline{Del} \cdot \underline{R}(x, y, z) + \underline{R}(x, y, z) \cdot \underline{Del}\, f(x, y, z)$$

This vectorial or cross product $\underline{Del} \times \underline{F}$ involving vectorial operator $\underline{Del}$ and vectorial function $\underline{F}$ defines $\underline{curl\ F}$, or $\underline{rot\ F}$, which is a vector function curl expressible as a tensor of second rank according to its nature as an axial-vectorial or pseudovectorial operator. Operating on the left of a vector field $\underline{F}$, we represent it from its definition of a vector product in cartesian coordinates in determinantal form:

$$\underline{curl\ F} \equiv \underline{Del} \ \textbf{x}\ \underline{F} = \; \begin{vmatrix} \underline{i} & \underline{j} & \underline{k} \\ \dfrac{d}{dx} & \dfrac{d}{dy} & \dfrac{d}{dz} \\ Q & R & S \end{vmatrix}$$

In this determinant appear unit vectors in the first row, partial differential operators as components of $\underline{Del}$ in cartesian coordinates in the second row, and components of $\underline{F}$ -- scalar functions of coordinates x, y, z -- in the third row, as $\underline{F} = Q\ \underline{i} + R\ \underline{j} + S\ \underline{k}$. Here operator $\textbf{x}$ connotes only a sense of a vectorial product, because $\underline{Del}$ is not a vector but merely a vectorial operator; although perhaps ambiguous, this determinantal notation is a useful mnemonic for a formula

$$\underline{curl\ F} = \underline{Del} \ \textbf{x}\ \underline{F} = (\frac{\partial}{\partial y}\, S\, - \frac{\partial}{\partial z}\, R\,)\,\underline{i}\; + \; (\frac{\partial}{\partial z}\, Q\, - \frac{\partial}{\partial x}\, S\,)\,\underline{j}\; + \; (\,\frac{\partial}{\partial x}\, R\, - \frac{\partial}{\partial y}\, Q\,)\,\underline{k}$$

that serves as a *vector* function; with *Maple*'s package **VectorCalculus** and cartesian coordinates, these unit vectors $\underline{i}, \underline{j}$ and $\underline{k}$ are expressed as $\textbf{e}_x$, $\textbf{e}_y$ and $\textbf{e}_z$ respectively.

For the flow of a fluid, a velocity field with $\underline{curl\ v} \neq 0$ has with it an associated rotational or whirling motion. A vectorial field with zero curl is thus irrotational.

A vectorial product of $\underline{Del}$ with itself operating on scalar function f is $\underline{Del} \ \textbf{x}\ \underline{Del}\, f\ =\ \underline{curl} \ \textbf{x}$ $\underline{grad}\, f\ =\ 0$ for any f. For any vector function $\underline{F}$ for which $\underline{Del} \ \textbf{x}\ \underline{F}\ =\ 0$, which signifies that $\underline{F}$ is irrotational, $\underline{F}$ is therefore expressible as a gradient of a scalar function f, i.e. $\underline{F} = \underline{grad}\, f$. For any three vectorial functions $\underline{F}$, $\underline{G}$ and $\underline{H}$, because

$$\underline{F} \ \textbf{x}\ \underline{G} \cdot \underline{H}\ =\ \underline{G} \ \textbf{x}\ \underline{H} \cdot \underline{F}\ =\ \underline{H} \ \textbf{x}\ \underline{F} \cdot \underline{G},$$

for cyclic permutations, $\underline{F} \ \textbf{x}\ \underline{G} \cdot \underline{H}\ =\ -\ \underline{F} \ \textbf{x}\ \underline{H} \cdot \underline{G}$ et cetera, with the above result

$$\underline{Del} \cdot \underline{Del} \ \textbf{x}\ \underline{F} =\ \underline{div}\ \underline{curl}\ \underline{F}\ =\ 0,$$

and

$$\underline{Del} \ \textbf{x}\ \underline{Del} \ \textbf{x}\ \underline{F}\ =\ \underline{curl}\ \underline{curl}\ \underline{F}\ = \underline{Del}\,(\,\underline{Del} \cdot \underline{F}\,) - \underline{Del}^{\,2}\ \underline{F}.$$

The important result $\underline{curl}\ \underline{grad}\, f\ =\ 0$ is related directly to the independence of a value of a line integral on a path of integration between two points in pairs: as the differential

$$\underline{u} \cdot d\underline{r} = u_x\, dx + u_y\, dy + u_z\, dz$$

is exact and independent of that path of integration only if

$$\frac{\partial}{\partial y}\, u_x = \frac{\partial}{\partial x}\, u_y\ ,\quad \frac{\partial}{\partial z}\, u_x = \frac{\partial}{\partial x}\, u_z\ ,\quad \frac{\partial}{\partial z}\, u_y = \frac{\partial}{\partial y}\, u_z$$

which are precisely the conditions under which vector field $\underline{u}$ is irrotational.

For l a path that encloses a surface S, according to a theorem of Stokes, the integral of $\underline{u} \cdot d\underline{r}$ along that path equals a double integral of $\underline{Del} \times \underline{u}$ over that surface with differential $d\underline{S} = \underline{n}\, dA$; here $\underline{n}$ is the outward normal to the surface at a location of a surface element of area dA, and $d\underline{S}$ is called a vectorial element of surface. From the result of that integral according to the theorem of Stokes, if $\underline{Del} \times \underline{u} = 0$ for all points in space, the path integral on the left must be zero and $\underline{u} \cdot d\underline{r}$ is exact. The conditions expressed as three equalities of derivatives above constitute a special case of that theorem.

According to the equation that governs the diffusion of a substance through an isotropic medium as specified by concentration c of that substance,

$$\frac{\partial}{\partial t} c = D\,\underline{Del}^{\,2}\, c$$

net diffusion occurs perpendicular to a surface across which the concentration is constant, $c(x, y, z) = constant$, and in a direction of decreasing concentration such that diffusion occurs down a gradient of concentration; proportionality factor D is called the diffusion coefficient and $\underline{Del}^{\,2}$ is the laplacian operator explained further below. For values of c not too large, Fick's law expresses that a product of concentration c and a linear velocity $\underline{v}$ of diffusion is proportional to the negative gradient of that concentration; the proportionality factor is again D:

$$c\,\underline{v} = -\,D\,\underline{Del}\, c$$

The loss of mass per unit volume is $\dfrac{\partial}{\partial t} c = -\,\underline{Del} \cdot (\,-\,D\,\underline{Del}\, c\,) = D\,\underline{Del} \cdot (\,\underline{Del}\, c\,)$; with diffusion coefficient D independent of location, we hence obtain the diffusion equation above.

Among three further operators in package **VectorCalculus**, a jacobian is useful in transformations between systems of coordinates; if cartesian coordinates x and y be functions of other coordinates u and v through $x = f(u, v)$ and $y = g(u, v)$, a jacobian of x and y with respect to u and v, denoted $\dfrac{d(x, y)}{d(u, v)}$, is the determinant of a matrix, called a jacobian matrix, of partial derivatives of first order.

$$\begin{bmatrix} \dfrac{\partial}{\partial u} x & \dfrac{\partial}{\partial v} x \\[2ex] \dfrac{\partial}{\partial u} y & \dfrac{\partial}{\partial v} y \end{bmatrix}$$

Because of a property of a jacobian that

$$\begin{bmatrix} \dfrac{\partial}{\partial u} x & \dfrac{\partial}{\partial v} x \\[2ex] \dfrac{\partial}{\partial u} v & \dfrac{d}{dv} v \end{bmatrix} = \left(\dfrac{\partial}{\partial u} x \right)\Bigg|_{v = c}$$

for v held constant at value c, any partial derivative is expressible as a jacobian; an application of this result to chemical thermodynamics serves to extend the derivation of relations between state functions involved in Maxwell's relations, discussed in section 5.202.

As previewed in section 5.108, for function $f(x, y)$ of two variables, an hessian is a symmetric

matrix of partial derivatives of second order with respect to variables x and y:

$$\begin{bmatrix} \dfrac{\partial}{\partial x}\left(\dfrac{\partial}{\partial x}\,\mathrm{f}(x,y)\right) & \dfrac{\partial}{\partial x}\left(\dfrac{\partial}{\partial y}\,\mathrm{f}(x,y)\right) \\[2ex] \dfrac{\partial}{\partial y}\left(\dfrac{\partial}{\partial x}\,\mathrm{f}(x,y)\right) & \dfrac{\partial}{\partial y}\left(\dfrac{\partial}{\partial y}\,\mathrm{f}(x,y)\right) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial^2}{\partial x^2}\,\mathrm{f}(x,y) & \dfrac{\partial^2}{\partial x\,\partial y}\,\mathrm{f}(x,y) \\[2ex] \dfrac{\partial^2}{\partial x\,\partial y}\,\mathrm{f}(x,y) & \dfrac{\partial^2}{\partial y^2}\,\mathrm{f}(x,y) \end{bmatrix}$$

Operator laplacian, commonly written as Del^2, is not a vectorial operator although it can be formed as _Del_ **.** _Del_ that is a scalar product of _Del_ with itself; the latter form of a laplacian implies a compound operator _div grad_ described above that acts in three dimensions, according to cartesian coordinates, on a scalar function $\mathrm{f}(x,y,z)$ to yield

$$Del^2\,\mathrm{f}(x,y,z) = \frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\,\mathrm{f}(x,y,z)\right) + \frac{\partial}{\partial y}\left(\frac{\partial}{\partial y}\,\mathrm{f}(x,y,z)\right) + \frac{\partial}{\partial z}\left(\frac{\partial}{\partial z}\,\mathrm{f}(x,y,z)\right)$$

$$= \left(\frac{\partial^2}{\partial x^2}\,\mathrm{f}(x,y,z)\right) + \left(\frac{\partial^2}{\partial y^2}\,\mathrm{f}(x,y,z)\right) + \left(\frac{\partial^2}{\partial z^2}\,\mathrm{f}(x,y,z)\right).$$

tensor

In mathematics, a tensor is a geometric entity of a particular kind, or a generalized *quantity*; a tensor concept includes the ideas of scalar, vector and linear operator. Tensors are expressible in terms of coordinate systems, as arrays of scalars, but are defined so as to be independent of a particular frame of reference. Although tensors are represented as components in multi-dimensional arrays, the justification of a tensor theory is to explain the further implications of stating that a quantity is a tensor beyond that it comprises indexed components of some number. In particular, tensors behave in specific manners under transformations of coordinates; the abstract theory of tensors is a branch of linear algebra called multilinear algebra. In a physical or chemical context, a tensorial property, unlike a scalar property, expresses a dependence on a direction of a response of molecules or solid samples to an external stress. In an isotropic medium -- in which properties are independent of direction, a vector such as force $\underline{F}$ is related to another vector such as acceleration $\underline{a}$ through a formula involving a scalar quantity m as factor of proportionality, as in

$$\underline{F} = m\,\underline{a}\,,$$

with m denoting inertial mass. For an anisotropic medium, the response of one vector to another vectorial property might depend strongly on direction; in this case, a scalar quantity is inadequate to describe the response relation, and a tensor must serve instead. A tensorial property thus depends on the orientation of a system of interest. A tensor that might represent a molecular property has generally multiple elements or components that depend on the orientation of that molecule with respect to axes in a system of coordinates, typically cartesian, although the molecular property is independent of coordinates according to a particular system. What a chemist or physicist might generally call a tensor is a tensor field, such as a tensor for stress of a body to which a torque is applied or a tensor for moment of inertia.

A quantity $a_{i,j,k,\,\ldots}^{(r,\,s,\,t,\,\ldots)}$, having lower indices or subscripts $i, j, k, \ldots$ numbering p and upper indices or superscripts $r, s, t, \ldots$ numbering q, for which each index takes values $1, 2, 3, \ldots, n$, in a set, might represent a tensor of orders (p, q), or a component of such a tensor. In a system of

cartesian coordinates in three spatial dimensions, a polar vector is such a tensor; a vector $\underline{v} = v_x\,\underline{i} +$ $v_y\,\underline{j} + v_z\,\underline{k}$, in which $\underline{i}$, $\underline{j}$ and $\underline{k}$ are unit vectors along positive x, y and z axes or Ox, Oy and Oz respectively, is specifically a tensor of first rank, for which each component of the three carries an index for one of three directions in space. A cartesian tensor of second rank requires two such indices for each component, hence amounting to nine components in total. A scalar quantity, which is invariant to a transformation, is considered to constitute a tensor of rank zero, a (polar) vector a tensor of rank unity, an axial vector and a dyadic a representation of a tensor each of rank two, and a polyadic a representation of a tensor of rank greater than two. A representation of a physical property, such as electric dipolar polarizability, as a tensor of second rank resembles a square matrix of order three; the mathematical operations involving such a tensor also resemble those of a matrix. From an algebraic point of view, a tensor is an extension of a concept of vector and matrix to an array of higher order; For cartesian tensors we devote attention to only tensors with indices appearing in subscript form, but in systems of other than orthogonal coordinates a more general tensorial form might be required.

A matrix, representing a tensor of rank two, serves as a means to effect a transformation or mapping. For two vector spaces U and V, a transformation T of U into V, expressed as $T: U \rightarrow V$, is a rule that assigns to each vector $\underline{u}$ in U a unique vector $\underline{v}$ in V. The domain of T is U; because T $(\underline{u}) = \underline{v}$, the image of $\underline{u}$ under T is $\underline{v}$. For example, for a transformation $T: \underline{\mathbf{R}}^3 \rightarrow \underline{\mathbf{R}}^2$ defined by
$$T(x, y, z) = (2\,x, y + z),$$
for which the domain of T is $\underline{\mathbf{R}}^3$, the image of vector $(1,2,-3)$ is $(2,-1)$.

For a vector space two operations are defined -- addition and scalar multiplication. The most important transformation between vector spaces preserves linear structures as follows: for two vector spaces U and V with vectors $\underline{u}_1$ and $\underline{u}_2$ in U and scalar c, for a linear transformation T: $U \rightarrow V$,
$$T(\underline{u}_1 + \underline{u}_2) \;=\; T(\underline{u}_1) \;+\; T(\underline{u}_2)$$
$$T(c\,\underline{u}_2) \;=\; c\,T(\underline{u}_2)$$
Of these two conditions, the former implies that T maps a sum of two vectors into a sum of the images of those vectors, and the latter implies that T maps a scalar multiple of a vector into the same scalar multiple of the image; such a mapping preserves the operations addition and scalar multiplication. For matrix $\mathbf{M}$ of dimensions $m \times n$, and column matrix $\mathbf{x}$ (or vector) an element in $\underline{\mathbf{R}}^n$, a mapping T: $\underline{\mathbf{R}}^n \rightarrow \underline{\mathbf{R}}^m$ defined in $T(\mathbf{x}) = \mathbf{M} \cdot \mathbf{x}$ is linear, and is called a *matrix transformation*. For example, for matrix $\mathbf{M} = \begin{bmatrix} 3 & 1 \\ 2 & -2 \\ 1 & 1 \end{bmatrix}$ and vector $\underline{v} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$ a linear transformation $T : \underline{\mathbf{R}}^2 \rightarrow \underline{\mathbf{R}}^3$

yields

$$T\left(\begin{bmatrix} v_x \\ v_y \end{bmatrix}\right) = \begin{bmatrix} 3 & 1 \\ 2 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} 3\,v_x + v_y \\ 2\,v_x - 2\,v_y \\ v_x + v_y \end{bmatrix}$$

which according to the specified vector $\underline{v} = \begin{bmatrix} 4 \\ -2 \end{bmatrix}$ with $v_x = 4$ and $v_y = -2$ becomes $\begin{bmatrix} 10 \\ 12 \\ 2 \end{bmatrix}$.

   A non-singular linear transformation is important because it preserves the linearity of a vector space in the sense that it transforms

- lines into lines,

- segments of lines into segments of lines,

- parallel lines into parallel lines, and

- lines through the origin into lines through the origin.

   Under an orthogonal transformation, which is a mapping with an orthogonal matrix that preserves linearity, a scalar conforms to a law

$$\phi' = \phi,$$

whereas a vector conforms to a law

$$\underline{v}' = \sum_{i=1}^{3} R_{i,j} v_j$$

or, in matrix notation,

$$\underline{v}' = \mathbf{R} \cdot \underline{v}$$

with components and elements explicitly expressed as

$$\begin{bmatrix} v_x' \\ v_y' \\ v_z' \end{bmatrix} = \begin{bmatrix} R_{x,x} & R_{x,y} & R_{x,z} \\ R_{y,x} & R_{y,y} & R_{y,z} \\ R_{z,x} & R_{z,y} & R_{z,z} \end{bmatrix} \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}.$$

in which appears matrix $\mathbf{R}$ that rotates axes of coordinates in cartesian systems. In a cartesian space of three dimensions, one defines a tensor of rank N as a quantity having 3^N components $T_{i,j,k,\ldots}$, with N subscripts, that transforms as

$$T_{i,j,k,\ldots}' = \sum_{l=1}^{3} \left(\sum_{m=1}^{3} \left(\sum_{n=1}^{3} R_{i,l} R_{j,m} R_{k,n} \ldots T_{l,m,n,\ldots} \right) \right)$$

with accordingly N sums. Consistent with this definition, a scalar is a tensor of rank zero and a vector is a tensor of rank one. The most common for chemists, and most tractable, other case is a tensor of rank two:

$$T_{i,j}' = \sum_{k=1}^{3} \left(\sum_{l=1}^{3} R_{i,k} R_{j,l} T_{k,l} \right)$$

A distinction between a tensor of rank two and a square matrix of order three is that a tensor is defined only in terms of its properties under transformation of coordinates, whereas a matrix can contain arbitrary elements. A matrix restricted to undergo an orthogonal transformation is equivalent to, and serves as a representative of, a tensor; as an orthogonal matrix is non-singular, it preserves the linearity of a vector space, as described above. Components of a tensor of rank two and elements of a matrix are manipulated in a similar fashion, and each equation for a tensor of rank two might correspond to a corresponding equation for a matrix. All terminology and

operations of matrix algebra, such as transpose, hermitian, antisymmetric et cetera, are applicable to these tensors without modification. We express the latter equation in a matrix notation,

$$\mathbf{T'} = \mathbf{R}\ \mathbf{T}\ \mathbf{R}^T$$

or

$$
\begin{bmatrix}
T_{x,x}{}' & T_{x,y}{}' & T_{x,z}{}' \\
T_{y,x}{}' & T_{y,y}{}' & T_{y,z}{}' \\
T_{z,x}{}' & T_{z,y}{}' & T_{z,z}{}'
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
R_{x,x} & R_{x,y} & R_{x,z} \\
R_{y,x} & R_{y,y} & R_{y,z} \\
R_{z,x} & R_{z,y} & R_{z,z}
\end{bmatrix}
\begin{bmatrix}
T_{x,x} & T_{x,y} & T_{x,z} \\
T_{y,x} & T_{y,y} & T_{y,z} \\
T_{z,x} & T_{z,y} & T_{z,z}
\end{bmatrix}
\begin{bmatrix}
R_{x,x} & R_{y,x} & R_{z,x} \\
R_{x,y} & R_{y,y} & R_{z,y} \\
R_{x,z} & R_{y,z} & R_{z,z}
\end{bmatrix}
$$

in which $\mathbf{R}$ is again a rotation matrix and $\mathbf{R}^T$ is its transpose. As an example of such a rotation matrix, to effect rotation about axis z, i.e. within plane xy such that if we rotate axes x and y counterclockwise through angle θ about axis z, and if we label the transformed axes as X and Y, the following matrix operates on coordinates of a fixed point P in systems of coordinates to rotate (x, y, z) into (X, Y, Z), with $z = Z$, as illustrated in section 5.109.

$$
\mathbf{R} =
\begin{bmatrix}
R_{x,x} & R_{x,y} & R_{x,z} \\
R_{y,x} & R_{y,y} & R_{y,z} \\
R_{z,x} & R_{z,y} & R_{z,z}
\end{bmatrix}
=
\begin{bmatrix}
\cos(\theta) & \sin(\theta) & 0 \\
-\sin(\theta) & \cos(\theta) & 0 \\
0 & 0 & 1
\end{bmatrix}
$$

For this matrix $R^T = R^{(-1)}$.

In euclidean geometry, a rotation is an instance of an isometry, which is a transformation that moves points without altering the distances between them. Rotations are distinguished from other isometries according to properties that they leave at least one point fixed and that they leave chirality or handedness unaltered. In contrast, a translation moves each point, a reflexion exchanges left-handed and right-handed ordering, and a *glide reflexion* effects both properties. A rotation that alters handedness is an improper rotation or a *rotoinversion*, corresponding to a rotation and an inversion through a centre of symmetry; such an operation enters the discussion of the properties of unit cells of crystalline substances. The product of two rotation matrices is likewise a rotation matrix because $(\mathbf{R}_1\,\mathbf{R}_2)^t\,(\mathbf{R}_1\,\mathbf{R}_2) = \mathbf{R}_2{}^t\,(\mathbf{R}_1{}^t\,\mathbf{R}_1)\,\mathbf{R}_2 = \mathbf{I}$, the identity matrix, and the determinant of the product of two matrices is equal to the product of the determinant of the separate matrices, $|\mathbf{R}_1\,\mathbf{R}_2| = |\mathbf{R}_1|\,|\mathbf{R}_2|$. For $n > 2$, a multiplication of rotation matrices of order n is not commutative, but a rotation matrix commutes with its transpose.

In a system of cartesian coordinates with spatial directions with labels x, y, z, a vector such as electric dipolar moment $\underline{p} = \underline{i}\,p_x + \underline{j}\,p_y + \underline{k}\,p_z$ is a tensor of first rank, with each of its three components p_x, p_y, p_z indexed according to one of those spatial directions. Like a vector, a tensor of rank greater than first is characterized by the way in which its components transform between coordinate systems. A more typical tensorial quantity is a tensor of second rank, which requires two indices, only subscripts in our usage for a cartesian tensor, for each component; as each index runs over three spatial dimensions, there are thus nine components, and this tensor of second rank

thus resembles, or is represented by, a square matrix of order three. To convert such a tensor of second rank into its most meaningful form resembles working with a quadratic form that might likewise be converted into a form with fewest terms.

Quadratic form Q in two dimensions x, y is expressible compactly in matrix notation as
$$Q = \mathbf{x}^T . \mathbf{A} . \mathbf{x}.$$
We seek a transformation of axes that reduces $\mathbf{A}$ to diagonal form. Thus, as
$$\mathbf{U} . \mathbf{U}^T = \mathbf{U}^T . \mathbf{U} = \mathbf{I}$$
with $\mathbf{I}$ an identity matrix of order two, we have
$$Q = \mathbf{x}^T . \mathbf{U}^T . \mathbf{U} . \mathbf{A} . \mathbf{U}^T . \mathbf{U} . \mathbf{x} = \mathbf{X}^T . \lambda . \mathbf{X}$$
in which λ is a diagonal matrix of eigenvalues; components of $\mathbf{X}$ yield new coordinates of P in a system of rotated axes.   As $\mathbf{A}$ is a symmetric matrix, we choose $\mathbf{U}$ such that $\mathbf{U} . \mathbf{A} . \mathbf{U}^T$ is diagonal: $\mathbf{U}^T$ is a transposed matrix of eigenvectors of $\mathbf{A}$. A method to determine $\mathbf{U}$ and to effect transformation of principal axes becomes understood on considering a specific example, which concludes with a graphical realization of a transformation of axes.

A quadratic form is thus an expression of form $\underline{x}^T . \mathbf{A} . \underline{x}$ in which appears square matrix $\mathbf{A}$, column matrix $\underline{x}$ and its transpose $\underline{x}^T$; this expression is a general extension of expressions in two dimensions of form
$$f(x, y) = a\,x^2 \qquad \text{or} \qquad a\,x^2 + b\,y^2 + 2\,c\,x\,y$$
et cetera. Matrix $\mathbf{A}$, the matrix of this quadratic form, is relatable to an hessian of $f(x, y)$. A stationary point or critical point is classified as a maximum, minimum or col depending on whether the quadratic form is positive, negative or positive in some direction and negative in another direction. On the basis of quadratic forms and eigenvalues of a matrix, the following classification of that matrix becomes feasible:

eigenvalues	quadratic form	description
all positive	> 0	positive definite
all negative	< 0	negative definite
non-negative	≥ 0	positive semidefinite
non-positive	≤ 0	negative semidefinite
positive and negative	positive and negative	indefinite

For a positive definite hessian, a stationary point is a minimum, or a maximum for a negative definite hessian. For both positive and negative eigenvalues a critical point is a col. With a zero eigenvalue, the point is a minimum for a positive semidefinite matrix or a maximum for a negative semidefinite matrix.

Components of a covariant tensor of rank two commonly satisfy a relation $T_{m,n} = T_{n,m}$ whereby such a tensor is symmetric: there exist only six independent components of this tensor for space of three dimensions. Because specification of a quadric surface, introduced in section 5.102, requires, after linear terms are eliminated by a translation of axes, six independent parameters in a reduced form, such as coefficients $a, b, ... f$ in
$$g(x, y, z) = a\,x^2 + b\,y^2 + c\,z^2 + d\,x\,y + e\,x\,z + f\,y\,z = 1 ,$$
a surface for which a defining equation is expressible equivalently as

$$[x \quad y \quad z] \begin{bmatrix} a & \dfrac{d}{2} & \dfrac{e}{2} \\ \dfrac{d}{2} & b & \dfrac{f}{2} \\ \dfrac{e}{2} & \dfrac{f}{2} & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = 1$$

represents uniquely any symmetric tensor of rank two. An example of a quantity that has a character of a symmetric tensor is the moment of inertia of a rotating body, even a rotating molecule.

A real symmetric tensor can provably be diagonalized with real eigenvalues; a procedure, called *transformation to principal axes*, for this purpose is identical to that of diagonalizing an hermitian or self-adjoint matrix. If a tensor be neither symmetric nor skew symmetric, a given tensor can be invariably written as a *sum* of a symmetric tensor and a skew symmetric tensor. A sum or difference of two or more tensors of the same rank and type is another tensor of the same rank and type. A tensor of second rank, possessing the form of a matrix, operates on a vector to yield another vector, such as is shown in the matrix of transformation above, whereas a tensor of third rank operates on a vector to yield a tensor of second rank.

As discussed above, a <u>dyadic</u>, which is a <u>direct</u> or outer product of two vectors, possesses elements the location of which within an array depends upon two indices that specify directions; such a mathematical object is an instance of a *tensor* of second rank that is represented by a matrix of third order. A tensor is characterized by its *rank* or *order*, the latter of which is distinct from order of a matrix: a tensor of which a matrix of order three is representative has rank two. A tensor of rank r in a space of d dimensions has d^r elements. A tensor of rank N having components along cartesian axes in three-dimensional space corresponds to an array containing 3^N numbers, but this array implies extension in N dimensions. An essential property of a tensor is a way that its components, referred to one system of coordinates, are related to its components in another system, according to a law of transformation, as distinct from an array or matrix of which elements might take arbitrary values. Recalling a link between a tensor of rank two, that one can represent as a matrix of order three, and a dyadic, we describe a *cartesian tensor* $\underline{\mathbf{T}}$ of rank two with respect to a particular system of cartesian coordinates x, y, z, with unit vectors respectively $\underline{i}, \underline{j}$, $\underline{k}$, as follows,

$$\underline{\mathbf{T}} = \begin{bmatrix} t_{x,x} & t_{x,y} & t_{x,z} \\ t_{y,x} & t_{y,y} & t_{y,z} \\ t_{z,x} & t_{z,y} & t_{z,z} \end{bmatrix}$$

or

$$\begin{aligned} \underline{\mathbf{T}} = \quad & t_{x,x}\,\underline{i}\,\underline{i} + t_{x,y}\,\underline{i}\,\underline{j} + t_{x,z}\,\underline{i}\,\underline{k} \\ + \; & t_{y,x}\,\underline{j}\,\underline{i} + t_{y,y}\,\underline{j}\,\underline{j} + t_{y,z}\,\underline{j}\,\underline{k} \\ + \; & t_{z,x}\,\underline{k}\,\underline{i} + t_{z,y}\,\underline{k}\,\underline{j} + t_{z,z}\,\underline{k}\,\underline{k} \end{aligned}$$

in which $\underline{i}\,\underline{i}$, $\underline{i}\,\underline{j}$ et cetera are dyads corresponding to direct or outer products of unit vectors $\underline{i}, \underline{j}$ et cetera.

A *cartesian tensor* $\underline{\mathbf{T}}$ of rank three is expressible as three matrices in a stack; each matrix is

square and of order three, and in each an element has three indices or subscripts, such as $t_{x,x,x}$; a corresponding expression in terms of unit vectors contains *triadics* $\underline{i}\,\underline{i}\,\underline{i}$, $\underline{i}\,\underline{i}\,\underline{j}$, $\underline{i}\,\underline{j}\,\underline{k}$ et cetera, with analogous expressions for tensors of greater order.

When a tensor represents a property of a system in which axes are not necessarily orthogonal, as occurs not only in relation to physical relativity, for instance, but also to crystals of monoclinic and triclinic classes, considerable complication arises because, although only two indices are required for a tensor of second rank, there are four combinations of their placement according to which a tensor can become subclassified as contravariant or covariant or mixed; here we avoid such complication.

Scalar products of basis vectors $\underline{a}_i \cdot \underline{a}_j$ occur in various formulae; for computational purposes these products have greater utility than either the vectors themselves or angles between them. A scalar quantity $g_{i,j} = \underline{a}_i \cdot \underline{a}_j$ becomes a component of a metric tensor, of second rank, in which *metric* implies that this tensor pertains to properties of measurement of a space; these scalar quantities contain information about the lengths of basis vectors and the angles between them, and their description as a tensor element implies properties of transformation from basis vectors in one set to those in another. This metric tensor is symmetric: with three physical dimensions, only six independent components of this tensor exist.

When vectorial operator $\underline{Del}$ is applied to a vectorial function in three spatial dimensions, the result is a tensor of second rank, as an axial or pseudovector. For vector function $\underline{F}$ with cartesian components, an expression of vector product $\underline{Del}\,\mathbf{x}\underline{F} = \mathbf{G}$ as a tensor, alternative to that presented above in the material on vector calculus, is

$$\underline{Del}\,\mathbf{x}\,\underline{F} = (\,\underline{i}\,\frac{\partial}{\partial x} + \underline{j}\,\frac{\partial}{\partial y} + \underline{k}\,\frac{\partial}{\partial z}\,)\,(\,\underline{i}\,F_x + \underline{j}\,F_y + \underline{k}\,F_z\,)$$

$$= \underline{i}\,\underline{i}\,\frac{\partial}{\partial x}F_x + \underline{i}\,\underline{j}\,\frac{\partial}{\partial x}F_y + \underline{i}\,\underline{k}\,\frac{\partial}{\partial x}F_z$$

$$+ \underline{j}\,\underline{i}\,\frac{\partial}{\partial y}F_x + \underline{j}\,\underline{j}\,\frac{\partial}{\partial y}F_y + \underline{j}\,\underline{k}\,\frac{\partial}{\partial y}F_z$$

$$+ \underline{k}\,\underline{i}\,\frac{\partial}{\partial z}F_x + \underline{k}\,\underline{j}\,\frac{\partial}{\partial z}F_y + \underline{k}\,\underline{k}\,\frac{\partial}{\partial z}F_z$$

As a cartesian tensor of second rank, $\mathbf{G}$ has thus nine components: $G_{xx} = \frac{\partial}{\partial x}F_x$, $G_{xy} = \frac{\partial}{\partial x}F_y$, et cetera. Operating twice with $\underline{Del}$ to the left of a vector, $\underline{Del}\,\mathbf{x}\,\underline{Del}\,\mathbf{x}\,\underline{F}$ generates a tensor of third rank with 27 components as G_{xxx}, G_{xxy}, et cetera to denote the corresponding third derivatives.

In chemistry, calculations of moments of inertia, molecular electric polarisability or electric quadrupolar moment involve properties of type tensor of rank two that can be represented with a matrix of order three, the trace of which is invariant under rotation of axes; an inertial tensor expresses a response, to a torque, of a molecule that has access to discrete rotational states whereas a polarisability tensor expresses a response of a molecule to an applied electric field. Like operations with matrices, the sum or difference of two or more tensors of the same rank and type is a further tensor of the same rank and type. A key problem is that a coordinate system that serves to define other molecular properties might yield a polarisability matrix of non-diagonal

form: conventional practice is to reorient axes, according to a linear transformation that yields a property matrix in diagonal form; such a *transformation to principal axes* is achieved on using eigenvectors of a property matrix to define coordinates according to a new system. Electric dipolar moment behaves simply like a *vector*, which is a tensor of rank one as each component of dipolar moment involves only one direction, whereas electric octupolar moment or first electric hyperpolarisability, each of which is a property depending upon three directions, is defined in terms of a tensor of rank three that is represented by matrices in stacks or layers, with indices for row, column and layer; for completeness, a *scalar* is formally considered to be a *tensor* of rank zero, as it is represented by a matrix 1 x 1, involves zero indices, and is invariant under rotation of coordinates. A molecular property of type *tensor* implies a multilinear function invariant to changes of coordinates; the most important property of a tensor is that its magnitude remains unaltered under some transformation of a measuring scale or frame, although its components vary with such a transformation. Components of a *tensor of rank two*, such as an inertial tensor of a molecule or molecular dipolar electric polarizability, are commonly represented in *matrix* form. Whereas a tensor of rank two is invariably expressible in matrix form, the elements of a general matrix need not transform in the same way as the components of a tensor. A cartesian tensor is based on coordinate axes in an orthonormal set. As a tensor of rank zero, a scalar quantity has no subscript or superscript, whereas a component of a polar vector as a tensor of first rank requires one subscript, or, in a less common depiction, one superscript; an axial vector, which arises as a result of a vector or cross product of two polar vectors, is, however, an antisymmetric tensor of second rank, as explained below. A component of a cartesian tensor of second rank requires in total two subscripts and is called covariant, or two superscripts and called contravariant, or one subscript and one superscript and called mixed, and can be represented with a matrix. As, for instance, the non-linear optical properties of crystals depend on first electric hyperpolarizability, which is a tensor of rank three, such quantities have chemical relevance. Among chemical and physical topics and properties of crystals susceptible to analysis according to tensor methods are diamagnetic and paramagnetic susceptibility, relative permittivity, double refraction, elasticity, electric conductivity, electric polarization, ferroelectricity, optical activity, photoelasticity, piezoelectricity and pyroelectricity, pyromagnetism, stress and strain, thermal conductivity and thermal expansion. If the conductance tensor for a crystal were not symmetric, the electric conduction in crystals of low symmetry would follow a spiral path. In analytical chemistry, an example of a tensor arises in the fluorescence spectrum of a sample comprising a mixture of emitting compounds, so that the measured intensity of fluorescent emission depends on the compound, the wave length of excitation and the wave length of emission, so represented with a tensor of order three; the chemical shift of a particular nucleus in a solid or oriented sample such as a liquid crystal is represented with a tensor of order two.

The electric dipolar moment $\underline{p}$ of a molecule in a space of three dimensions depends on strength of electric field; we write this moment as a sum of its contributions of various orders:

$$\underline{p} \;=\; \underline{p}^{\,(0)} \;+\; \underline{p}^{\,(1)} \;+\; \underline{p}^{\,(2)} \;+\; \underline{p}^{\,(3)} \;+\; \ldots$$

in which $\underline{p}^{\,(0)}$ is the permanent electric moment, a vector or a tensor of first rank. We express further terms in a power series,

$$p = p^{(0)} + \alpha \cdot \underline{E} + \frac{1}{2} \beta : \underline{E}\,\underline{E} + \frac{1}{6} \gamma :: \underline{E}\,\underline{E}\,\underline{E} + \dots$$

in which α that is a tensor of second rank and has the form of a matrix of order three is electric dipolar polarisability that produces a contribution to induced dipolar moment linear in strength $\underline{E}$ of electric field, β that is a tensor of third rank and has the form of three matrices, each of order three, in a stack is first electric dipolar hyperpolarisability that produces a contribution to induced dipolar moment quadratic in strength of electric field, γ that is a tensor of fourth rank is second electric dipolar hyperpolarisability that produces a contribution to induced dipolar moment cubic in strength of electric field, and so forth. A quantity $\underline{E}\,\underline{E}$ implies an outer product of vector $\underline{E}$ with itself, and $\underline{E}\,\underline{E}\,\underline{E}$ analogously. We express explicitly the sums for each contribution, for each component:

$$p_i^{(1)} = \sum_j \alpha_{i,j}\, E_j$$

$$p_i^{(2)} = \sum_j \left(\sum_k \beta_{i,j,k}\, E_j\, E_k \right)$$

$$p_i^{(3)} = \sum_j \left(\sum_k \left(\sum_l \gamma_{i,j,k,l}\, E_j\, E_k\, E_l \right) \right)$$

In a particular case of a tensor α for electric polarisability that relates the electric dipolar moment $\underline{p}$ induced in a molecule to an externally applied electric field $\underline{E}$ as described above, we express that dipolar moment as

$$\begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} \alpha_{x,x} & \alpha_{x,y} & \alpha_{x,z} \\ \alpha_{y,x} & \alpha_{y,y} & \alpha_{y,z} \\ \alpha_{z,x} & \alpha_{z,y} & \alpha_{z,z} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}.$$

Knowing the values of the nine components of this tensor α, one can calculate the polarization in a particular direction when an electric field is applied in any direction. The induced dipolar moment is exactly parallel to that applied field for a field applied along one of three directions that become the principal axes of the tensor. The three components of α along these directions are the principal components of the polarisability tensor, and define radii of a polarisability ellipsoid that describes how the polarisability varies as the molecule rotates; the radius of this ellipsoid in any direction specifies the magnitude of the dipolar moment when an electric field is applied in that direction. A tensor is decomposible into irreducible components according to their weights, which are orders of an associated Legendre polynomial in a spherical system. The properties of a tensor of second rank, such as that for polarisabililty, in relation to symmetry are the isotropic polarisability,

$$\frac{\alpha_{x,x} + \alpha_{y,y} + \alpha_{z,z}}{3},$$

three measures of anisotropy,

$$\frac{\alpha_{x,y} - \alpha_{y,x}}{2}, \quad \frac{\alpha_{x,z} - \alpha_{z,x}}{2} \quad \text{and} \quad \frac{\alpha_{y,z} - \alpha_{z,y}}{2}$$

and five symmetric components,

$$\frac{\alpha_{x,y} + \alpha_{y,x}}{2}, \quad \frac{\alpha_{x,z} + \alpha_{z,x}}{2}, \quad \frac{\alpha_{y,z} + \alpha_{z,y}}{2}, \quad \frac{\alpha_{x,x} - \alpha_{y,y}}{2}, \quad \alpha_{z,z} - \frac{\alpha_{x,x} + \alpha_{y,y} + \alpha_{z,z}}{3} \quad ;$$

terms in these three sets or representations are hence a scalar, an axial vector and a traceless symmetric tensor of second rank, with weights $j = 0$, 1 and 2 with $2j + 1$ components respectively, corresponding to theory for angular momentum.

A linear molecule has only two principal components $\alpha_{||}$ and $\alpha_{\perp}$ that are respectively parallel and perpendicular to the internuclear axis and that in general depend on the quantum state of that molecule; the isotropic polarisability is accordingly

$$\alpha = \frac{\alpha_{||} + 2\,\alpha_{\perp}}{3}$$

and the anisotropy is

$$\Delta\,\alpha = \alpha_{||} - \alpha_{\perp}$$

If $T_{m,n} = -T_{n,m}$, a tensor of rank two is skew, *skew symmetric* or *antisymmetric*; this condition implies that diagonal elements are zero and thus that only three independent components require evaluation. An axial vector or pseudovector is hence describable as an antisymmetric tensor of rank two; an antisymmetric tensor can represent an axial or pseudovector in only three spatial dimensions, and then only in rectangular components. An antisymmetric tensor of rank two is characterized by three independent quantities as components, such as $T_{x,y}$, $T_{y,z}$, $T_{z,x}$; an axial vector or pseudovector $\underline{A}$ can be associated with such a tensor according to relations

$$A_x = -T_{y,z}, \quad A_y = -T_{x,z}, \quad A_z = -T_{x,y}$$

or explicitly

$$\begin{bmatrix} 0 & -A_z & A_y \\ A_z & 0 & -A_x \\ -A_y & A_x & 0 \end{bmatrix}$$

A property of a tensor being symmetric or antisymmetric is unaltered upon transformation of that tensor from one system of coordinate axes to another.

A quantity that conforms to this description is density $\underline{B}$ of magnetic flux, with components (B_x, B_y, B_z), for which an explicit correspondence is

$$\text{axial vector } \underline{B} = \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} \quad \text{-->} \quad \mathbf{B} = \begin{bmatrix} 0 & -B_z & B_y \\ B_z & 0 & -B_x \\ -B_y & B_x & 0 \end{bmatrix} \text{ skew symmetric tensor}$$

or its transpose, which is equivalent to the preceding general case with particular values of components $T_{i,j}$. Although this tensor of second rank for density of magnetic flux is evidently represented by a matrix of third order, one can represent a tensor, also of second rank and antisymmetric, for the electromagnetic field -- hence including both electric and magnetic components -- with a matrix of fourth order,

$$\begin{bmatrix} 0 & \dfrac{E_x}{c} & \dfrac{E_y}{c} & \dfrac{E_z}{c} \\[2mm] -\dfrac{E_x}{c} & 0 & -B_z & B_y \\[2mm] -\dfrac{E_y}{c} & B_z & 0 & -B_x \\[2mm] -\dfrac{E_z}{c} & -B_y & B_x & 0 \end{bmatrix}$$

or its transpose, in which c denotes the speed of electromagnetic radiation in free space. A tensor representing a real physical property need thus not necessarily be represented by only a vector in three dimensions or a matrix of third order.

A vectorial or cross product of two polar vectors is representable as a scalar product of an antisymmetric tensor of rank two and a polar vector; the result of such a cross product is an axial vector. A distinction between polar and axial vector is important in relation to properties of transformation of these vectors; for instance, a polar vector changes sign upon inversion through the origin of coordinates in a cartesian system as the sign of each component is changed, whereas an axial vector remains unchanged upon such inversion, according to the property of the skew symmetric tensor shown above. The cross product of two polar vectors is an axial vector, the cross product of an axial vector and a polar vector is a polar vector, and the cross product of two axial vectors is an axial vector. The scalar product of a polar vector and an axial vector, which changes sign upon the same inversion, is called a *pseudo scalar*; a triple scalar product of polar vectors is an example of such a pseudo scalar. In contrast, the scalar product of two axial vectors is a scalar that retains its sign under inversion, just as a scalar product of two polar vectors is a scalar quantity. Axial vectors, in either vectorial or tensorial form, hence serve to represent quantities associated with rotation, for which purpose an axial vector has a screw sense with its magnitude and its direction according to the axis of rotation.

For situations in which the interest lies in the transformation of a tensor under a rotation, one works preferably with spherical rather than cartesian tensors. We denote a spherical tensor with two indices, but here one subscript and one superscript, as T_m^{l} in which index m adopts $2\,l+1$ values from $-l, -l+1, \ldots , +l-1, +l$. The components of a spherical tensor are simply the cartesian components in particular combinations. For instance, for a tensor of first rank, or vector, with cartesian components T_x, T_y, T_z, the components of the spherical tensor are conventionally

$$T_1^{l} = -\frac{1}{\sqrt{2}}(T_x + i\,T_y)$$

$$T_0^{l} = T_z$$

$$T_{-1}^{l} = \frac{1}{\sqrt{2}}(T_x - i\,T_y)$$

in which direction z is taken as the special axis of rotation and $i = \sqrt{-1}$. For a cartesian tensor of

second rank, the irreducible components of the spherical tensor have $l = 0, 1, 2$, of which each has $2\,l + 1$ components, thus accounting for the nine components of the cartesian tensor. One decomposes an arbitrary cartesian tensor T of second rank into these nine irreducible components of the corresponding spherical tensor:

$$T_0^{\;0} = -\frac{1}{\sqrt{3}}\,(T_{x,x} + T_{y,y} + T_{z,z})$$

$$T_0^{\;1} = \frac{i}{\sqrt{2}}\,(T_{x,y} - T_{y,x})$$

$$T_1^{\;1} = \frac{T_{z,x} - T_{x,z} + i\,(T_{x,y} - T_{y,x})}{2}$$

$$T_{-1}^{\;1} = \frac{T_{z,x} - T_{x,z} - i\,(T_{x,y} - T_{y,x})}{2}$$

$$T_0^{\;2} = \frac{2\,T_{z,z} - T_{x,x} - T_{y,y}}{\sqrt{6}}$$

$$T_1^{\;2} = \frac{-(T_{z,x} + T_{x,z} + i\,(T_{y,z} + T_{z,y}))}{2}$$

$$T_{-1}^{\;2} = \frac{T_{z,x} + T_{x,z} - i\,(T_{y,z} + T_{z,y})}{2}$$

$$T_2^{\;2} = \frac{T_{x,x} - T_{y,y} + i\,(T_{x,y} + T_{y,x})}{2}$$

$$T_{-2}^{\;2} = \frac{T_{x,x} - T_{y,y} - i\,(T_{x,y} + T_{y,x})}{2}$$

A pseudo-tensor is an quantity of which formulae for transformation involve square roots of matrix determinants rather than the *common rules*; mathematical applications of these quantities exist in differential geometry, which is the study of geometry according to methods of calculus -- for instance to evaluate the area of a surface, and in connection with Fourier integral operators, and in physics in the properties of deformed solid objects or in general relativity.

Maple provides a specific package tensor to manipulate general tensorial quantities, designed to facilitate calculations in general relativity; for many chemical applications that involve typically tensors of rank two -- or less, we can instead, for simplicity, apply matrix, or vectorial, operations; another package physics provides further facilities to work with tensors. An arbitrary square matrix of order three does not necessarily represent a tensor of rank two: to be such an object, a matrix must have elements that are defined in terms of a relation with an underlying spatial geometry. We employ *Maple* to construct a **Matrix** according to package **LinearAlgebra** to represent properties of a mathematical object that transforms as a *tensor* of rank two in chemical systems, whereas an **Array** according to package **LinearAlgebra**, or an **array** according to package **linalg**, might serve to represent a tensor of greater rank. Many applications of tensors of rank two that pertain to systems of orthogonal coordinates can be implemented with commands involving operations on matrices.

spreadsheet

As an invention that originates in this era of an electronic digital computer, a *spreadsheet* lacks a traditional arithmetical or mathematical counterpart, apart from resemblance to a static ledger or similar document for accounting. On examining how a spreadsheet functions, one observes that many operations that it facilitates resemble those on a list, vector or a matrix, but one executes these operations without explicit reference to formal constructs of linear algebra. A spreadsheet represents a convenient, if implicit, means to work with *lists*, *arrays*, *vectors* and *matrices* that might contain not only *numeric* quantities but also *algebraic* expressions, or their combinations, consistent with general symbolic capabilities of *Maple*.

Linear algebra is thus a branch of mathematics concerned with linear equations, matrices, determinants, vectors, vectorial spaces, eigenvectors, tensors, vector fields and related topics. All these abstract mathematical objects have important roles in a chemical context: for instance, in relation to absorption spectra of a liquid solution that becomes a problem central to analytical chemistry for quantitative characterization of that solution, matrix methods are efficient. Building quantitative models in physical chemistry, such as a treatment of nuclear vibrations in relation to a discussion of infrared and Raman spectra or the motions of electrons in a molecule, likewise benefits from concise and powerful notation in terms of matrix, vectorial and tensorial quantities. An application of vectorial properties arises, for instance, in associating an electric dipolar moment with each conventional bond, or link between adjacent atoms, in a molecule; with vector summation of such postulated bond dipoles, one estimates a total molecular electric dipolar moment.

In discussing these topics we generally employ *Maple*'s packages **LinearAlgebra** and **VectorCalculus** that have features additional to obsolescent package **linalg**; extensions and additional features made available in **LinearAlgebra** affect no underlying principle but facilitate an application of matrices, vectors and tensors, especially for fully numerical operations through embedded procedures (supplied by Numerical Algorithms Group) and invoked within *Maple* transparently to a user. As after *Maple* release 8, package capabilities of **linalg** are essentially duplicated in packages **LinearAlgebra** and **VectorCalculus**, we employ only the latter packages; for work with symbolic matrices and vectors, package **linalg** might provide advantages in some applications, because by default elements of a matrix or components of a vector are symbolic with package **linalg** but zero with package **LinearAlgebra**, but general use of obsolescent package **linalg** is otherwise deprecated. Structures in package **linalg** are based on a **table**, whereas structures in these subsequent packages are based on **rtable** or rectangular table. For calculations in general relativity, *Maple* package **tensor** still employs operations from package **linalg**. These topics and their applications we treat in this chapter. references C. A. Hollingsworth, *Vectors, Matrices and Group Theory for Scientists and Engineers*, McGraw-Hill, New York USA, 1967; D. D. Fitts, *Vector Analysis in Chemistry*, McGraw-Hill, New York USA

>

summary of chapter 6

In this chapter our concern is to establish key constructs of linear algebra -- matrix, determinant, vector, dyad, dyadic and tensor -- and the calculus of vectors. Of special interest are

ways to combine such objects of the same and other kinds, and how their associated properties relate to chemical applications. In subsequent chapters we discern that the eigenvalue problem, which requires a detailed understanding of interlinked properties of vectors, determinants and matrices, forms a foundation to model electronic and vibrational processes of great importance in physical chemistry. A concept of dual vectors, used to establish a concept of reciprocal space, underpins understanding of structure of a crystalline chemical compound. A spreadsheet, which, apart from being a display of tabular matter of convenient format, is an array of special form, renders great service in many chemical applications through their structure and properties; we apply an unique feature of a *Maple* spreadsheet -- a capability of symbolic operations -- in subsequent chapters.

chapter 7 Differential and integral equations

7.0 overview and principles

Science is a differential equation.
Alan Turing

Like differential and integral calculus, the study of the solution of differential equations began with Newton and Liebniz, but major development was due to Jakob and Johann Bernoulli, Leonhard Euler, Joseph-Louis Lagrange and Pierre-Simon de Laplace during the eighteenth century.

A differential equation is an equation that involves one or more independent variables, their known and unknown functions or formulae and their derivatives of finite number. An ordinary-differential equation is a differential equation in a single independent variable, such as this one with dependent variable y denoting distance as a function of time t as its only independent variable,

$$\frac{d^2}{dt^2} y(t) = y(t) \cos(t)$$

whereas a partial-differential equation involves multiple independent variables and partial derivatives of an unknown function or formula with respect to those variables, such as this one with a derivative of unknown formula $f(x, y)$ of spatial coordinates in two dimensions with respect to two spatial variables x and y,

$$\frac{\partial^2}{\partial y\, \partial x} f(x, y) = c\, f(x, y)$$

To solve a differential equation, we must find a formula or function for which the equation is true; we must thus manipulate the differential equation so as to eliminate all derivatives, leaving a relation between independent and dependent variables. Differential equations have their origin in geometric and physical problems, and, just like other equations, they occur also in systems -- multiple equations involving the same independent and dependent variables and their derivatives; systems of ordinary-differential equations have thus only a single independent variable and systems of partial-differential equations have multiple independent variables.

An integral equation likewise involves a solution for an unknown formula that occurs within an integrand, such as x(t) in this definite integral,

$$\int_0^a x(t)^2\, dt = t$$

There exist also integro-differential equations that contain both derivatives and integrals. All such equations are each a special case of an operator *equation* because it contains a differential operator or an integration. A differential equation can be invariably reformulated as an integral equation, but the converse is not necessarily true. Finding an exact algebraic solution to a differential or integral equation occurring in a chemical context is likely to be an exception rather than a rule, but a symbolic computation with *Maple* maximizes a chance of success with exact methods, and facilitates an application of approximate methods, apart from purely numeric approaches for which *Maple* also caters.

differential equation

   Poincare described the study of differential equations to have both qualitative and quantitative aspects; we here consider both, and naturally emphasize the latter in subsequent sections involving direct calculations. We customarily consider a derivative in a context of calculus to be a differential operator $\dfrac{\partial}{\partial x}$ operating on a variable such as y that depends on variable x, rather than as a ratio of differential quantities, such as $\dfrac{dy}{dx}$, also introduced in chapter 3. According to an extended development of calculus, we separate numerator and denominator, to become differential quantities such as dy and dx, to opposite sides of an equality sign for instance, and handle each quantity separately; for this purpose we describe applications in chapters 3 and 5; in the same way, we employ properties of a differential to guide us to solve differential and integral equations that we treat in this chapter. The order of a differential equation is that of the highest derivative appearing in that equation; the degree of the differential equation equals the greatest power of the function to be evaluated or its derivatives.

To solve this differential equation, one looks at it until the solution occurs to one.

adapted from George Polya

   With advanced mathematical software such as *Maple*, no longer is Polya's dictum valid: if no algebraic solution be found by *Maple* -- which is the likely case for a general differential equation because no such solution exist, a numerical solution is alternatively practicable, and this software caters for both possibilities.

ordinary-differential equation

   For a general ordinary-differential equation of first order such as $\dfrac{d}{dx} y(x) = f(x, y)$, or analogously for other order, a solution might exist providing that $f(x, y)$ is continuous and has a single value over a region of points (x, y), and that $\dfrac{\partial}{\partial x} f(x, y)$ exists and is continuous at all points in that region. The solution or integral of a differential equation is defined as a set of all formulae of which the derivatives satisfy identically that differential equation. The general solution of a differential equation contains arbitrary constants, equivalent to constants of integration for an indefinite integral, that number the same as the order of that differential equation, but such a

general solution might not contain all possible solutions. A particular solution contains no arbitrary constant, like the result of evaluating a definite integral, and might result from the application of initial values or boundary conditions of sufficient number, or under other conditions.

The general solution of a linear ordinary-differential equation of order n having a form $L_n(y) = g(x)$ is hence a sum of the homogeneous solution or complementary function $y_h(x)$ and any particular solution $y_p(x)$. A solution of a differential equation is called singular if it be unobtainable from the general solution according to the choice of a particular parameter. The graphical representation of a differential equation of order n comprises curves in a family with n parameters; each family of curves has, conversely, its differential equations. A particular solution corresponds to one curve among the family of curves, which is described as the curve of a solution or an integral curve. A differential equation of first order determines, at each point (x, y) of the domain of definition of the function, the direction $\dfrac{d}{dx} y(x) = \tan(\theta)$ of the curve through this point and included in a curve of the family of the general solution of the differential equation $f\left(x, y, \dfrac{d}{dx} y(x)\right) = 0$ or $\dfrac{d}{dx} y(x) = f(x, y)$. Three values $(x, y, \dfrac{d}{dx} y(x))$ in a set define one line element of a curve of a solution set; all line elements produce a direction field according to the coordinates in a cartesian system. The family of curves of a solution set includes all curves of which the directions at each point correspond to the direction field. Lines connecting all points with the same direction of the line elements are called isoclines, for which $\dfrac{d}{dx} y(x) = constant$; for the nullcline the slope is zero. A differential equation of second order determines both the direction and the curvature of arc elements at each point of the domain of definition. A trajectory is a curve that intersects each curve of a family exactly once; if the intersection occur at angle $\dfrac{\pi}{2}$ rad, the trajectory is orthogonal.

A differential equation contains a derivative of some order, for instance appearing in a form $\dfrac{dy}{dx}$ or $\dfrac{\partial}{\partial x} y$ or even y' for a derivative of first order of dependent variable y with respect to independent variable x; a derivative with respect to time, such as a velocity or speed $\dfrac{dx}{dt} = \dfrac{\partial}{\partial t} x$, might alternatively be expressed as $\dot{x}$ as in Newton's own notation, with hence a point directly above the symbol (which is difficult to represent here). Derivatives of greater order might accordingly be expressed as $\dfrac{\partial^2}{\partial x^2} y = y''$, or for acceleration as $\ddot{x}$ with two points directly above the symbol to signify a second derivative with respect to time, or $\dfrac{\partial^3}{\partial x^3} y = y'''$ for a third derivative, $\dfrac{\partial^4}{\partial x^4} y = y^{(4)}$ for a fourth derivative, et cetera, but, for recognition in a *Maple* command for differential equations,

the dependence on a particular independent variable must be explicit, as in $\dfrac{d}{dx}\,y(x)$, and analogously for other derivatives. A standard form for a differential equation contains typically the derivatives of the dependent variable for an ordinary-differential equation, or dependent variables in the case of a partial-differential equation, appearing on the left side of an equality in descending order of derivative from left to right, with any other terms containing the dependent variable and other terms of the differential equation containing only the independent variable and constants on the right side of that equality, such as in

$$\left(\frac{\partial^2}{\partial x^2}\,y\right) - (x^2 + 5)\left(\frac{\partial}{\partial x}\,y\right) + (x^4 + \sin(4\,x))\,y(x) = e^{(-7\,x)}\cos(2\,x)$$

The corresponding homogeneous differential equation, for which the left side containing all derivatives equals zero, might be described as *normal* over some finite interval if the coefficient of the derivative of greatest order, here $\dfrac{\partial^2}{\partial x^2}\,y$, is never zero over that interval. The formula on the right side of the equality might be described as a driving term or *external source*. At any point of a curve of that function, the second derivative of a function is a measure of that function's concavity, which is related to -- but not the same as -- the curvature, the first derivative is a measure of the slope, and the zero derivative -- the function itself -- is a measure of its magnitude; the solution of that homogeneous equation, called an *homogeneous solution*, is thus a function of which a sum of the concavity multiplied by its coefficient in the differential equation plus the slope multiplied by its coefficient plus the magnitude multiplied by its coefficient must be zero. On any interval on which the homogeneous ordinary-differential equation of order *n* is *normal*, as defined above, the solution as a vector space has *n* dimensions, so is hence finite; there exist *n* linearly independent solution vectors $y_1(x), y_2(x), ..., y_n(x)$. For these *n* solutions over interval *I*, if the wronskian, explained below, differ from zero everywhere within that interval, these *n* solutions are linearly independent and form a basis of that space.

For a differential equation of first order, a typical notation might alternatively be $\dfrac{\partial}{\partial x}\,y = f(x, y)$ or $y' = f(x, y)$. For only a differential equation of first order, the descriptor homogeneous implies that, for each real number *a*, $f(a\,x, a\,y) = a\,f(x, y)$, or even just $f(a\,x, a\,y) = f(x, y)$, or alternatively $\dfrac{\partial}{\partial x}\,y = f\!\left(\dfrac{y}{x}\right)$.

Consider these steps, in which a derivative results from differentiation of function $y = F(x)$ of one independent variable *x*, such that $f(x) = F'(x)$. We begin with a differential equation of first order containing a derivative to the left of an equality operator;

$$\frac{dy}{dx} = f(x)$$

we separate the differential quantities within that quotient, one to each side of that operator,

$$dy = f(x)\ dx$$

and integrate both sides,

$$\int dy = \int f(x)\,dx$$

$$y = \int f(x)\,dx$$

neglecting the constant of integration required for an indefinite integral. A result of these four steps is the production of a relation between variable y on the left, dependent on x, and an indefinite integral on the right that we might in principle evaluate fully to yield $F(x)$, according to methods described in chapter 4. What we achieve formally here is to demonstrate that a solution of a differential equation of first order, present in the first step, appears in the fourth step to involve an indefinite integration of $f(x)$ -- the derivative function of $F(x)$. As we note in chapter 4, the solution of an indefinite integral, as in this fourth step, involves inclusion of a constant of integration: solution of the original differential equation, present in the first step, likewise necessitates inclusion of a constant, to be evaluated. That solution on an interval is a function $y = F(x)$ that satisfies identically the differential equation for all x on that interval. In a general solution of a single differential equation the arbitrary constants number the same as the order of that differential equation, and subsidiary conditions known as initial conditions or boundary conditions are applied to eliminate some or all such constants for a particular chemical or physical problem; a particular solution might contain no such arbitrary constant. A particular solution of a differential equation is any one solution, or the solution of the non-homogeneous differential equation that contains, beside derivatives and expressions containing the dependent variable, also terms with only the independent variable; the general solution of a differential equation comprises all solutions in a set. The resolution of a general solution into the homogeneous and the particular solutions is a characteristic of a linear equation, both algebraic and differential. A differential equation might have solutions of uncountable number, or only one solution, or no solution; for

instance, differential equation $2\left(\dfrac{\partial}{\partial x}y\right)^4 + 3\,y^2 = -5$ must have no real solution for real function

$y(x)$ because, under such a condtion, the left side must evaluate to a positive quantity whereas the right side is a purely negative quantity, independent of the magnitudes of the numerical coefficients that appear in this equation. To test whether a formula $y = f(x)$ for a dependent variable be a solution of a differential equation, or to confirm whether a solution proffered by software or a table in a book is correct, that formula is simply inserted into the pertinent differential equation and the result simplified. A differential equation given without initial values of dependent variables or without boundary conditions has in general multiple solutions, related to the presence of an arbitrary parameter in those solutions equivalent to an integration constant, but a proper substitution of that proposed solution into the differential equation and evaluation of the resulting expressions must eliminate that parameter or constant.

In a chemical context one has generally information to evaluate this constant: for example, if we record a concentration of a compound or species over time in a sequence of intervals, we typically know its initial concentration. One generally refers to such information as an initial value for the solution. If subsidiary conditions be provided at multiple values of an independent variable, these conditions constitute boundary conditions.

A simple approach to obtain an exact algebraic solution of a differential equation of first order

of form described as standard,

$$\frac{\partial}{\partial x} y = h(x, y)$$

yields a solution through direct integration only if function $h(x, y)$ has a <u>factorisable</u> form $f(x)\, g(y)$ that enables <u>separation of variables</u>, in which case a <u>substitution</u> $h(x, y) = f(x)\, g(y)$ and subsequent rearrangement yield a form

$$\int \frac{1}{g(y)}\, dy = \int f(x)\, dx$$

For instance, in this simple case,

$$\frac{d}{dx} y(x) = x\, e^{(-y)}$$

rearrangement yields

$$\int e^y\, dy = \int x\, dx$$

which has an obvious solution. In all other cases, in which variables are inseparable in this way, such an *ordinary-differential equation* might in general be solved through another method. We express such an ordinary-differential equation of first order also even more generally as

$$G\left(y, x, \frac{\partial}{\partial x} y\right) = constant,$$ and analogously with further derivatives for an ordinary-differential equation of greater order.

Similarly, for a more general differential equation of first order in one independent variable,

$$g_1(x)\left(\frac{d}{dx} y(x)\right) + g_2(x)\, y(x) = f(x)$$

the corresponding homogeneous equation,

$$g_1(x)\left(\frac{d}{dx} y(x)\right) + g_2(x)\, y(x) = 0$$

has a general solution

$$y_1(x) = c\, e^{\left(\int -\frac{g_2(x)}{g_1(x)}\, dx\right)}$$

as a single basis vector; the solution of the homogeneous equation contains the arbitrary constant. To solve the non-homogeneous equation we assume the solution to be a non-linear product of this general solution with unknown variable function $u(x)$,

$$y(x) = u(x)\, y_1(x)$$

such that we force a condition of linear independence between $y(x)$ and $y_1(x)$. Substitution of this assumed solution into the original differential equation yields

$$g_1(x)\left(\frac{d}{dx} u(x)\right) y_1(x) + g_1(x)\, u(x)\left(\frac{d}{dx} y_1(x)\right) + g_2(x)\, u(x)\, y_1(x) = f(x).$$

Because $y_1(x)$ is a solution of the homogeneous equation, the latter two terms on the left side of

the equality vanish, leaving an equation first order in u(x),

$$g_1(x)\left(\frac{d}{dx}u(x)\right)y_1(x) = f(x)$$

that has as solution

$$u(x) = \int \frac{f(x)}{g_1(x)\,y_1(x)}\,dx$$

A [particular solution] of the non-homogeneous differential equation is thus

$$y_p(x) = y_1(x)\int \frac{f(x)}{g_1(x)\,y_1(x)}\,dx$$

Defining a [Green's function] of first order as

$$G_1(x, s) = \frac{y_1(x)}{g_1(x)\,y_1(x)}$$

we express that particular solution as

$$y_p(x) = \int G_1(x, s)\,f(x)\,dx,$$

which possesses a general utility in that an evaluation of this Green's function for a particular differential equation provides a solution that accommodates a *driving* or *source function* f(x) of any type. In evaluating the latter integral, one should perform integration first with respect to *s* and then substitute *x* for *s* in the solution of that integral.

Differential equations, each of first order, with multiple [dependent variables] for a single independent variable such as time, comprise a [system] of simultaneous equations, having a form

$$\frac{\partial}{\partial t}x1 = f_1(t, x1, x2, x3, ..., xn)$$

$$\frac{\partial}{\partial t}x2 = f_2(t, x1, x2, x3, ..., xn)$$

$$\cdots$$

$$\frac{\partial}{\partial t}xn = f_n(t, x1, x2, x3, ..., xn)$$

in which the number of equations equals the number of dependent variables *x1 ... xn*.

Another differential equation, of first order and first degree and of type described as being in [differential form],

$$M(x, y)\,dx + N(x, y)\,dy = 0,$$

as opposed to a standard form of the same content,

$$\frac{\partial}{\partial x}y = -\frac{M(x, y)}{N(x, y)}$$

is directly integrable if the left side be an [exact differential], or if an [integrating factor] be deducible, as discussed in section 5.113 with examples. For a differential equation of first order, an integrating factor $\lambda(x, y)$ such that

$$\lambda(x, y)\,(M(x, y)\,dx + N(x, y)\,dy)$$

becomes an exact differential, even though $M(x, y)\,dx + N(x, y)\,dy$ is not, invariably exists, but a general procedure to evaluate that factor is unknown. For these three particular cases, an integrating factor is readily found. If

$$\frac{1}{N(x, y)}\left(\left(\frac{\partial}{\partial y}(\lambda(x, y)\,M(x, y))\right) - \left(\frac{\partial}{\partial x}(\lambda(x, y)\,N(x, y))\right)\right) = g(x)$$

in which $g(x)$ is a function of only x, the integrating factor is

$$\lambda(x, y) = \mathbf{e}^{\left(\int g(x)\,dx\right)}$$

or if

$$\frac{1}{M(x, y)}\left(\left(\frac{\partial}{\partial y}(\lambda(x, y)\,M(x, y))\right) - \left(\frac{\partial}{\partial x}(\lambda(x, y)\,N(x, y))\right)\right) = h(x)$$

in which $h(y)$ is a function of only y, the integrating factor is

$$\lambda(x, y) = \mathbf{e}^{\left(\int h(y)\,dy\right)}$$

If $M(x, y) = y\,f(x\,y)$ and $N(x, y) = x\,g(x\,y)$, the integrating factor is

$$\lambda(x, y) = \frac{1}{x\,M(x, y) - y\,N(x, y)}\,.$$

If independent variables number more than one and with partial derivatives of dependent variable with respect to them, a *partial-differential equation* might be separable into ordinary-differential equations each involving only a single independent variable as an approach to a solution.

If a differential equation be expressible in a form $f_1(x)\,g_1(y)\,dy + f_2(x)\,g_2(y)\,dx = 0$, an integrating factor $\dfrac{1}{f_2(x)\,g_2(y)}$ reduces the preceding equation to a form $\dfrac{f_1(x)}{f_2(x)}\,dx + \dfrac{g_1(y)}{g_2(y)}\,dy = 0$, from which a primitive is obtained on integrating each term separately.

Differential equations either *ordinary-* or *partial-* can be classified further as to order, whether they are linear, homogeneous, exact or autonomous, and whether they have constant coefficients et cetera; as an example of a linear homogeneous ordinary-differential equation with constant coefficients,

$$\left(\frac{d^2}{dx^2}y(x)\right) + c\left(\frac{d}{dx}y(x)\right) - y(x) = 0,$$

and we provide other examples with their solutions in ensuing sections. The descriptor linear in these cases refers to the dependent variable, y in the above case; a product of that dependent variable or its derivatives with the independent variable or formulae thereof does not affect that

linearity, whereas any term in a differential equation such as $y\left(\dfrac{d}{dx}y(x)\right)$ or $y(x)^2$ or $\sqrt{\dfrac{d}{dx}y(x)}$

would make a differential equation non-linear. In a linear differential equation, there thus appears neither a product of derivatives nor a product of a derivative with the dependent variable, nor does a derivative appear as an argument of a transcendental function. For a linear differential equation

of first order such as $\left(\dfrac{d}{dx}\,\text{y}(x)\right)+\dfrac{x}{\text{y}(x)}=0$, with an initial condition $y\big|_{x=x_0}=y_0$, we rewrite the

equation as $x\,dx+y\,dy=0$, which upon direct integration yields $x^2+y^2=a$, a constant such that

$x_0^{\,2}+y_0^{\,2}=\sqrt{a}$. This solution corresponds to concentric circles of radius $\sqrt{a}$, which is variable,

and is alternatively expressed explicitly as $y=\pm\,(a-x^2)^{\left(\frac{1}{2}\right)}$, so that y is doubly valued and also a
function of constant a of integration. A linear differential equation of second order has two
distinct or linearly independent solutions, not proportional to one another. For instance, for

$\left(\dfrac{d^2}{dt^2}\,\text{y}(t)\right)+\omega^2\,\text{y}(t)=0$, the complete or general solution is $\text{y}(t)=\alpha_1\,y_1(t)+\alpha_2\,y_2(t)$ in which α_1

and α_2 are arbitrary constant multiplicands, not zero, and subject to evaluation according to an
initial value or boundary condition; the values of these parameters have no effect on the period

$\dfrac{2\,\pi}{\omega}$ of the oscillation of the mechanical system to which that differential equation pertains. If a

term be added to this equation, to yield $\left(\dfrac{d^2}{dt^2}\,\text{y}(t)\right)+\omega^2\,\text{y}(t)+\beta\,y^2=0$ for instance, the general

form of solution of that non-linear equation can not take the form of the general solution above
because a non-linear differential equation can not have linearly independent solutions; the period
corresponding to this differential equation is a function of initial conditions, such that that period
decreases with increasing initial amplitude y_0.

 Dimensional analysis is useful in many applications in chemistry and physics, including
differential equations. Regarding variable x as having <u>dimension</u> L^1 and y dimension L^k for some

unit L that is immaterial for the purpose, we take the dimension of $\dfrac{\partial}{\partial x}y$ to have a dimension $L^{(k-1)}$

; hence $\dfrac{y}{x^k}$ and $\dfrac{\dfrac{\partial}{\partial x}y}{x^{(k-1)}}$ are both dimensionless, or pure numbers. A term $x^m\,y^n$ has dimension

$L^{(m+kn)}$, which corresponds to a <u>weight</u> $m+k\,n$ of that term; formulae such as $\mathbf{e}^u$ or $\sin(u)$ have
zero weight when the weight of u is zero, but a weight is not assignable otherwise. A differential
equation is described as *isobaric* when all terms therein have the same weight for appropriately
chosen weights of variables. For instance, for an equation $2\,x\,y\,dx+(2\,x^2-3\,y)\,dy=0$, the
weights of terms in an expanded form are $2+k$, $2+k$ and $2\,k$; when $k=2$ all terms have weight 4.
Dividing the original equation by $2\,x^2\,y+2\,y\,(2\,x^2-3\,y)=6\,y\,(x^2-y)$ thus converts it into an

<u>exact</u> equation $\dfrac{2\,x\,dx}{x^2-y}+\dfrac{(2\,x^2-3\,y)\,dy}{y\,(x^2-y)}=0$, which we integrate as $\displaystyle\int_0^x\dfrac{2\,x}{x^2-y}\,dx+3\int\dfrac{1}{y}\,dy=c$,

with constant c. A differential equation of first order and of form $f\left(x, y, \dfrac{\partial}{\partial x} y\right) = 0$ might be simplified according to a point transformation of variables on introducing new variables $u = g(x, y)$ and $v = h(x, y)$ such that a point (x, y) in plane xy becomes transformed to a point (u, v) in plane uv, assuming that a <u>jacobian</u> that is a determinant of the derivatives, $\begin{vmatrix} \dfrac{\partial}{\partial x} u & \dfrac{\partial}{\partial y} u \\ \dfrac{\partial}{\partial x} v & \dfrac{\partial}{\partial y} v \end{vmatrix} \neq 0.$

In either case of an isobaric equation or a point transformation, *Maple* undertakes the corresponding operations to solve the equation if practicable.

   If there be no function of independent variable apart from the derivatives thereof, as above, a linear differential equation is <u>homogeneous</u>, otherwise <u>non-homogeneous</u>. A particular case of such a linear differential equation arises when the coefficients of derivatives or dependent variable are only constants -- hence a homogeneous linear differential equation with constant coefficients.

A differential operator D implies $\dfrac{\partial}{\partial x}$; we apply this operator, that Heaviside originated, in the solution of differential equations as follows. With each linear differential equation that has constant coefficients, of form $\phi(D)(y) = f(x)$, there is associated an <u>auxiliary equation</u> that is simpler to solve because it contains no derivatives; this auxiliary equation is formed on replacing, in the homogeneous equation $\phi(D)(y) = 0$, a first derivative D by λ, a second derivative D^2 by λ^2, and so forth, to form a polynomial in λ, of which the roots obtained on solving that polynomial set equal to zero become coefficients λ_j of x in terms of form $\mathbf{e}^{(\lambda_j x)}$, and a linear combination of these terms constitutes the general solution. For an auxiliary equation $\phi(\lambda) = 0$ associated with a homogeneous linear differential equation $\phi(D)(y) = 0$ containing constant coefficients, if $\phi(\lambda) = 0$ have a <u>real</u> <u>root</u> λ_j of <u>multiplicity</u> k, a solution of the differential equation is

$(c_0 + c_1 x + ... + c_{k-1} x^{(k-1)}) \, \mathbf{e}^{(\lambda_j x)}$, or if $\phi(\lambda) = 0$ have <u>complex conjugate</u> roots in a couple $\lambda_j = a \pm b\,i$ each of multiplicity k, a solution of the differential equation is obtained from expanding into trigonometric form the complex exponent as

$\quad (c_0 + c_1 x + ... + c_{k-1} x^{(k-1)}) \cos(b\,x) \, \mathbf{e}^{(\lambda_j x)} + (d_0 + d_1 x + ... + d_{k-1} x^{(k-1)}) \sin(b\,x) \, \mathbf{e}^{(\lambda_j x)}$;

superposing all such solutions in both cases yields a general solution of the differential equation, as elaborated below.

   In a context of finding a particular solution of a differential equation with the use of operator D when $f(x)$ has a polynomial or exponential or sine or cosine form and sums or products of such terms, we convert $\phi(D)(y) = f(x)$ into $y(x) = \dfrac{1}{\phi(D)} f(x)$; as $\phi(D)$ that is a linear operator might contain a sum of contributions of form D, D^2, ..., we regard the application of $\dfrac{1}{\phi(D)}$ through an <u>expansion</u> of the denominator into the numerator and the subsequent application of the

contributions in that sum to f(x). If $\phi(D)$ be simply D, we regard $\dfrac{1}{D}$ as implying an integration

$\int$ f(x) dx, and $\dfrac{1}{D^2}$ as implying two integrations in sequence $\int\int$ f(x) dx dx, and so forth: for instance, if $\phi(D) = D^2\,(1 + 3\,D - D^2)$, after transfer to the right side of the equality factor D^2 remains in the denominator, and $1 + 3\,D - D^2$ becomes expanded in the numerator there to $1 - 3\,D + 10\,D^2 - 33\,x^3 + ...$; first the latter operations are applied to f(x) and then the other factor D^2 is taken into account with integration twice of the result of the preceding operations.

An order of a differential equation signifies the greatest order of derivative that appears therein, after the equation has been rationalized; a degree of a differential equation is equal to the greatest exponent or power of a derivative of greatest order: for instance, a differential equation of form

$$\left(\frac{d\,y(x)}{dx} \right)^2 = y(x)$$

has first order but second degree. A differential equation of form

$$\left(\frac{\partial^3}{\partial x^3} y \right) + x \sqrt{\frac{\partial}{\partial x} y} + x^2\,y = 0$$

has order third and degree second, because after rationalization it contains a term $\left(\dfrac{\partial^3}{\partial x^3} y \right)^2$. The solutions of a differential equation of order n are represented by curves filling a space of $n + 1$ dimensions; to describe the shape of such curves is the qualitative nature of the problem.   In a *linear* equation, there are no squares or products involving an independent variable and its derivatives; an unknown function y(x) and its derivatives $\dfrac{d^n\,y(x)}{dx^n}$ appear with coefficients possibly containing a function of independent variable x but with no function of dependent variable y or its product except with a constant: such an equation has thus first degree in y(x); a function of independent variable x might include terms in x^2 or x to other powers, but with y to only the first power. The differential equations in a system are linear if each equation is linear in dependent variables. A linear equation is *homogeneous* if one can express it in a form such that unknown function y(x) and all its derivatives appear on the left side of an equality sign but only zero appears on the right side; the equation is otherwise *inhomogeneous*. For an homogeneous differential equation of order n greater than unity, or for a system of n linear homogeneous differential equations, n solutions in linear combination -- sum or difference with coefficients to be evaluated -- also constitute a solution. For an non-homogeneous differential equation, the general solution is a sum of the general solution of the corresponding homogeneous equation -- the complementary function or homogeneous solution -- and any solution of the non-homogeneous equation -- a particular integral. Explicitly, for a non-homogeneous linear differential equation $\phi(D)(y) = $ f(x) containing constant coefficients, the associated homogeneous equation $\phi(D)(y) = 0$ is called a homogeneous or complementary or reduced equation; a general solution

$y(x) = y_h(x) + y_p(x)$ of that non-homogeneous equation contains $y_h(x)$ that is a general solution of the associated homogeneous equation and $y_p(x)$ that is any particular solution of the entire differential equation.

An ordinary-differential equation of second order, such as $\dfrac{\partial^2}{\partial t^2} x = \dfrac{f}{m}$, is invariably expressible as a system of two differential equations of first order, through a <u>transformation</u>

$$v = \frac{\partial}{\partial t} x, \ \frac{\partial^2}{\partial t^2} x = \frac{\partial}{\partial t} v,$$

so that the two equations become $v = \dfrac{\partial}{\partial t} x$ and $\dfrac{\partial}{\partial t} v = \dfrac{f}{m}$. The order of differential equations in a system is a sum of orders of equations in that system; according to the preceding method of a reduction of the order, the differential equations in an arbitrary system of order n are expressible as n differential equations of first order in a corresponding system. One can in principle hence express all ordinary-differential equations and systems thereof in exactly a form of simultaneous linear differential equations in a system displayed above. By regarding $x1, x2, x3, ..., xn$ as <u>components</u> of a <u>vector</u>, we might consider such a system to represent a single vectorial differential equation. This reduction of order is practicable for any linear differential equation, which becomes thereby equivalent to multiple equations of first order in a matrix system; the linear differential equations in a set, and with initial conditions, are also reducible to a system of equations of first order.

In an <u>autonomous</u> differential equation, such as $\dfrac{\partial}{\partial x} y = f(y)$, independent variable, such as x, appears explicitly only in a derivative. Such autonomous differential equations have the following properties:

- if $u(x)$ be a solution of that equation, $u(x + c)$ with a constant c is also a solution;

- if $u(x)$ be a solution of that equation and if $\left. \left(\dfrac{\partial}{\partial x} u \right) \right|_{x = x_0} = 0$ for some x_0, $u(x)$ is a constant solution;

- a solution of that autonomous differential equation must be either constant or <u>monotonically</u> increasing or decreasing with increasing x, hence possessing no <u>relative</u> <u>maximum</u> or <u>minimum</u> or <u>oscillatory</u> behaviour;

- if a solution remains <u>bounded</u>, it must be <u>asymptotic</u> to a constant solution.

In chemical or physical terms, a solution represents either a growth or decay, or a constant that corresponds to an equilibrium state. Such an <u>equilibrium</u> is either <u>stable</u> or <u>unstable</u>: if a small <u>displacement</u> from a <u>state</u> of equilibrium cause a return to that equilibrium state, the equilibrium is stable, whereas a displacement, no matter how slight, from a state of unstable equilibrium causes a large departure from that state and an approach to a state of stable equilibrium, perhaps leading to oscillatory <u>motion</u> about the latter state.

A solution of an ordinary-differential equation of order n requires an equivalent of n

integrations, each of which generates a constant of integration; an evaluation of these constants requires further information about the system, such as initial conditions or boundary values of variables. A linear differential equation of second order hence requires two integrations, each generating a constant; there are two linearly independent solutions of which a general solution is a linear combination thereof and that form a basis of solutions: if $y_1(x)$ and $y_2(x)$ are the independent solutions of a differential equation of homogeneous linear type, a general solution is $A\,y_1(x) + B\,y_2(x)$ in which A and B are arbitrary parameters or constants.

For this linear differential equation of order n,

$$a_n(x)\left(\frac{\partial^n}{\partial x^n}y\right) + a_{n-1}(x)\left(\frac{\partial^{n-1}}{\partial x^{n-1}}y\right) + \ldots + a_1(x)\left(\frac{\partial}{\partial x}y\right) + a_0(x)\,y = f(x)$$

$f(x)$ and coefficients $a_j(x), j = 0, 1, 2, \ldots, n$, depend only on independent variable x, hence on neither y nor a derivative of y. If $f(x) = 0$, this differential equation is homogeneous, otherwise *non-homogeneous*. If all $a_j(x)$ be constants -- i.e. containing neither x nor y, the differential equation has *constant coefficients*; otherwise it has *variable coefficients*.

For such a differential equation, there are typically initial conditions applicable to its solution. If there be n initial conditions of form such that values of the unknown formula or function $y(x)$ and its derivatives are given at point x_0,

$$y(x)\big|_{x=x_0} = c_0, \left(\frac{d}{dx}y(x)\right)\Big|_{x=x_0} = c_1, \left(\frac{d^2}{dx^2}y(x)\right)\Big|_{x=x_0} = c_2, \ldots, \left(\frac{d^{n-1}}{dx^{n-1}}y(x)\right)\Big|_{x=x_0} = c_{n-1}$$

if $f(x)$ and all coefficients $a_j(x), j = 0, 1, 2, \ldots$, be continuous on some interval containing x_0, and if $a_n(x) \neq 0$ on that interval, the initial-value problem specified by these relations has a single and unique solution defined throughout that interval. When the above conditions hold, we reduce the differential equation by dividing by $a_n(x)$ to obtain

$$\left(\frac{\partial^n}{\partial x^n}y\right) + b_{n-1}(x)\left(\frac{\partial^{n-1}}{\partial x^{n-1}}y\right) + \ldots + b_1(x)\left(\frac{\partial}{\partial x}y\right) + b_0(x)\,y = g(x)$$

in which $\dfrac{a_j(x)}{a_n(x)} = b_j(x), j = 0, 1, 2, \ldots$, and $\dfrac{f(x)}{a_n(x)} = g(x)$. By defining a differential operator $L_n(y)$ such that

$$L_n(y) = \left(\frac{\partial^n}{\partial x^n}y\right) + b_{n-1}(x)\left(\frac{\partial^{n-1}}{\partial x^{n-1}}y\right) + \ldots + b_1(x)\left(\frac{\partial}{\partial x}y\right) + b_0(x)\,y$$

we express the linear differential equation of order n as $L_n(y(x)) = g(x)$ to exhibit the explicit dependent variable, with its corresponding linear homogeneous differential equation $L_n(y(x)) = 0$ that has invariably n linearly independent solutions. Representing these solutions in a set as $\{y_1(x), y_2(x), \ldots, y_n(x)\}$, the general solution of this homogeneous equation $L_n(y) = 0$ is

$$y_h(x) = c_1\,y_1(x) + c_2\,y_2(x) + \ldots + c_n\,y_n(x)$$

in which coefficients $c_1, c_2, \ldots, c_n$ denote arbitrary constants. According to the principle of superposition, if $y_1(x), y_2(x), \ldots, y_n(x)$ be separately solutions of a homogeneous linear differential

equation, any linear combination thereof is likewise a solution, because the differential operator $L_n(y)$ for such an equation is a linear operator. If $y_1(x), y_2(x), ..., y_n(x)$ be, separately, both solutions of a homogeneous linear differential equation of number equal to the order of the equation and linearly independent, their linear combination constitutes a general solution of that equation.

The wronskian of formulae or functions in a set $\{w_1(x), w_2(x), ..., w_n(x)\}$ on an interval $x_1 \le x \le x_2$, for which each $w_j(x)$ possesses $n - 1$ derivatives on this interval, is a determinant

$$W(w_1, w_2, ..., w_n) = \begin{bmatrix} w_1 & w_2 & ... & w_n \\ \dfrac{\partial}{\partial x} w_1 & \dfrac{\partial}{\partial x} w_2 & ... & \dfrac{\partial}{\partial x} w_n \\ ... & ... & ... & ... \\ \dfrac{\partial^{n-1}}{\partial x^{n-1}} w_1 & \dfrac{\partial^{n-1}}{\partial x^{n-1}} w_2 & ... & \dfrac{\partial^{n-1}}{\partial x^{n-1}} w_n \end{bmatrix}$$

with a property that, if $W(w_1, w_2, ..., w_n) \ne 0$ for at least one point on that interval, the functions or formulae in that set are linearly independent there. If the wronskian be identically zero and if formulae $w_j(x)$ be not known to be solutions of the same linear differential equation, one must test directly for linear dependence -- whether

$$c_1 y_1(x) + c_2 y_2(x) + ... + c_n y_n(x) = 0$$

is satisfied for constants c_j in some set provided that not all $c_j = 0$. For instance, for homogeneous linear differential equation $\left(\dfrac{d^2}{dx^2} y(x) \right) + y(x) = 0$, two solutions, or solution vectors, are

$y_1(x) = \cos(x)$ and $y_2(x) = \sin(x)$; the wronskian is

$$\begin{bmatrix} \cos(x) & \sin(x) \\ \dfrac{d}{dx} \cos(x) & \dfrac{d}{dx} \sin(x) \end{bmatrix}$$

which evaluates to $\cos(x)^2 + \sin(x)^2$ that is nowhere zero. For that reason and because the differential equation is *normal* because the coefficient of the second derivative is unity, the two functions considered as vectors form a basis for the solution space of that equation, according to a set $\{\cos(x), \sin(x)\}$, and the general solution is $y(x) = c_1 \cos(x) + c_2 \sin(x)$. An alternative solution, proferred in *Maple*'s output, is the basis $\{e^{(ix)}, e^{(-ix)}\}$ of which the two functions as vectors are linearly independent, but the two basis sets are linearly dependent because each of $\sin(x)$ or $\cos(x)$ is expressible according to Euler's relations.

For a general linear homogeneous differential equation of second order with non-constant coefficients,

$$\left(\dfrac{d^2}{dx^2} y(x) \right) + p(x) \left(\dfrac{d}{dx} y(x) \right) + q(x) y(x) = 0$$

the general solution is a sum of two linearly independent functions, such as $u(x)$ and $v(x)$, of form,

$$y(x) = c_1 u(x) + c_2 v(x).$$

The wronskian is the determinant of this matrix,

$$W(x) = \begin{bmatrix} u(x) & v(x) \\ \dfrac{d}{dx}u(x) & \dfrac{d}{dx}v(x) \end{bmatrix}$$

which equals $u(x)\left(\dfrac{d}{dx}v(x)\right) - v(x)\left(\dfrac{d}{dx}u(x)\right)$. If we differentiate that wronskian, we obtain

$$\frac{d}{dx}W(x) = u(x)\left(\frac{d^2}{dx^2}v(x)\right) - v(x)\left(\frac{d^2}{dx^2}u(x)\right)$$

Substituting each solution separately into the original differential equation yields

$$\frac{d^2}{dx^2}u(x) = -p(x)\left(\frac{d}{dx}u(x)\right) - q(x)\,u(x)$$

and

$$\frac{d^2}{dx^2}v(x) = -p(x)\left(\frac{d}{dx}v(x)\right) - q(x)\,v(x)$$

Substituting those results into the derivative of the wronskian yields

$$\frac{d}{dx}W(x) = p(x)\left(-u(x)\left(\frac{d}{dx}v(x)\right) + v(x)\left(\frac{d}{dx}u(x)\right)\right)$$

of which the right side is just $-p(x)$ multiplied by the wronskian. The derivative of the wronskian is hence equal to $-p(x)$ multiplied by the wronskian,

$$\frac{\partial}{\partial x}W(x,y) = -p(x)\,W(x,y)\,;$$

of which the solution is

$$W(x) = c\,e^{\left(\int -p(x)\,dx\right)}$$

in which appears c as an arbitrary constant of integration as a pre-exponential factor coefficient of an exponential function of an indefinite integral. If the original differential equation of second order lack a term in a first derivative, so that $p(x) = 0$, the wronskian hence takes a value zero, if the solutions $u(x)$ and $v(x)$ be linearly dependent, or a positive or negative constant, if those solutions be linearly independent. If $p(x) \neq 0$, the wronskian has a fixed sign for x between two adjacent singularities; for instance, if $p(x) = -\dfrac{1}{x}$, $W(x) = c\,x$ that reverses sign across a singularity.

As a special case of a linear homogeneous differential equation, we consider those equations with constant coefficients. For such a differential equation of second order,

$$\left(\frac{d^2}{dx^2}y(x)\right) + c_1\left(\frac{d}{dx}y(x)\right) + c_0\,y(x) = 0$$

which becomes in operator form,

$$(D^2)(y) + c_1\,D(y) + c_0\,D^0\,y(x) = 0$$

in which c_1 and c_0 are real constants and $D^0\,y(x) = y(x)$,

$$deq := \left(\frac{d^2}{dx^2} y(x)\right) + c_1 \left(\frac{d}{dx} y(x)\right) + c_0\, y(x) = 0$$

we write a corresponding algebraic equation on replacing operator D with a scalar quantity λ,

$$\lambda^2 + c_1\, \lambda + c_0 = 0$$

which we either derive on substituting $y = e^{(\lambda x)}$,

$$deq := e^{(\lambda x)}(\lambda^2 + \lambda\, c_1 + c_0) = 0$$

or simply replace D^2 with λ^2, D with λ, D^0 with unity, yielding this quadratic equation,

$$deq := \lambda^2 + \lambda\, c_1 + c_0 = 0$$

that has this solution for the roots.

> solve(deq, lambda);

$$-\frac{1}{2}\frac{c_1}{1} + \frac{1}{2}\frac{(c_1^2 - 4\, c_0)^{\left(\frac{1}{2}\right)}}{1} \qquad -\frac{1}{2}\frac{c_1}{1} - \frac{1}{2}\frac{(c_1^2 - 4\, c_0)^{\left(\frac{1}{2}\right)}}{1}$$

The roots thus become $\lambda_1 = -\dfrac{c_1}{2} + \dfrac{(c_1^2 - 4\, c_0)^{\left(\frac{1}{2}\right)}}{2}$ and $\lambda_2 = \dfrac{c_1}{2} - \dfrac{(c_1^2 - 4\, c_0)^{\left(\frac{1}{2}\right)}}{2}$. In the sollution of this quadratic equation, three possible cases hence arise.

- The two roots λ_1 and λ_2 are both real and distinct; this case implies that $c_1^2 \neq 4\, c_0$ and $c_1^2 > 4\, c_0$.
  The two linearly independent solutions become $e^{(\lambda_1 x)}$ and $e^{(\lambda_2 x)}$, and the general solution is
  $y(x) = A_1\, e^{(\lambda_1 x)} + A_2\, e^{(\lambda_2 x)}$ in which A_1 and A_2 are integration constants typically evaluated according to initial conditions. If $\lambda_1 = -\lambda_2$, implying that $c_1 = 0$ and $c_0 < 0$, the general solution is expressible as $y(x) = B_1\, \cosh(\lambda_1\, x) + B_2\, \sinh(\lambda_1\, x)$.

- The two roots λ_1 and λ_2 are <u>complex conjugates</u> of each other, expressed as $\alpha + i\, \beta$ and $\alpha - i\, \beta$ with $i = \sqrt{-1}$, because, with real constants c_0 and c_1, if one root be <u>complex</u>, the other must be its conjugate. The two linearly independent solutions become $e^{((\alpha + i\beta)x)}$ and $e^{((\alpha - i\beta)x)}$ and the general solution in complex form is $y(x) = A_1 e^{((\alpha + i\beta)x)} + A_2\, e^{((\alpha - i\beta)x)}$, or
  $y(x) = e^{(\alpha x)}(B_1\, \cos(\beta\, x) + B_2\, \sin(\beta\, x))$ in trigonometric form.

- The two roots λ_1 and λ_2 are equal, $\lambda_1 = \lambda_2$, in which case the two linearly independent solutions become $e^{(\lambda_1 x)}$ and $x\, e^{(\lambda_1 x)}$; the general solution is thus $y(x) = A_1\, e^{(\lambda_1 x)} + A_2\, x\, e^{(\lambda_1 x)}$.

  For a general linear homogeneous differential equation of order n but with constant

coefficients, the characteristic equation is analogously derived on replacing a derivative $\dfrac{d^j}{dx^j} y(x)$

with D^j and thence D^j with λ^j, $0 \leq j \leq n$, so generating a polynomial in λ of which the roots become the linearly independent solutions, and their sum becomes the general solution; for this purpose, a term $y(x)$ that is deemed a derivative of order zero becomes replaced by D^0, and that by unity accordingly at the next stage of replacement. If these roots be distinct, the solution has a form $y(x) = A_1\, e^{(\lambda_1 x)} + A_2\, e^{(\lambda_2 x)} + ... + A_n\, e^{(\lambda_n x)}$. If root λ_j have multiplicity k, such that $(\lambda - \lambda_j)^k$ be a factor of the characteristic equation, but $(\lambda - \lambda_j)^{(k+1)}$ not such a factor, the k linearly independent solutions $e^{(\lambda_j x)}$, $x\, e^{(\lambda_j x)}$, ..., $x^k\, e^{(\lambda_j x)}$ are combined with the $n - k$ other linearly independent solutions $e^{(\lambda_1 x)}$, ... to form the complete general solution. For differential equations of large order n, factoring the characteristic equation might be difficult, requiring numerically imprecise roots, but with the employment of sufficient digits the imprecision might be made as small as desired in a domain of interest.

For a homogeneous linear differential equation of second order,

$$\left(\frac{d^2}{dx^2}\, y(x) \right) + P(x)\left(\frac{d}{dx}\, y(x) \right) + Q(x)\, y(x) = g(x) \text{ with } g(x) = 0$$

and for which $x = 0$ is an <u>ordinary point</u>, the general solution on an interval containing this point is $y(x) = a_0\, y_1(x) + a_1\, y_2(x)$, in which appear linearly independent functions $y_1(x)$ and $y_2(x)$ <u>analytic</u> at $x = 0$ and their arbitrary constants a_0 and a_1 as coefficients. A series $\displaystyle\sum_{j=0}^{\infty} a_j\, x^j$ in x to various powers is a trial solution of this differential equation, for which substitution into this equation and on collection of coefficients of x to each power yields an equation containing terms a_j of finite number; solution of the latter equation for coefficient a_j of greatest value of j produces a <u>recurrence relation</u> with which one determines sequentially a_j, $j = 2, 3, 4, ...$, in terms of a_0 and a_1. Substitution of the latter values into the power series then yields a solution of form $y(x) = a_0\, y_1(x) + a_1\, y_2(x)$. This method of power series is applicable only when $x = 0$ is an ordinary point. For the homogeneous linear differential equation above, point x_0 is a <u>regular singular point</u> if x_0 be not an ordinary point but both $(x - x_0)\, P(x)$ and $(x - x_0)^2\, Q(x)$ are analytic at x_0. For a regular singular point elsewhere than at $x_0 = 0$, a translation $u = x - x_0$ to the origin allows direct application of a solution in series, or, for a regular singular point at $x_0 = \infty$, independent variable x is replaced by $\chi = \dfrac{1}{x}$ so that the resulting equation becomes solved in series near $\chi = 0$, if practicable.

When the origin is a regular singular point of a linear differential equation of order n, a solution in series invariably exists for $y(x) = x^m \left(\displaystyle\sum_{j=0}^{\infty} a_j\, x^j \right)$ in which $a_0 \neq 0$ and m and coefficients a_j remain to be determined. Substitution of this series into the differential equation yields a term in $x^{(m-1)}$ of least order, of which the coefficient is a product a_0 with an expression in m; that expression set

equal to zero is called an <u>indicial equation</u>. Each root of that indicial equation corresponds to a separate particular solution; the general solution of the homogeneous equation is $\sum_{k=1}^{n} x^{m_k} y_k(x)$ in

which each $y_k(x)$ is a sum $\sum_{j=0}^{\infty} a_j x^j$, in which coefficients a_j in each set correspond in turn to a

particular value of m_k. If two roots of the indicial equation be equal, only one solution is obtainable; if the two roots differ by other than an integer, two solutions are obtainable. If the two roots differ by an integer, the larger integer yields a solution but the smaller integer might or might not yield a solution; a test by inserting a trial solution into the original differential equation is required for a decision on this matter.

For the non-homogeneous equation with $g(x) \neq 0$, if $g(x)$ have an expansion as a Maclaurin series, the preceding approach is modifiable to solve that non-homogeneous equation on expressing $g(x)$ as a Maclaurin series and setting the coefficients of x to each power on the left side equal to their counterparts on the right side. The general solution has a form $y(x) = a_0 y_1(x) + a_1 y_2(x) + y_3(x)$ in which the former two terms constitute the general solution of the associated homogeneous differential equation and the latter term is a particular solution of the non-homogeneous equation.

For a homogeneous or non-homogeneous linear differential equation for which initial values are given, this equation is solved first as above; the initial values then serve to enable an evaluation of the arbitrary constants a_0, a_1, ... For the solution about an ordinary point x_0 other than $x = 0$, an altered variable $u = x - x_0$ according to which the ordinary point is translated to the origin might simplify the algebra involved in application of the method of power series about $u = 0$.

These solutions have little value in the context of an algebraic processor such as *Maple* with powerful facilities to solve differential equations both algebraically and numerically, and are thus of mostly heuristic interest and value; if an attempt to solve an ordinary-differential equation fail to yield a solution in elementary or special functions, recourse to numerical methods to solve this equation is recommended. In 1725, D. Bernoulli showed that even an ordinary-differential equation of first order might not have a solution expressible in terms of elementary functions of finite number; for thar reason special functions become defined to extend the range of functions applicable to the solution of physical and chemical problems. As distinct from <u>elementary functions</u> that comprise exponential, logarithmic and all trigonometric functions, <u>special functions</u> generally arise as solutions of partial-differential equations of second order that become separable into ordinary-differential equations with variable coefficients, of which many are described through internet at *www.dlmf.nist.gov*.

For a non-homogeneous linear differential equation of order n as derived above to have a form $L_n(y(x)) = g(x)$, a particular solution might be found if $g(x)$ have a form

- $g(x) =$ polynomial of degree $m \leq n$, in which case a particular solution is

$$y_p(x) = e^{(\alpha x)} (A_1 x + A_2 x^2 + ... + A_m x^m);$$

- $g(x) = k\,e^{(q\,x)}$, in which case a particular solution might have a form $y_p(x) = A\,e^{(\alpha\,x)}$;

- $g(x) = q_1 \cos(\alpha\,x) + q_2 \sin(\alpha\,x)$ in which q_1, q_2, α are known constants, in which case a particular solution might have a form $y_p(x) = A_1 \cos(\alpha\,x) + A_2 \sin(\alpha\,x)$; even if one of q_1 and q_2 be zero -- i.e. $g(x)$ lacks the corresponding term, both sine and cosine terms must be present in a particular solution until one might be proved to be redundant;

- a product of these forms, in which case a particular solution might have a form of a corresponding product of the individual particular forms above.

These prospective particular solutions contain arbitrary parameters α, A, A_1, A_2, ..., to be evaluated on substitution of the particular solution into the non-homogeneous equation. If $g(x)$ and all its derivatives be expressible in terms of linearly independent functions in the same finite set, a method of undetermined coefficients is applicable; an assumed form $y_p(x)$ is thereby substituted into the differential equation, and the arbitrary parameters become evaluated again on substituting the prospective particular solution into the non-homogeneous equation and equating coefficients of like terms. These and other appropriate operations are automatically undertaken in *Maple*'s solution of a proffered ordinary-differential equation.

An ordinary-differential equation for which initial conditions are specified is solvable on applying these conditions to the general solution of the non-homogeneous differential equation, which enables an evaluation of the arbitrary constants appearing in a solution as constants of integration.

For a differential equation that is inseparable into a form directly integrable as indicated above, there exist two general methods of solution, either finding an integrating factor -- if practicable -- or a change of variables that maps one differential equation of which a solution is sought into another equation of which a solution is known. For an equation of first order, a transformation resulting in an integral is a likely route to solution, whereas, for equations of order beyond first, to decrease that order by one unit is a likely intermediate objective. Both these approaches are attempted automatically by *Maple*. All ordinary-differential equations of first order are integrable in that a solution is expressible in implicit form, with expressions involving algebraic operations, special functions and integrals. A solution of an ordinary-differential equation posed without initial condition or boundary value contains integration constants of number equal to the order of that equation: an initial condition specifies the value of a dependent variable, or a value of a derivative if required, at zero value of independent variable; a boundary condition specifies the value of a dependent variable at an end point of a particular region of interest, or at a point within such a region of interest, for the solution of a differential equation, for instance of second order for various cases of chemical and physical interest. A solution to a differential equation for which are provided initial conditions or boundary conditions both solves that equation and satisfies all subsidiary conditions.

In mathematics, introducing a transformation of some kind proves a useful strategy when a problem of interest is more amenable to solution through deployment of a transformed variable. As a simple example, consider a problem of multiplying two positive numbers together: if we transform each number into its logarithm and add the two logarithms,

$$\log(x\,y) = \log(x) + \log(y)$$

the resulting number can be transformed through an *inverse* function -- an *exponential* function -- to yield an answer to the original problem, thus converting a multiplication into a sum, apart from transformations. Although in this particular example this procedure seems cumbersome, the principle of using an *isomorphism* to transform between two distinct modes of calculation provides a useful tool. As a second example involving vectors, as discussed in section 6.205, we can use an isomorphism between all vectors, as a set in three-dimensional space, and *matrices*, each comprising a single column, in a corresponding set; an addition of vectors thus becomes transformed from a *geometric* problem to a problem of matrix *algebra*: we add, for instance, two vectors by evaluating sums of appropriate matrix representatives, followed by inverse transformation to produce a corresponding vector in space. For a further and chemically important example, a model to interpret results obtained from measurements of diffraction of xrays from a crystal involves constructing a function for the electronic *density* from the structure factors defined through a transformation to reciprocal space. Regardless what transformation we might undertake to reformulate a problem, on obtaining a solution of that reformulated problem, we subsequently generate an answer of interest by applying the inverse transformation.

For this non-homogeneous linear differential equation of second order,

$$\left(\frac{d^2}{dx^2}\,y(x)\right) + p(x)\left(\frac{d}{dx}\,y(x)\right) + q(x)\,y(x) = f(x)$$

for which $p(x)$, $q(x)$ and $f(x)$ are continuous in a domain $[a, b]$, these boundary conditions might be applicable,

$$\alpha_1\,y(x)\big|_{x=a} + \beta_1\left(\frac{d}{dx}\,y(x)\right)\bigg|_{x=a} = c_1, \qquad \alpha_2\,y(x)\big|_{x=b} + \beta_2\left(\frac{d}{dx}\,y(x)\right)\bigg|_{x=b} = c_2$$

in which α_1, α_2, β_1, β_2, c_1, c_2 are all real constants, with α_1 and β_1 not both zero, and α_2 and β_2 not both zero. If $f(x)$, c_1 and c_2 all be zero, this problem is homogeneous, otherwise non-homogeneous. To solve this problem, by standard methods one finds a solution to the general differential equation, and then applies the boundary conditions to evaluate the arbitrary constants in that solution. For two linearly independent solutions $y_1(x)$ and $y_2(x)$, this problem has non-trivial solutions, in addition to the trivial solution $y(x) = 0$, only if the determinant of this matrix

$$\begin{bmatrix} \alpha_1\,y_1(x)\big|_{x=a} + \beta_1\left(\frac{d}{dx}\,y_1(x)\right)\bigg|_{x=a} & \alpha_1\,y_2(x)\big|_{x=a} + \beta_1\left(\frac{d}{dx}\,y_2(x)\right)\bigg|_{x=a} \\ \alpha_2\,y_1(x)\big|_{x=b} + \beta_2\left(\frac{d}{dx}\,y_1(x)\right)\bigg|_{x=b} & \alpha_2\,y_2(x)\big|_{x=b} + \beta_2\left(\frac{d}{dx}\,y_2(x)\right)\bigg|_{x=b} \end{bmatrix}$$

equal zero. The non-homogeneous problem thus defined with these boundary values has a unique solution only if the associated homogeneous differential equation with $f(x) = 0$ have only trivial solution $y(x) = 0$, which is a unique solution.

For a more general homogeneous linear differential equation

$$\left(\frac{d^2}{dx^2}\,y(x)\right) + p(x, \lambda)\left(\frac{d}{dx}\,y(x)\right) + q(x, \lambda)\,y(x) = 0,$$

non-trivial solutions exist for only particular values of λ called *eigenvalues*, for which the

corresponding solutions $y(x)$ are called <u>eigenfunctions</u>. As a simple case of such an equation, we take $p(x, \lambda) = 0$ and $q(x, \lambda) = -\lambda$, yielding $\dfrac{d^2}{dx^2} y(x) = \lambda \, y(x)$; a general solution of this equation is

$$y(x) = c_1 \, e^{(-\sqrt{\lambda} \, x)} + c_2 \, e^{(\sqrt{\lambda} \, x)}$$

This solution is consistent with the differential equation being linear with constant coefficients, for which an alternative expression is

$$(D^2)(y)(x) - \lambda \, y(x) = 0;$$

when we replace D with scalar quantity m to obtain $m^2 - \lambda = 0$ that has roots $\pm\sqrt{\lambda}$, we proceed to the general solution as above. If $\lambda = 0$, the general solution of $\dfrac{d^2}{dx^2} y(x) = 0$ is simply

$y(x) = c_1 + c_2 \, x$; when we apply boundary conditions $y(x)\big|_{x=0} = 0$ and $y(x)\big|_{x=L} = 0$ to evaluate the arbitrary constants c_1 and c_2, we obtain $c_1 = c_2 = 0$, which as a trivial solution $y(x) = 0$ precludes $\lambda = 0$ from being an eigenvalue. For $\lambda \neq 0$, the general solution containing exponential terms above likewise generates a trivial solution unless $\lambda = -\dfrac{n^2 \, \pi^2}{L^2}$ with n integer, so that the general

solution becomes $y(x) = c_n \sin\left(\dfrac{n \, \pi}{L}\right)$; the set of these solutions in an infinite sequence constitutes a <u>Fourier sine series</u> with <u>Fourier coefficients</u> c_n. If, instead of boundary condition $y(x)\big|_{x=0}$, we

have $\left(\dfrac{d}{dx} y(x)\right)\Big|_{x=0}$, the eigenfunctions become cosine instead of sine functions, so constituting a Fourier cosine series.

A differential equation of <u>Sturm-Liouville form</u>, which arises in many problems of mathematical physics and engineering and which is a homogeneous equation subject to boundary conditions, is written in standard form as

$$\left(\dfrac{d}{dx}\left(p(x)\left(\dfrac{d}{dx} y(x)\right)\right)\right) + (\lambda \, w(x) + q(x)) \, y(x) = 0$$

or in expanded form,

$$p(x)\left(\dfrac{d^2}{dx^2} y(x)\right) + \left(\dfrac{d}{dx} p(x)\right)\left(\dfrac{d}{dx} y(x)\right) + (\lambda \, w(x) + q(x)) \, y(x) = 0$$

in which $p(x)$, $\dfrac{d}{dx} p(x)$, $q(x)$ and $w(x)$ are continuous on interval $[a, b]$ and also both $p(x)$ and $w(x)$ are rigorously positive on that interval and for which apply these boundary conditions,

$$\alpha_1 \, y(x)\big|_{x=a} + \beta_1 \left(\dfrac{d}{dx} y(x)\right)\Big|_{x=a} = 0 \;, \qquad \alpha_2 \, y(x)\big|_{x=b} + \beta_2 \left(\dfrac{d}{dx} y(x)\right)\Big|_{x=b} = 0 \;,$$

or these periodic boundary conditions,

$$y(x)\big|_{x=a} = \left(\dfrac{d}{dx} y(x)\right)\Big|_{x=a} \;, \qquad y(x)\big|_{x=b} = \left(\dfrac{d}{dx} y(x)\right)\Big|_{x=b} \;,$$

and has eigenfunctions as solutions for which the eigenvalues are real and non-negative, in a rigorously increasing infinite sequence, analogously for the simple eigenvalue problem above;

each eigenvalue has only one linearly independent eigenfunction. The allowed values of λ for which the differential equation satisfies the boundary conditions are called <u>eigenvalues</u> and the corresponding solutions $y(x)$ constitute the <u>eigenfunctions</u>. The operator

$$\left(\frac{d}{dx}\left(p(x)\left(\frac{d}{dx}y(x)\right)\right)\right) + q(x)\,y(x)$$

is <u>self-adjoint</u> if it be equal to its complex conjugate. Any two eigenfunctions $y_j(x)$ and $y_k(x)$ in this set satisfy this relation,

$$\int_a^b w(x)\,y_j(x)\,y_k(x)\,dx = 0$$

in which $w(x)$ is called a <u>weight function</u>, called <u>orthogonality</u>; if furthermore each function $y_j(x)$ satisfy this integral,

$$\int_a^b w(x)\,y_j(x)\,y_j(x)\,dx = 1$$

with the same weight function, called *normalization,* the two descriptors become merged into <u>orthonormal</u>. For a function $f(x)$ that is <u>piecewise</u> <u>smooth</u> on an open interval $]a,\,b[$ because both $f(x)$ and $\dfrac{d}{dx}f(x)$ are there piecewise <u>continuous</u>, and with $\{y_j(x)\}$ as a set of all eigenfunctions that satisfy a particular differential equation of Sturm-Liouville form, one can form an expansion

$$f(x) = \sum_{j=1}^{\infty} c_j\,y_j(x)$$

to represent $f(x)$ on that open interval; coefficients c_j are evaluated with this ratio of integrals,

$$c_j = \frac{\displaystyle\int_a^b w(x)\,f(x)\,y_j(x)\,dx}{\displaystyle\int_a^b w(x)\,y_j(x)^2\,dx}$$

For regular Sturm-Liouville problems,

- there exist uncountable eigenvalues of which the magnitudes can be ordered and that can hence be indexed with non-negative integers 0,1,2,3...;

- all eigenvalues are real;

- for each eigenvalue there corresponds a unique eigenfunction; these eigenfunctions form a complete set with respect to any piecewise smooth function over a finite interval; over such an interval the function is representable as a generalized Fourier expansion in terms of eigenfunctions $\displaystyle\sum_{j=0}^{\infty} F(j)\,\phi_j(x)$, in which coefficients $F(j)$ are appropriately evaluated Fourier coefficients;

- the latter infinite series converges to the mean of the left and right limits of the function at any point in the interval;

- eigenfunctions having distinct eigenvalues are orthogonal relative to a weight function $w(x)$ over the interval; if, further, eigenfunctions be normalized, the statement of orthonormal functions makes the weighted inner or scalar product equal to Kronecker's δ function.

A Sturm-Liouville system is one composed of a Sturm-Liouville equation and supplementary conditions, in a set, on $p(x)$, $q(x)$, $w(x)$ and $y(x)$ that satisfy certain limits; such conditions of an important system can include that $w(x)$ be positive and $q(x)$ be continuous over an interval $[a, b]$.

We consider an ordinary linear homogeneous differential equation of second order of form

$$L_z y(z) \;=\; P_0(z)\,y''(z) \;+\; P_1(z)\,y'(z) \;+\; P_2(z)\,y(z) = 0$$

with polynomial coefficients $P_0(z)$, $P_1(z)$, $P_2(z)$ considered in the complex z plane, y'' implying a second derivative with respect to z and y' a first derivative with respect to z. We suppose that those polynomials possess no common factor depending on z; the degrees of those polynomials are for $P_0(z)$ k_0, for $P_1(z)$ k_1 and for $P_2(z)$ k_2. Beyond plane z as the range of definition of this differencial equation, the Riemann sphere allows inclusion of point $z = \infty$ in our treatment; complex plane z including that point is denoted $\mathbf{C}$. Being second order, the differential equation has two linearly independent solutions, of which the general solution is an arbitrary linear combination of these particular solutions. All points of $\mathbf{C}$ in the vicinity of which the initial-value problem with arbitrary initial data $y(z_0) = y_0$, $y'(z_0) = y'_0$ is solvable are called ordinary points of the equation. Point $z = \infty$ becomes studied after substitution $z = \dfrac{1}{\zeta}$ with further consideration of whether $\zeta = 0$ is an ordinary point. In the vicinity of an ordinary point, all particular solutions of the differential equation are holomorphic or analytic functions. For example, Bessel's equation

$$z^2\,y''(z) \;+\; z\,y'(z) \;+\; (z^2 - u^2)\,y(z) \;=\; 0$$

has two singularities, explicitly at $z = 0$ and $z = \infty$; all other points of $\mathbf{C}$ are ordinary points of Bessel's equation. If both linearly independent solutions of $L_z y(z)$ be functions of finite order at singularity $z = z^*$, the singularity is regular, otherwise irregular. If function $P(z) = \dfrac{P_1(z)}{P_0(z)}$ have a pole at point $z = z^*$ of at most first order and if function $Q(z) = \dfrac{P_2(z)}{P_0(z)}$ have a pole at point $z = z^*$ of at most second order, this point is called a Fuchsian singularity. The singular point at $z = \infty$ is a Fuchsian singularity if $P(z) = O(z^{(-1)})$ and $Q(z) = O(z^{(-2)})$; the definition of the singularity at $z = \infty$ is reduced to the definition at a finite point z^* on substituting $z^{(-1)}$ --> $z - z^*$. For instance, the singular point $z = 0$ is a Fuchsian singularity of Bessel's equation but the point at $z = \infty$ is not a Fuchsian singularity. Euler's equation

$$z^2\,y''(z) \;-(a - b - 1)\,z\,y'(z) \;+ a\,b\,y(z) \;=\; 0$$

that has two singularities $z = 0$ and $z = \infty$ is a Fuchsian equation whereas Bessel's equation is not.

An integral transform converts a differential equation into an algebraic equation. In the context

of solving a differential equation -- especially one with boundary conditions specified, integral transforms associated with french mathematicians [Laplace](#) and [Fourier](#) are useful in transforming each term of a differential equation in an independent variable, such as t for time, into another function in a reciprocal space in which the independent variable is a frequency ν, with dimensions hence of inverse time. For a Laplace transform in particular, two pertinent features ease the problem of obtaining a solution of an original differential equation with its appropriate boundary conditions: application of a Laplace transform is naturally suited to those differential equations that involve initial values as boundary conditions, but a simple transformation of a variable might serve to convert a differential equation into one with initial conditions specified with respect to the new variable. In either case, the boundary conditions become embedded in an algebraic equation produced through a Laplace transform; in this way one circumvents a search for a general solution. A Laplace transform is useful also to solve a linear differential equation with constant coefficients, especially if there be a driving term defined [piecewise](#), because a Laplace transform converts a differential equation into an algebraic equation, which might be solved more easily. A *Laplace transform* $F(s)$ of a function of time, $f(t)$, defined on an interval $0 \le t < \infty$, is formed as an [improper integral](#):

$$F(s) = \int_0^\infty e^{(-ts)} f(t) \, dt$$

Here [integration](#) is along the [positive](#) [real](#) [axis](#) for t; a [factor](#) $e^{(-ts)}$ in this integrand is a [kernel](#) of a Laplace transformation; variable s might be real or [complex](#). Because the upper end point is [infinity](#), this integral defining $F(s)$ is an improper integral, for which [convergence](#) is achieved on an imposition of two technical restrictions on $f(t)$: its [domain](#) $[0, \infty]$ can be [divided](#) into [intervals](#) with $f(t)$ [continuous](#) in the [interior](#) of each interval and approaching finite [limits](#) at each end point thereof, so that $f(t)$ is at least piecewise continuous; $f(t)$ is of exponential order, such that there exists a positive constant κ for which a product

$$e^{(-\kappa t)} \left| f(t) \right| < c$$

remains bounded by finite constant c as $t \to \infty$. The greatest smaller bound on the value of κ is called an abscissa of convergence of $f(t)$. In practice, to solve a differential equation arising from a chemical model, these technical details are of little or no concern.

Laplace transforms have these important properties:

- linearity -- if Laplace transforms of $f(x)$ and $g(x)$ exist and that of $f(x)$ be $F(s)$ and of $g(x)$ be $G(s)$, the Laplace transform of the sum is the sum of the Laplace transforms, $L(f(x) + g(x)) = L(f(x)) + L(g(x)) = F(s) + G(s)$, in which $L(\ )$ implies a Laplace transform;

- distributivity -- for constant c, the Laplace transform of $c\,f(x)$ is $c\,F(s)$;

- multiplication by $e^{(\alpha x)}$ -- if a Laplace transform of $f(x)$ exist and be $F(s)$, the Laplace transform of $e^{(\alpha x)} f(x)$ is $F(s - \alpha)$, for α any constant;

- multiplication by x^n -- if a Laplace transform of $f(x)$ exist and be $F(s)$, the Laplace transform of

$x^n f(x)$ is $(-1)^n \left(\dfrac{d^n}{ds^n} F(s) \right)$ for n any positive integer;

- division by x -- if a Laplace transform of $f(x)$ exist and be $F(s)$ and if $\lim\limits_{x \to 0} \dfrac{f(x)}{x}$ exist with $x > 0$, the Laplace transform of $\dfrac{f(x)}{x}$ is $\displaystyle\int_s^\infty F(u)\, du$;

- integral -- if a Laplace transform of $f(x)$ exist and be $F(s)$, the Laplace transform of $\displaystyle\int_0^x f(u)\, du$ is $\dfrac{F(s)}{s}$;

- periodicity -- if $f(x)$ be periodic with period a such that $f(x + a) = f(x)$, the Laplace transform of $f(x)$ is $\dfrac{1 \displaystyle\int_0^\infty e^{(-s\,x)} f(x)\, dx}{1 - e^{(-a\,s)}}$.

In these cases, if x be a distance or have a dimension length, corresponding variable s in the Laplace transform has dimension reciprocal length, or wavenumber.

Just as a derivative of a product differs from a product of derivatives of its factors, a Laplace transform of a product of two functions differs from a product of the individual transforms of those functions. There exists instead a Laplace transform of a convolution of two functions -- multiplication of a special type -- that is a product of individual transforms; in this way one can invert some Laplace transforms and obtain convolutions without explicit evaluation of a convolution integral.

A convolution of two functions $f(x)$ and $g(x)$ is defined as $f(x) * g(x) = \displaystyle\int_0^x f(u)\, g(x - u)\, dt$, in which $*$ denotes a convolution operator, so that $f(x) * g(x) = g(x) * f(x)$; the Laplace transform of that convolution is hence $L(f(x) * g(x)) = L(f(x))\, L(g(x)) = F(s)\, G(s)$, and the inverse Laplace transform of $F(s)\, G(s)$ is thus $f(x) * g(x) = g(x) * f(x)$. An inverse Laplace transform of a product is computed with such a convolution.

Special functions of many kinds arise in mathematical chemistry and physics in the solution of differential equations -- ordinary-differential and partial-differential -- that govern the behaviour of various physical quantities. A commonly occurring equation of this type is Laplace's equation of which examples appear above. The mathematical problem comprises finding functions that not only satisfy the particular differential equation but also fulfill particular conditions on the surfaces bounding the region being considered. The shape of these boundaries commonly makes preferable working in curvilinear coordinates q_1, q_2, q_3 instead of rectangular cartesian coordinates x, y, z. Relations

$$x = x(q_1, q_2, q_3), \qquad y = y(q_1, q_2, q_3), \qquad z = z(q_1, q_2, q_3)$$

express the cartesian coordinates in terms of curvilinear coordinates. For

$$\left(\frac{\partial}{\partial q_j}x\right)\left(\frac{\partial}{\partial q_k}x\right)+\left(\frac{\partial}{\partial q_j}y\right)\left(\frac{\partial}{\partial q_k}y\right)+\left(\frac{\partial}{\partial q_j}z\right)\left(\frac{\partial}{\partial q_k}z\right)=0$$

when $j \ne k$, these curvilinear coordinates are orthogonal. Element dl of length is then found as

$$dl^2 = h_1^{\ 2}\ dq_1^{\ 2} + h_2^{\ 2}\ dq_2^{\ 2} + h_3^{\ 2}\ dq_3^{\ 2}$$

in which

$$h_j^{\ 2} = \left(\frac{\partial}{\partial q_j}x\right)^2 + \left(\frac{\partial}{\partial q_j}y\right)^2 + \left(\frac{\partial}{\partial q_j}z\right)^2$$

which yields, for instance, this laplacian operator operating on quantity Φ:

$$\underline{\text{del}}\ ^2\Phi = \frac{1}{h_1\ h_2\ h_3}\left(\frac{\partial}{\partial q_1}\left(\frac{h_2\ h_3\left(\frac{\partial}{\partial q_1}\Phi\right)}{h_1}\right)\right)+\left(\frac{\partial}{\partial q_2}\left(\frac{h_3\ h_1\left(\frac{\partial}{\partial q_2}\Phi\right)}{h_2}\right)\right)+\left(\frac{\partial}{\partial q_3}\left(\frac{h_1\ h_2\left(\frac{\partial}{\partial q_3}\Phi\right)}{h_3}\right)\right)$$

Adjective *special* is applied to particular functions in this context not because our concern is not with the general properties of functions as in analysis but with the properties of functions that arise in the solution of chemical and physical problems.

As defined above, a function is analytic at a point if expansion as a Taylor series is valid in some vicinity of that point, such that the function be singly valued and possess there derivatives of all orders. In differential equations to be considered, the coefficients of derivatives are analytic functions except possibly at particular isolated points. An ordinary point $x = a$ of a differential equation

$$\left(\frac{d^2}{dx^2}\ y(x)\right) + \alpha(x)\left(\frac{d}{dx}\ y(x)\right) + \beta(x)\ y(x) = 0$$

of second order is a point at which coefficients $\alpha(x)$ and $\beta(x)$ are analytic functions; at such an ordinary point, every solution of the equation is analytic. If Taylor expansions of $\alpha(x)$ and $\beta(x)$ be valid in range $|x - a| < R$, the Taylor expansion is valid for the same range; if $\alpha(x)$ and $\beta(x)$ be polynomials in x, a solution in series of that differential equation is valid for all values of x. If either function $\alpha(x)$ or $\beta(x)$ be not analytic at point $x = a$, this is a singular point of the differential equation; when these functions $\alpha(x)$ and $\beta(x)$ have a nature such that the differential equation is expressible as

$$(x - a)^2\left(\frac{d^2}{dx^2}\ y(x)\right) + (x - a)\ p(x)\left(\frac{d}{dx}\ y(x)\right) + q(x)\ y(x) = 0$$

in which $p(x)$ and $q(x)$ are analytic at point $x = a$, this point is a regular singular point of the differential equation. A singularity of an ordinary-differential equation is a singularity of a solution, such as a pole.

For a differential equation of second order,

$$p(x)\left(\frac{d^2}{dx^2}\ y(x)\right) + q(x)\left(\frac{d}{dx}\ y(x)\right) + r(x)\ y(x) = 0$$

of which coefficients $p(x)$, $q(x)$, $r(x)$ are commonly polynomials of small degree, this equation

becomes <u>self-adjoint</u> if $q(x) = \dfrac{d}{dx} p(x)$ such that it is expressible as

$$\left(\frac{d}{dx}\left(p(x)\left(\frac{d}{dx} y(x) \right) \right) \right) + r(x)\, y(x) = 0 .$$

    Some special functions notable in chemical applications that arise in the solution of differential equations include the following examples. <u>Hermite</u>'s differential equation, in which k is not necessarily integer,

$$\left(\frac{d^2}{dx^2} y(x) \right) - 2\, x \left(\frac{d}{dx} y(x) \right) + 2\, k\, y(x) = 0$$

with k integer yields <u>Hermite's polynomials</u> that conform to this recurrence relation and $H(0, x) = 1$, $H(1, x) = 2\, x$,

$$H(k + 1, x) = 2\, x\, H(k, x) - 2\, k\, H(k - 1, x)$$

    The associated Laguerre equation,

$$x\left(\frac{d^2}{dx^2} y(x) \right) + (m + 1 - x)\left(\frac{d}{dx} y(x) \right) + (k - m)\, y(x) = 0, \;\; k \text{ and } m \text{ integer with } k \geq m,$$

yields associated <u>Laguerre</u> polynomials, $L(k, m, x)$, which are related to confluent hypergeometric series ${}_1F_1(a; c; x)$ through

$$L(k, m, x) = \frac{(-1)^m\, (k!)^2}{m!\, (k - m)!}\; {}_1F_1(-k + m;\, m+1;\, x)$$

whereas Laguerre's differential equation,

$$x\left(\frac{d^2}{dx^2} y(x) \right) + (1 - x)\left(\frac{d}{dx} y(x) \right) + k\, y(x) = 0, \;\; k \text{ integer}$$

yields Laguerre polynomials $L(k, x)$, related to confluent hypergeometric series through

$$L(k, x) = k!\; {}_1F_1(-k;\, 1;\, x)$$

    Analogously to the Laguerre system, we express the associated Legendre differential equation,

$$\left((1 - x^2)\left(\frac{d^2}{dx^2} y(x) \right) - 2\, x\left(\frac{d}{dx} y(x) \right) + k\, (k + 1) - \frac{m^2}{1 - x^2} \right) y(x) = 0, \;\; k \text{ and } m \text{ integer}$$

which yields <u>Legendre</u>'s associated functions of four types. Legendre's differential equation, with $m = 0$ in the associated equation,

$$(1 - x^2)\left(\frac{d^2}{dx^2} y(x) \right) - 2\, x\left(\frac{d}{dx} y(x) \right) + k\, (k + 1)\, y(x) = 0, \;\; k \text{ integer},$$

yields as solution <u>Legendre polynomials</u>, $P(k, x)$.

    Ordinary linear differential equations of second order with polynomial coefficients have many applications in physics and chemistry. To classify these equations, one considers the singular points, the parameters and the confluent cases of equations and their standard forms. Beginning with

$$P_0(z)\left(\frac{d^2}{dz^2} y(z) \right) + P_1(z)\left(\frac{d}{dz} y(z) \right) + P_2\, y(z) = 0 \qquad *$$

in which polynomials $P_0(z)$, $P_1(z)$ and $P_2(z)$ have degrees m_0, m_1, m_2 respectively that lack zeros

common to all three, we rewrite this equation with rational coefficients by dividing through with $P_0(z)$,

$$\left(\frac{d^2}{dz^2}\,y(z)\right) + P(z)\left(\frac{d}{dz}\,y(z)\right) + Q(z)\,y(z) = 0$$

We consider the singularities or poles of rational functions $P(z)$ and $Q(z)$. With a transformation $t = \dfrac{1}{z - z_0}$, in which z_0 is an arbitrary finite point in the complex or z-plane and subsequent consideration of the point $t = 0$, we express the preceding equation in terms of independent variable t,

$$\left(\frac{d^2}{dt^2}\,y(t)\right) + \left(\frac{2}{t} - \frac{P(t^{(-1)} + z_0)}{t^2}\right)\left(\frac{d}{dt}\,y(t)\right) + \frac{1}{t^4}\,Q(t^{(-1)} + z_0)\,y(t) = 0$$

The points in the finite z-plane at which P and Q are holomorphic are ordinary points of the differential equation *; the point at ∞ is an ordinary point of that equation only if

$$Q(z) = O(z^{(-4)}) \qquad \text{and} \qquad P(z) = \frac{2}{z} + O(z^{(-2)}) \qquad \text{as } z \to \infty;$$

a point z_i that is not an ordinary point is a singular point of that equation *. In a vicinity of each such singularity z_i, each function $P(z)$ and $Q(z)$ is expandable in a Laurent series in $(z - z_i)$; the set of singular points of differential equation *, never empty, divides into regular singular points and irregular singular points. The functions of Bessel, Legendre, Laguerre, Weber, Hermite and others appear as special or confluent cases of the hypergeometric function.

  From the differential equation *, one might obtain *confluent equations* through the *coalexcence of singularities*. If the distance between two zeros of equation * depend on a parameter, we suppose that as this parameter tends to zero the two zeros of $P_0(z)$ coalesce. A combination of that coalescence of singularities and the limiting process of the coefficients in the Laurent series tending to ∞ constitutes a *confluence*, according to which one coalescing singularity is at ∞, yielding a confluent case of equation *. For instance, Euler's equation,

$$(z - z_1)^2\left(\frac{d^2}{dz^2}\,y(z)\right) + A\,(z - z_1)\left(\frac{d}{dz}\,y(z)\right) + B\,y(z) = 0$$

can be transformed into the confluent case,

$$\frac{d^2}{dz^2}\,y(z) + E\left(\frac{d}{dz}\,y(z)\right) + D\,y(z) = 0$$

which might be solvable more easily.

  Five linear ordinary-differential equations of second order with multiple parameters, attributed to Heun, include as particular cases the Lame, Mathieu, spheroidal-wave and hypergeometric equations -- and thereby most known equations of mathematical physics. Five Heun functions are defined as the solutions to each of five Heun equations of which this be the most general,

$$\left(\frac{d^2}{dx^2}\,y(x)\right) + \left(\frac{\gamma}{x} + \frac{\delta}{x-1} + \frac{\varepsilon}{x-a}\right)\left(\frac{d}{dx}\,y(x)\right) + \frac{(\alpha\,\beta\,x - q)\,y(x)}{x\,(x-1)\,(x-a)} = 0$$

containing variables x independent and y dependent with parameters α, β, γ, δ, ε, a and q, of which

$\alpha + \beta - \gamma - \delta - \varepsilon = 1$, and an accessory parameter q generally serves as a *spectral parameter*; from the general equation of Fuchsian type with four singularities at $x = 0$, $x = 1$, $x = a$, $x = \infty$, four other equations are derivable from this equation on coalescing singularities. As there are four singularities and two Frobenius solutions at each singularity, the local solutions for the Heun equation number eight in total, but with a transformation

$$y(x) = (x - x_j)^{\rho(x_j)} v(x)$$

in which x_j is an arbitrary regular singularity and $\rho(x_j)$ is the non-zero characteristic exponent at this singularity, 24 solutions located at singularities at finite points might be constructed, but the exchange of the singularity at ∞ with any finite singularity increases the total number of local solutions to 72.

Kamke's book, *Differential equations: Methods of Solution and Solutions* (Chelsea Publishing Co., New York, USA, 1959), which with its succeeding editions is regarded as a definitive, even if not comprehensive, compilation of differential equations that have known algebraic solutions, includes a collection of 446 linear ordinary-differential equations of second order, all but four of them being of Heun type. All but those four are hence solvable merely through a single factorization of a polynomial of fourth degree; essentially all applications behind Kamke's linear examples are thereby formulated using Heun equations. Developments in the algebraic solution of differential equations with software such as *Maple* expand the range of solvable equations based on this formalism, but at present only hybrid algebraic and numerical methods are applicable.

The hypergeometric series arises as a solutions of the hypergeometric differential equation

$$x(x-1)\left(\frac{d^2}{dx^2} y(x)\right) + ((\alpha + \beta + 1)x - \gamma)\left(\frac{d}{dx} y(x)\right) + \alpha \beta \, y(x) = 0$$

which has a solution for γ not a positive integer and $|x| < 1$, according the explanation of hypergeometric series in section 2.403, 2.405 and 4.0. Independent variable x might in general lie in complex plane $\mathbf{C}$. The regular singular points of the hypergeometric equation are at

- $x = 0$ with exponents 0, $1 - \gamma$,

- $x = \infty$ with exponents α, β, and

- $x = 1$ with exponents 0, $\gamma - \alpha - \beta$.

Corresponding to the root $r = 0$, a solution exists of form $y = \sum_{j=0}^{\infty} c_j x^j$; corresponding to the root $r = 1$, a solution exists according to substitution $\xi = 1 - x$, and corresponding to the root $x = \infty$, a solution exists of form $y = \sum_{j=0}^{\infty} c_j x^{(-\alpha - j)}$, yielding in all cases a sum of two hypergeometric series as the two independent solutions of the differential equation of second order.

On replacing x by $\dfrac{x}{\beta}$ in the hypergeometric differential equation above and letting $\beta \to \infty$, the differential equation is converted into

$$x\left(\frac{d^2}{dx^2}y(x)\right)+(\gamma-x)\left(\frac{d}{dx}y(x)\right)-\alpha\,y(x)=0$$

known as the *singly conflluent hypergeometric equation*, and also as Kummer's differential equation; it yields solutions of form $_1F_1(\alpha;\gamma;x)$ called a *confluent hypergeometric function*, obtained through the coalescence of singularities such that this equation has only two singularities, regular at $z=1$ and irregular at $z=\infty$ Under a homotopic transformation, this equation becomes

$$\left(\frac{d^2}{dx^2}y(x)\right)+\left(-\frac{1}{4}-\frac{a-\dfrac{c}{2}}{x}+\frac{1-(1-c)^2}{4\,x^2}\right)y(x)=0$$

which is known as Whittaker's equation. The biconfluent hypergeometric equation

$$\left(\frac{d^2}{dx^2}y(x)\right)-x\left(\frac{d}{dx}y(x)\right)-a\,y(x)=0$$

that has only one, irregular, singularity at $z=\infty$, is known as Hermite's differential equation, as presented above. A confluence hence implies a coalescence of two singularities of this equation,

$$\left(\frac{d^2}{dz^2}y(z)\right)+P(z)\left(\frac{d}{dz}y(z)\right)+Q(z)\,y(z)=0$$

in complex plane z such that the corresponding coefficients of singular parts of Laurent series for $P(z)$ and $Q(z)$ in the vicinities of those singularities rearrange to yield the singular part of the Laurent series in the vicinity of the resulting singularity. Hypergeometric functions $F(a,b,c;z)$ satisfy this difference equation,

$$z\,F(a,b,c;z)=\frac{a\,(b-c)}{(a-b)\,(a-b+1)}F(a+1,b-1,c;z)+\frac{c\,(a+b-1)-2\,a\,b}{(a-b+1)\,(a-b-1)}F(a,b,c;z)+$$
$$\frac{b\,(a-c)}{(a-b)\,(a-b-1)}F(a-1,b+1,c;z)$$

and this differential-difference equation.

$$z\,(1-z)\,F'(a,b,c;z)=\frac{a\,b\,(b-c)}{(a-b)\,(a-b+1)}F(a+1,b-1,c;z)+\frac{a\,b\,(2\,c-a-b-1)}{(a-b+1)\,(a-b-1)}F(a,b,$$
$$c;z)+\frac{a\,b\,(a-c)}{(a-b)\,(a-b-1)}F(a-1,b+1,c;z)$$

The importance of these differential equations for chemistry and physics lies in the fact that Schroedinger's temporally independent equation with varied functions of potential energy assumes the form of one or other of these equations.

   Other strategies to solve ordinary-differential equations, which are likely implemented in *Maple*, include the following:

- if the dependent variable, such as y, be absent, let $\dfrac{d}{dx}y(x)=p$ become the dependent variable, hence diminishing the order of the equation by unity;

- if the independent variable, such as x, be absent, let y become the independent variable and

$\dfrac{d}{dx}\text{y}(x) = p$ become the dependent variable, hence also diminishing the order of the equation by unity;

- if the differential equation be homogeneous in y, let $v = \ln(y)$ become a dependent variable so that the resulting equation lacks v and a substitution $\dfrac{d}{dx}\text{v}(x) = p$ then diminishes the order by unity.

For two unknown functions $f(t, x, y)$ and $g(t, x, y)$ in a system of differential equations of first order,

$$\frac{d}{dt}\text{x}(t) = \text{f}(t, x, y), \quad \frac{d}{dt}\text{y}(t) = \text{g}(t, x, y),$$

numbers in an ordered pair (x, y) might be regarded as rectangular cartesian coordinates of a point in plane xy. For functions $x(t)$ and $y(t)$ that constitute a solution of those differential equations, relations $x = \text{x}(t)$ and $y = \text{y}(t)$ become interpreted as parametric equations of a curve in plane xy, which is called a *phase plane* or a plane in phase space of two spatial dimensions. That curve in this plane is called a trajectory of the system. A point (x_0, y_0) such that $\text{f}(t, x_0, y_0) = \text{g}(t, x_0, y_0) = 0$ is called a critical point of the system; for such a point, the system possesses a constant solution $x = x_0$, $y = y_0$, and the trajectory of such a solution comprises that single point. A trajectory might be a closed curve or an open curve.

Most ordinary-differential equations that might arise or that one might devise have no direct algebraic solution, and even solution in terms of established special functions might not be practicable. In such cases one must have recourse to numerical methods, for which one absolutely requires initial values or boundary conditions of number equal to the order of the differential equation, because a numerical approach inherently fails to cope with symbolic quantities. Such a numerical method yields approximate solutions at particular points using only operations addition, subtraction, multiplication, division and functional evaluations. Points are typically chosen as $x_0, x_1, x_2, \dots$ with a constant increment $x_j - x_{j-1} = h$. For a general differential equation of first order written $\dfrac{d}{dx}\text{y}(x) = \text{f}(x, y)$, with an initial value $\text{y}(x_0) = y_0$, a curve of the solution $\text{y}(x)$ must pass a point (x_0, y_0). For a point near x_0, the corresponding value of $y = y_0 + \displaystyle\int_{x_0}^{x} \text{f}(x, y)\, dx$.

According to Picard's method, a first approximation of y in the integrand is y_0, which generates a value y_1 on the left side from the integration on the right side; replacing y in the integrand a second time with y_1 and integrating again yields an improved value y_2 on the left side, and continuing in an iterative manner yields a sequence of values of y, each a better approximation than the preceding one. As a practical procedure, Picard's method suffers from the difficulty of performing the necessary integrations.

According to Euler's method, we calculate $y_{j+1} = y_j + h\,\text{f}(x, y_j) = y_j + h\left(\dfrac{d}{dx}\text{y}(x)\right)_j$, iteratively

beginning at $j = 0$, which corresponds to extrapolation along tangent lines; this process is continued until sufficient points to define a solution curve in a domain of interest are derived. If h be chosen too large, error might be appreciable, whereas if h be chosen too small, numerical error with the arithmetic of real numbers with a finite precision might accumulate or the number of points to cover a particular domain of interest might be excessive.

   To improve the accuracy of Euler's method that might rapidly become inaccurate, more sophisticated methods involving predictor and corrector have been devised, so that the result of a prediction, such as that tangential extrapolation in Euler's method, is subjected to correction; the corrector depends in general on the predictor. Such a modification to Euler's method for which a predicted value is

$$y_{j+1,p} = y_j + h\left(\frac{d}{dx}y(x)\right)_j$$

with increment h and predicted value $y_{j+1,p}$ might have as correction

$$y_{j+1} = y_j + \frac{h}{2}\left(\left(\frac{d}{dx}y(x)\right)_{j+1,p} + \left(\frac{d}{dx}y(x)\right)_j\right).$$

For a correction of increased accuracy, the methods of Runge and Kutta include further terms of a nature of those in Simpson's rule for a numerical integration; for their method of fourth order, such that

$$y_{j+1} = y_j + \frac{g_1 + g_2 + g_3 + g_4}{6}, g_1 = h\,f(x_j, y_y), g_2 = h\,f\left(x_j + \frac{h}{2}, y_j + \frac{g_1}{2}\right), g_3 = h\,f\left(x_j + \frac{h}{2}, y_j + \frac{g_2}{2}\right),$$

$$\text{and } g_4 = h\,f(x_j + h, y_j + g_3),$$

but this approach is formally not of type predictor and corrector. Other methods, of type predictor and corrector, due to Adams, Bashford and Moulton and to Milne require four values $y_j, j = 0 .. 3$, as initial values, which are obtained according to the method of Runge and Kutta. For a numerical solution of differential equations, *Maple* uses by default a method similar to that of Runge and Kutta but modified for enhanced accuracy.

partial-differential equation

   A partial-differential equation contains one or more partial derivatives and must hence involve at least two independent variables; with independent variables x and y and dependent variable z, its general form is hence

$$f\left(x, y, z, \frac{\partial}{\partial x}z, \frac{\partial}{\partial y}z, \frac{\partial^2}{\partial x^2}z, \frac{\partial^2}{\partial y\,\partial x}z, \frac{\partial^2}{\partial y^2}z, ...\right) = 0.$$

A linear partial-differential equation is linear with respect to quantities $z, \frac{\partial}{\partial x}z,$

$\frac{\partial}{\partial y}z, \frac{\partial^2}{\partial x^2}z, \frac{\partial^2}{\partial y\,\partial x}z, \frac{\partial^2}{\partial y^2}z, ...$ The order of a partial-differential equation equals the order of the highest partial derivative within it. A partial-differential equation of first order is homogeneous in the absence of a term free of z and its derivatives; this equation is otherwise non-homogeneous. The general solution of a partial-differential equation differs from that of an ordinary-differential equation in that arbitrary formulae or functions of independent variables occur instead of arbitrary

Among partial-differential equations important in chemistry and physics are

- wave equation in one dimension x and with variable time t, of hyperbolic type,

$$\frac{\partial^2}{\partial t^2} u(x, t) = c^2 \left(\frac{\partial^2}{\partial x^2} u(x, t) \right);$$

- equation for diffusion of mass or heat in one dimension x and with variable time t, of parabolic type, $\frac{\partial}{\partial t} u(x, t) = \lambda \left(\frac{\partial^2}{\partial x^2} u(x, t) \right);$

- Laplace's equation in three dimensions x, y, z,

$$\left(\frac{\partial^2}{\partial x^2} u(x, y, z) \right) + \left(\frac{\partial^2}{\partial y^2} u(x, y, z) \right) + \left(\frac{\partial^2}{\partial z^2} u(x, y, z) \right) = 0, \text{ which is a special case of}$$

- Poisson's equation in three dimensions x, y, z, of elliptic type,

$$\left(\frac{\partial^2}{\partial x^2} u(x, y, z) \right) + \left(\frac{\partial^2}{\partial y^2} u(x, y, z) \right) + \left(\frac{\partial^2}{\partial z^2} u(x, y, z) \right) = -f(x, y, z) \text{ in which } f(x, y, z)/ \text{ might}$$

  represent $\rho(x,y,z)/\varepsilon_0$ with charge density $\rho(x,y,z)$ distributed in space

- Helmholtz's equation in three dimensions, $\left(\frac{\partial^2}{\partial x^2} u(x, y, z) \right) + \left(\frac{\partial^2}{\partial y^2} u(x, y, z) \right) + \left(\frac{\partial^2}{\partial z^2} u(x, y, z) \right)$

  $+ k^2 u(x, y, z) = 0$, in which appear wavenumber k and amplitude $u(x, y, z)$

- Schroedinger's equation in three spatial dimensions x, y, z and dependent on time t for a single particle of mass m subject to a potential energy $V(x, y, z)$ involving only spatial coordinates,

$$-\frac{h^2}{8 \pi^2 m} \left(\left(\frac{\partial^2}{\partial x^2} \right) + \left(\frac{\partial^2}{\partial y^2} \right) + \left(\frac{\partial^2}{\partial z^2} \right) + V(x, y, z, t) \right) \Psi(x, y, z, t) = \frac{i h}{2 \pi} \Psi(x, y, z, t)$$

- For an isotropic body, the thermal conductivity at each point is independent of the direction of flow of thermal energy through that point; the temperature $T = T(x, y, z, t)$ is obtained on solving this partial- differential equation,

$$\left(\frac{\partial}{\partial x} \left(k \left(\frac{\partial}{\partial x} T \right) \right) \right) + \left(\frac{\partial}{\partial y} \left(k \left(\frac{\partial}{\partial y} T \right) \right) \right) + \left(\frac{\partial}{\partial z} \left(k \left(\frac{\partial}{\partial z} T \right) \right) \right) = c \rho \left(\frac{\partial}{\partial t} T \right)$$

  in which appear thermal conductivity k, specific heat c and density ρ; when the latter three parameters are constant, this equation is known as the heat equation in three dimensions,

$$\left(\frac{\partial^2}{\partial x^2} T \right) + \left(\frac{\partial^2}{\partial y^2} T \right) + \left(\frac{\partial^2}{\partial z^2} T \right) = \frac{c \rho}{k} \frac{\partial}{\partial t} T,$$

  which is amenable to solution with Fourier series. The laplacian operator hence appears in not only Laplace's equation but also Helmholtz's equation and Schroedinger's equation, and assumes an appropriate form in a particular system of coordinates.

An harmonic function in region R of plane xy satisfies Laplace's equation in two dimensions,

$$\left(\frac{\partial^2}{\partial x^2}\phi(x,y)\right)+\left(\frac{\partial^2}{\partial y^2}\phi(x,y)\right)=0$$

If a complex function $\phi(z)=u(x,y)+i\,v(x,y)$ be <u>analytic</u> in region R, both $u(x,y)$ and $v(x,y)$ are harmonic functions; these are also harmonic conjugates in that one is determinable from the other through integration and addition of an <u>arbitrary constant</u> according to the <u>Cauchy-Riemann equations</u>:

$$\frac{\partial}{\partial x}u(x,y)=\frac{\partial}{\partial y}v(x,y)\quad\text{and}\quad\frac{\partial}{\partial y}u(x,y)=-\left(\frac{\partial}{\partial x}v(x,y)\right)$$

If function $\phi(z)$ be analytic, it has no dependence on $\overline{z}$; if these Cauchy-Riemann equations be satisfied, function $\phi(z)$ is analytic; $u(x,y)$ is then called the harmonic conjugate of $v(x,y)$, and vice versa.

   The <u>order</u> of a partial-differential equation is that of the derivative of greatest order in that equation; for instance, as the order of the derivatives with respect to spatial variables in all above equations is second, and as a derivative with respect to no other variable exceeds second order, all these equations have second order. These partial-differential equations are also all <u>linear</u> because the total degree of the dependent variable u and of each derivative thereof corresponds to the first power at most, but only the former three equations are <u>homogeneous</u> because each term in those equations contains only the dependent variable or its derivatives; the latter equation, Poisson's equation, is non-homogeneous. A <u>solution</u> of a partial-differential equation in some region of space and time is a function of all spatial and temporal variables for which derivatives appear in that equation, and satisfies the equation throughout that region. For instance, for Laplace's equation in two spatial dimensions,

$$\left(\frac{\partial^2}{\partial x^2}u(x,y)\right)+\left(\frac{\partial^2}{\partial y^2}u(x,y)\right)=0$$

$u(x,y)=x^2-y^2$, $u(x,y)=\cos(x)\,e^y$, $u(x,y)=\arctan\left(\frac{y}{x}\right)$ and $u(x,y)=\ln(x^2+y^2)$ are all

solutions. An application of <u>initial conditions</u>, such as for $t=0$ for the temporal variable, or <u>boundary conditions</u>, which define the dependent variable at particular points of the solution <u>domain</u> that form boundaries of that domain, of appropriate number, yield an <u>unique</u> solution to that differential equation; these conditions, which should number equal to the order of the highest partial derivative, enable the evaluation of arbitrary constants, or even arbitrary functions, that arise from the general solution of that differential equation.

   For a partial-differential equation of form

$$\alpha\left(\frac{\partial^2}{\partial x^2}u(x,y)\right)+2\,\beta\left(\frac{\partial^2}{\partial y\,\partial x}u(x,y)\right)+\gamma\left(\frac{\partial^2}{\partial y^2}u(x,y)\right)=f\left(x,y,u(x,y),\frac{\partial}{\partial x}u(x,y),\frac{\partial}{\partial y}u(x,y)\right),$$

by analogy with profiles of quadric surfaces these conditions describe the nature of this equation:

- $\alpha\gamma-\beta^2=0$, <u>parabolic</u>, for the transfer of mass or heat for instance,

- $\alpha\gamma-\beta^2<0$, <u>hyperbolic</u>, for wave motion and vibrations for instance, and

- $\alpha\gamma-\beta^2>0$, <u>elliptical</u>, for systems described with a potential energy or a steady state.

Unlike an ordinary-differential equation for which the order is also the dimension of the solution space, the dimension of a partial-differential equation with spatial boundary conditions is infinite, and likewise the corresponding basis of the solutions as a vector space. Other than originating in geometric and physical problems, a partial-differential equation is derivable on elimination of arbitrary constants from a given relation between the variables and on elimination of arbitrary functions of those variables. Whereas a general solution of an ordinary-differential equation involves arbitrary constants, a general solution of a partial-differential equation involves arbitrary functions. To illustrate this occurrence of arbitrary functions, we consider the form

$$w(x, y) = y\, f(x) + x\, g(y).$$

Forming partial derivatives with respect to x and y yields

$$\frac{\partial}{\partial x} w(x, y) = y\left(\frac{d}{dx} f(x)\right) + g(y) = p, \text{ and } \frac{\partial}{\partial y} w(x, y) = f(x) + x\left(\frac{d}{dy} g(y)\right) = q,$$

with p and q introduced as symbols to represent the partial derivatives.. Because elimination of $f(s)$, $g(y)$, $\frac{d}{dx} f(x)$ and $\frac{d}{dy} g(y)$ remains impracticable from these relations for p, q and $w(x, y)$, we proceed to evaluate the second partial derivatives:

$$\frac{\partial^2}{\partial x^2} w(x, y) = y\left(\frac{d^2}{dx^2} f(x)\right) = r, \quad \frac{\partial^2}{\partial y\, \partial x} w(x, y) = \left(\frac{d}{dx} f(x)\right) + \left(\frac{d}{dy} g(y)\right) = s \quad \text{and}$$

$$\frac{\partial^2}{\partial y^2} w(x, y) = x\left(\frac{d^2}{dy^2} g(y)\right) = t.$$

After algebraic manipulation, we obtain a partial-differential equation

$$x\, y\left(\frac{\partial^2}{\partial y\, \partial x} w(x, y)\right) = x\frac{\partial}{\partial x} w(x, y) + y\frac{\partial}{\partial y} w(x, y) - w(x, y)$$

that contains none of $f(x)$, $g(y)$ and their derivatives.

As another example, taking $p = \frac{\partial}{\partial x} z$ and $q = \frac{\partial}{\partial y} z$, we form two partial-differential equations from $f(x, y, z, \alpha, \beta) = 0$, in terms of independent variables x, y and z and two parameters α and β, on forming partial derivatives with respect to x and to y to obtain $\left(\frac{\partial}{\partial x} f\right) + \left(\frac{\partial}{\partial z} f\right)\left(\frac{\partial}{\partial x} z\right) = 0$ or

$$\left(\frac{\partial}{\partial x} f\right) + \left(\frac{\partial}{\partial z} f\right) p = 0 \text{ and } \left(\frac{\partial}{\partial y} f\right) + \left(\frac{\partial}{\partial z} f\right)\left(\frac{\partial}{\partial y} z\right) = 0 \text{ or } \left(\frac{\partial}{\partial y} f\right) + \left(\frac{\partial}{\partial z} f\right) q = 0,$$

of which the latter forms are partial-differential equations of order unity.

For a differential equation of type $f(x, y)\left(\frac{\partial}{\partial x} w\right) + g(x, y)\left(\frac{\partial}{\partial y} w\right) = 0$, a geometric interpretation is that, at each point in plane xy, a vector $f(x, y)\,\underline{i} + g(x, y)\,\underline{j}$ is orthogonal to a vector $\underline{del}\, w$, in which $\underline{i}$ and $\underline{j}$ are unit vectors parallel to axes x and y and $\underline{del}$ is the differential vectorial operator $\underline{i}\frac{\partial}{\partial x} + \underline{j}\frac{\partial}{\partial y}$ for a gradient. For point P, we choose a number c so that P is in the graph $w(x, y) = c$, which makes $\underline{del}\, w$ is perpendicular to this graph at point P. The graph is thus tangent to vector $f(x, y)\,\underline{i} + g(x, y)\,\underline{j}$. According to this reasoning, to find solutions of a given

differential equation, we find first all curves with a property that, at each point, the tangent line is parallel to $f(x, y) \underline{i} + g(x, y) \underline{j}$; in this way, we determine a function w such that it is constant along each curve. For instance, to solve this differential equation,

$$3 \left(\frac{\partial}{\partial x} w(x, y) \right) + 5 \left(\frac{\partial}{\partial y} w(x, y) \right) = 0,$$

we observe that curves with tangent lines parallel to to $3\,\underline{i} + 5\,\underline{j}$ are straight lines of equation $5\,x - 3\,y = c$. A function $w(x, y)$ that is constant along each such line is of type $w(x, y) = f(5\,x - 3\,y)$ for some function $f(x, y)$. Such a differentiable function $f(x, y)$ thus constitutes a solution of the differential equation, as we verify:

$$\frac{\partial}{\partial x} w(x, y) = 5\,f(5\,x - 3\,y) \text{ and } \frac{\partial}{\partial y} w(x, y) = 3\,f(5\,x - 3\,y),$$

$$\text{so } 3 \left(\frac{\partial}{\partial x} w(x, y) \right) + 5 \left(\frac{\partial}{\partial y} w(x, y) \right) = 15\,f(5\,x - 3\,y) - 15\,f(5\,x - 3\,y) = 0.$$

Lagrange found a general solution to an equation involving two independent variables x and y and one dependent variable $w(x, y)$ of form

$$P(w(x, y), x, y) \left(\frac{\partial}{\partial x} w(x, y) \right) + Q(w(x, y), x, y) \left(\frac{\partial}{\partial y} w(x, y) \right) = R(w(x, y), x, y)$$

with a reduction to an auxiliary system

$$\frac{dx}{P(w(x, y), x, y)} = \frac{dy}{Q(w(x, y), x, y)} = \frac{dw}{R(w(x, y), x, y)}$$

of ordinary-differential equations by demonstrating that $\phi(u, v) = 0$, in which $\phi(u, v)$ is an arbitrary function, is a general solution provided that $u(x, y, w)$ and $v(x, y, w)$ equal arbitrary constants. For example, for

$$x \left(\frac{\partial}{\partial x} w(x, y) \right) + y \left(\frac{\partial}{\partial y} w(x, y) \right) = 3\,w(x, y)$$

the auxiliary system is

$$\frac{dx}{x} = \frac{dy}{y} = \frac{dw}{3\,w}$$

From $\dfrac{dx}{x} = \dfrac{dw}{3\,w}$ we derive $u(x, y, w) = \dfrac{w}{x^3} =$ arbitrary constant a, and from $\dfrac{dx}{x} = \dfrac{dy}{y}$ we derive

$v = \dfrac{y}{x} =$ arbitrary constant b, so obtaining, as general solution, arbitrary function $\phi \left(\dfrac{w}{x^3}, \dfrac{y}{x} \right) = 0.$

With an alternative selection of equalities, we derive also $\chi \left(\dfrac{w}{x^3}, \dfrac{w}{y^3} \right) = 0$ and $\eta \left(\dfrac{w}{y^3}, \dfrac{y}{x} \right) = 0$, any of

which is equivalent to the other two and acceptable as a general solution.

Among partial-differential equations of great importance in chemistry and physics are these linear partial-differential equations of second order, to which allusion is made above with derivatives in an explicit form,

$$\text{diffusion equation } \kappa\,\underline{del}^{\,2}\,w(x, y, z, t) = \frac{\partial}{\partial t} w(x, y, z, t)$$

in three spatial dimensions in cartesian coordinates x, y, z and time t with concentration w and diffusion coefficient κ,

$$\text{wave equation } c^2\underline{del}^{\;2}\, w(x, y, z, t) \;=\; \frac{\partial^2}{\partial t^2}\, w(x, y, z, t)$$

with speed c of propagation of the waves of amplitude w of whatever type, and

$$\text{Laplace's equation } \underline{del}^{\;2}\, w(x, y, z) = 0,$$

all containing laplacian operator

$$\underline{del}^{\;2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

likewise in cartesian coordinates. Solutions of the diffusion equation are explained in sections 7.404 and 7.405, and of the wave equation in section 7.403, in restricted spatial dimensions. Schrodinger's temporally dependent equation in wave mechanics has the form of a diffusion equation; instances of its solution appear in chapter 12 in part II of this book. Laplace's equation has evidently a solution $w(x, y, z) = A\,x + B\,y + C\,z + \alpha$ for arbitrary values of constants A, B, C, α ; a particular solution might be expressed in a form of product $X(x)\,Y(y)\,Z(z)$, in which each multiplicand is a function of only the one specified coordinate variable. The latter form, explicitly a product of functions of separate variables, is a standard fixture of an attack on a problem requiring the solution of a linear partial-differential equation. The general solution of a non-homogeneous problem is a sum of any particular solution of the entire differential equation plus the general solution of the corresponding homogeneous equation provided that both the equation and the boundary conditions are homogeneous.

Of two important methods to obtain algebraic solutions of partial-differential equations, one is the separation of variables, which effectively converts the partial-differential equation into multiple ordinary-differential equations that are solved independently and of which the solutions are combined in a solution of the partial-differential equation. In some cases the solutions depend on the value of a parameter called an eigenvalue, in which case the solutions are called eigenfunctions; if that parameter appear in a sine or cosine formula, it might be called an eigenfrequency. Another method involves the use of integral transforms; a partial-differential equation in n independent variables is thereby to an equation that has only $n - 1$ independent variables.. For instance, such a reduction of a partial-differential equation with two independent variables yields an ordinary-differential equation.

A transformation of variables is a prospective approach to solve a partial-differential equation, and Fourier series, Fourier transforms, Laplace transforms and Green's functions play important roles in the solution of various partial-differential equations with particular initial values and boundary conditions, as explained by Y. Abe in *Essentials of Partial-differential Equations in Mathematical Physics*. Schrodinger's equation dependent on time has the same form as the diffusion equation above, whereas Schrodinger's equation independent of time resembles that for a standing wave; these equations arise in wave mechanics, which is one particular form of quantum mechanics that constitutes a collection of mathematical methods to treat problems in which the discreteness of various physical quantities, such as energy or angular momentum, is a distinguishing feature. The chemical applications of quantum mechanics are discussed in chapter

12 and elsewhere in part II of this book.

The methods of solution of partial-differential equations are in principle similar to those for ordinary-differential equations, but the increased number of independent variables greatly increases the difficulty of finding a solution. Efficient numerical methods have been devised involving finite differences or finite elements or boundary elements, but solutions in even two or three spatial dimensions typically require a mesh of many thousands of points, demanding efficient numerical processors, and might generally exceed the capabilities of a general algebraic processor such as *Maple*.

An advantage of the finite-element method over the finite-difference method is the ease with which boundary conditions are handled; for such conditions involving derivatives and irregularly shaped boundaries with the latter method, each boundary condition involving a derivative must be approximated with a difference quotient at the grid points; a boundary of irregular shape makes difficult the placing of the grid points. Because the finite-element method includes the boundary conditions as integrals in a functional that is subject to minimization, the construction procedure is independent of the particular boundary conditions.

integral equation

An equation that serves to determine a formula or expression constitutes an integral equation if the required formula occur in an integrand of an integral. Integral equations are complementary to differential equations in that for some chemical or physical problems, such as diffusion and transport phenomena, a representation by means of a differential equation is impracticable, but an integral equation is practicable; in other cases, a formulation as an integral equation might be convenient. Whereas a differential equation to treat some chemical or physical system is typically subject to particular initial values or boundary conditions, an integral equation relates an unknown formula or function not only to its values at neighbouring points, through derivatives, but also to its values throughout a region, including its boundaries; such boundary conditions are hence intrinsic within an integral equation rather than becoming imposed at an ultimate stage of solution. Because mathematical ramifications such as existence, uniqueness and completeness might be managed more readily in integral form than in differential form, and because some physical problems are not readily expressible as differential equations, integral equations comprise part of an armoury of mathematical techniques that a student of chemistry should encounter and understand. Like partial-differential equations, many integral equations that one encounters in practice lack an algebraic solution and are thus amenable to only numerical solution.

A linear differential equation of first order, $\dfrac{d}{dx} y(x) = f(x, y(x))$, is readily convertible to an integral equation,

$$\int_{x_0}^{x} f(x, y(x)) \, dx = y(x) - f(x_0)$$

for which $y(x)$ is known and $f(x, y(x))$ unknown. Even a linear differential equation of second order can be converted into an integral equation, involving integration by parts one time or more as follows in this instance in which $y(x)$ might be an arbitrary function of x and y but must include

no derivative $\dfrac{d}{dx} y(x)$, thus pertaining to both linear and non-linear differential equations of second order in a large class.

$$\frac{d^2}{dx^2} y(x) = f(x, y(x))$$

$$\frac{d}{dx} y(x) = \int_0^x f(x', y(x')) \, dx' + c_1$$

$$y(x) = \int_0^x dx'' \int_0^{x''} f(x', y(x')) \, dx' + c_1 x + c_2$$

As long as we maintain constant the region in plane $x' x''$ over which the double integration is performed, the order of integrations is reversible; altering the bounds of integration accordingly, we obtain

$$y(x) = \int_0^x f(x', y(x')) \, dx' \int_{x'}^x dx'' + c_1 x + c_2$$

and then combine the two integrals,

$$y(x) = \int_0^x (x - x') f(x', y(x')) \, dx' + c_1 x + c_2$$

into a non-linear integral equation of Volterra equation of second type because a bound of integration comprises a variable, as explained below. Such initial values as $y(x)\big|_{x=0} = a$ and $\left(\dfrac{d}{dx} y(x) \right)\bigg|_{x=0} = b$ would set $c_2 = a$ and $c_1 = b$.

   A transformation from a differential equation bestows benefits such as that the existence of a solution is more readily ascertained and that an integral equation can incorporate automatically any boundary conditions applicable to a solution. From a numerical point of view, a linear integral equation, in which an unknown function involving a particular variable occurs separately from other variables, is tractable as a system of linear equations with uncountable unknowns; eigenvalues and eigenfunctions play a significant role in the theory and practice of integral equations. Expansions in series, such as Fourier series or polynomials, are important aids in solution of integral equations. A linear integral equation, in which unknown formula or function $f(y)$ occurs to only the first power in the integrand, or a corresponding expression $f(x)$ outside the integrand analogously, is classifiable into these three types,

$$\int_a^x k(x, y) \, f(y) \, dy = g(x)$$

$$f(x) + \lambda \int_a^x k(x, y) \, f(y) \, dy = g(x)$$

$$h(x) \, f(x) + \lambda \int_a^x k(x, y) \, f(y) \, dy = g(x)$$

in which λ is a constant parameter; a multiplicand $k(x, y)$ of $f(y)$ within an integrand here is called a <u>kernel</u> of an integral equation. The former two integral equations arise as special cases of the latter with $h(x)$ equal to zero or a constant; the latter equation involving a single unknown function $f(x)$ becomes a *general linear integral equation*. If in any case $g(x) = 0$, the equation is <u>homogeneous</u>. An algebraic or symbolic approach to solution of such equations is applicable under certain conditions, such as those presented below. To warrant a descriptor linear, we regard

kernel $k(x, y)$ as an operator k such that $k\ f = \displaystyle\int_a^b k(x, y)\, f(y)\, dy$, analogous to linear differential

operator L_n defined above; just as eigenvalues and eigenfunctions play a role in the solution of those differential equations, constructs of linear algebra are applicable in the solution of integral equations. One strategy is applicable only when the kernel, such as $k(x, y)$, is separable into

functions of each variable, such as, in a <u>degenerate</u> case, into $k(x, y) = \displaystyle\sum_{j=1}^{n} P_j(x)\, Q_j(y)$;

Neumann's approach develops an unknown function $f(y)$ as a power series in λ, which is formally a constant.

Equations of these three types are further classified into linear integral equations of kinds that *Maple* recognizes, in which $f(x)$ or $f(y)$ is an unknown formula or expression. For Fredholm's equations, end points of integration are constants,

- Fredholm equation of first kind: $\lambda \displaystyle\int_a^b k(x, y)\, f(y)\, dy\ = g(x)$

- Fredholm equation of second kind: $f(x)\ +\ \lambda \displaystyle\int_a^b k(x, y)\, f(y)\, dy\ =\ g(x)$

- Fredholm equation of third and homogeneous kind: $f(x)\ +\ \lambda \displaystyle\int_a^b k(x, y)\, f(y)\, dy\ =\ 0$

- Fredholm equation of fourth kind: $h(x)\, f(x)\ +\ \lambda \displaystyle\int_a^b k(x, y)\, f(y)\, dy\ =\ g(x)$

whereas for Volterra's equations one end point is a variable.

- Volterra equation of first kind: $\lambda \displaystyle\int_a^x k(x, y)\, f(y)\, dy\ =\ g(x)$

- Volterra equation of second kind: $f(x)\ +\ \lambda \displaystyle\int_a^x k(x, y)\, f(y)\, dy\ =\ g(x)$

- Volterra equation of third and homogeneous kind: $f(x)\ +\ \lambda \displaystyle\int_a^x k(x, y)\, f(y)\, dy\ =\ 0$

- Volterra equation of fourth kind: $h(x)\, f(x) \;+\; \lambda \displaystyle\int_a^x k(x,y)\, f(y)\, dy \;=\; g(x)$

If kernel $k(x,y) = 0$ for $y > x$, a Fredholm equation reduces to the corresponding Volterra equation of the same kind. If an integral equation belong explicitly to no type among these -- for instance if an upper end point be x^2 instead of x, *Maple* might generate an erroneous solution, if any; verification of a prospective solution is invariably advisable on substitution of the result into the original integral equation and subsequent evaluation. Likewise for either Fredholm or Volterra equations of the fourth kind, in which an additional function of independent variable multiplies the unknown function outside the integral, *Maple* is unlikely to yield an answer. A numerical solution of an integral equation, involving the replacing of that integral equation by simultaneous algebraic equations in a set to which matrix techniques are applied, works well for Fredholm's equation of first kind, but poorly for Fredholm's equation of second kind. If an integral in a Fredholm equation become replaced with a more general form $\displaystyle\int_a^b k(x,y,\, f(x))\, dy$, such as $\displaystyle\int_a^b k(x,y)\, f(y)^2\, dy$

or $\displaystyle\int_a^b k(x,y)\, \cos(f(y))\, dy$, the equation is considered non-linear, and likely unsolvable directly with *Maple*.

For example, for this integral equation,

$$f(x) - \lambda \int_0^1 (x\,y^2 + x^2\, y)\, f(y)\, dy = x^2$$

in which the kernel is
```
> k(x,y) = - (x*y^2 + x^2*y);
```
we define

$$A = \int_0^1 y^2\, f(y)\, dy \qquad B = \int_0^1 y\, f(y)\, dy$$

so that the equation above becomes
```
> eq := f(x) =  x^2 - lambda*A*x - lambda*B*x^2;
```
On substituting this formula $f(x)$ for $f(y)$ into the defining equations above we obtain
```
> fA := A = Int(y^2*subs(x=y, rhs(eq)), y=0..1);
  fA := A = int(y^2*subs(x=y, rhs(eq)), y=0..1);
> fB := B =  Int(y*subs(x=y, rhs(eq)), y=0..1);
  fB := B =  int(y*subs(x=y, rhs(eq)), y=0..1);
```
We solve these two simultaneous linear equations for A and B,
```
> solve({fA,fB},{A,B});
```
assign the solution,
```
> assign(%);
```
and evaluate the original equation.
```
> simplify(eval(eq));
```
For these two values of λ,

```
> lambda = solve(denom(rhs(%)), lambda);
>
```

the value of f(x) becomes undefined; these values of lambda are called *eigenvalues* of the integral equation: the homogeneous equation, for which the term x^2 is absent, has non-trivial solutions only if λ be one or other eigenvalue; these solutions become the eigenfunctions of kernel $k(x, y) = -x\,y^2 - x^2\,y$. If the kernel be degenerate so that a procedure like that above becomes applicable, the solution of an integral equation becomes reduced to a solution of algebraic equations, which might be readily effected. On observing that a reasonably well behaved kernel is expressible as an infinite series of degenerate kernels, Fredholm deduced the following conditions that are applicable to real kernels.

- Either the inhomogeneous equation

$$f(x) \; + \; \lambda \int_a^x k(x, y)\, f(y)\, dy \; = \; g(x)$$

has a unique solution for an arbitrary function g(x), such that λ be not an eigenvalue, or the homogeneous equation, in which g(x) = 0, has at least one non-trivial solution, for which λ is an eigenvalue and the solution is an eigenfunction.

- If λ be not an eigenvalue, λ is also not an eigenvalue of the equation with a 'transposed' kernel,

$$f(x) \; + \; \lambda \int_a^x k(y, x)\, f(y)\, dy \; = \; g(x)$$

whereas if λ be an eigenvalue, λ is also an eigenvalue of the transposed equation, such that the transposed homogeneous equation,

$$f(x) \; + \; \lambda \int_a^x k(y, x)\, f(y)\, dy \; = \; 0$$

has at least one non-trivial solution.

- If λ be an eigenvalue, the inhomogeneous equation, as stated in the former condition above, has a solution if and only if

$$\int_a^b \alpha(x)\, g(x)\, dx = 0$$

for each function $\alpha(x)$ that obeys the transposed homogeneous equation just above.

If one express the general linear integral equation in symbolic form $h\,g + \lambda\,k\,f = g$, in which k denotes an operator that implies multiplication by a kernel k(x, y) and integration over y between bounds a and b, such a form is directly comparable with operator equations involving matrix or differential operators.

In some cases, a solution of integral equations might be effected with integral transforms, in particular those of Fourier, Laplace, Mellin and Hankel, but more general approaches include solution in series, following Neumann, Liouville and Volterra, and numerical solution involving conversion of an integral equation into simultaneous algebraic solutions in a set, which invokes matrix operations. Laplace transforms arise in a solution of differential equations in section 7.206

and Fourier transforms in section 7.207; Fourier transforms are discussed at length in chapter 14 with important applications to molecular and crystalline structure and to molecular spectra.

All these integral equations, and others, are analogues to equations involving matrices and vectors, and these integral equations have applications in a solution of problems in linear algebra.

Differential equations serve to model situations in chemistry, biology, physics, engineering, economics and medicine, and those of first order have extensive applications in chemical kinetics; for various kinetic orders, we explore these solutions at some length in section group 7.3. Methods algebraic and numeric to solve differential and integral equations with *Maple* we consider in other succeeding sections within this chapter. Partial-differential equations, with applications in physical chemistry and chemical physics, we discuss in further sections of this chapter. Further applications that require solution of differential or integral equations appear in chapters in part II, which includes a discussion of Fourier transforms and their applications in several important chemical experiments.

references *Heun's Differential Equations*, editor A. Ronveaux, Oxford University Press, Oxford UK, 1995; *Special Functions -- a Unified Theory Based on Singularities*, S. Y. Slavyanov and W. Lay, Oxford University Press, Oxford UK, 2000

>

[-] **summary of chapter 7**

A solution of a differential equation or an integral equation to yield an algebraic or symbolic formula or function is in general difficult, for the same reason as an expression in one or several variables might be difficult to integrate symbolically; for just such reasons, many special functions have arisen, such as Bessel functions introduced in section 7.102. According to the relative ease of solving mathematical quantities, conversion of an integral equation into a differential equation is preferable; in turn, conversion of a partial-differential equation into multiple ordinary-differential equations facilitates the solution, and conversion of a differential equation into an integral is generally desirable, because methods of treating quantities of simpler types are generally further developed and more reliable. One can naturally adopt a numerical approach when an algebraic result is elusive, and *Maple* includes many methods for this purpose. As differential equations and, to a lesser extent, integral equations are an active area of research, one expects that new algorithms and methods under current development will become implemented in forthcoming versions of software.

chapter 8 Probability, statistics, regression and optimization

[-] **8.0 overview and principles**

Although undoubtedly a physicist who anyhow became Nobel laureate for chemistry for achieving a transmutation of chemical elements, Lord Ernest Rutherford is reported to have asserted,

If your experiment needs statistics, *then you ought to have done a better experiment*. Like the quotation of Auguste Comte that appears in the Preface, this idea appears simplistic, much as a practitioner of chemistry might wish to the contrary, but an appropriate design of an

experiment might simplify greatly the subsequent analysis that must incorporate a proper statistical component. As a scientific field of inquiry, mathematical statistics originated largely through the work initiated by Legendre and became much extended by Gauss in the early nineteenth century, but most significant advances occurred during the early twentieth century, many from the work of Sir Ronald A. Fisher, who studied and was professor in Cambridge and is called "the father of modern statistics"; with a computer as a powerful tool to treat numeric data in a comprehensive and competent fashion, a chemist must not shirk his duty to treat all numeric data in a proper statistical manner, at least associating explicitly an uncertainty with each and every numerical datum or result or derived parameter: for many tasks in which a professional chemist is involved, such a statistical treatment is indeed a legal requirement. Almost invariably, the data available to a chemist are fewer than all possible data that would represent a total population or universe, so that the available data constitute merely a sample that is a subset of that population an alternative meaning of sample for a chemist, some material collected in a flask, is generally irrelevant in this chapter. A statistic is a value derived from a sample; such a statistic might serve as an estimate for a population, in which case it becomes an estimator. An unbiased estimator is a statistic of which the expected value equals the population parameter being estimated, and an efficient estimator has a smaller standard error than a less efficient estimator.   Beyond an actual use or treatment of numbers in a statistical evaluation of data, a chemist must, in the face of uncertainty, make inferences; an inference is a generalization based on incomplete or imperfect information, i.e. that from a sample rather than an entire population, that produces an estimate, prediction or decision, and that involves a probability rather than a certainty. Statistics constitutes the theory and procedures resulting therefrom that might be applied to numerical evidence for the purpose of an inference in the face of uncertainty. Statistical methods can serve in both experimental design and interpretation; the emphasis in this chapter is the latter, how to treat experimental data to reveal the maximal significance of the results of the experiments.

   A principal objective of this chapter is hence to present methods to treat numeric data, such as measurements resulting from chemical experiments, to yield both compact representations in a form of functional relation and significant chemical information through descriptive statistics. We conduct our experimental measurements with the expectation of deriving therefrom the maximum amount of useful information from the collected data. A practical scientific activity should have as an initial action the formulation of a hypothesis or multiple hypotheses about a defined scientific question or questions; on such a basis, the planning of experiments proceeds to ensure that any collected data are pertinent to the question, and that experimental variables are expected to influence the results of measurements. Experiments that inevitably involve an element of chance yield data that require a statistical treatment to warrant any inferences or conclusions therefrom; to measure or to assess a degree of uncertainty in drawing an inference from that statistical treatment of experimental data is a general task for which we here consider some concepts and procedures. A degree of uncertainty becomes subject to quantitative consideration on applying a concept probability. We proceed from a point of view of a chemist or physicist rather than that of a mathematician: the mathematician knows the parameters and the nature of the population, whereas, for an experimental scientist, an objective of a statistical analysis is to obtain values of pertinent parameters to describe adequately a population. A population is a collection of objects that have at least one common attribute or characteristice. Without a knowledge of an

entire population, one can not describe quantitatively a distribution of that population, but might express such a description in terms of probabilities. The most reliable results that one might derive from data in a particular set are those for which the estimated errors are the least; the techniques of analysis of data must hence include techniques of analysis of error: even the best efforts yield only estimates of the quantities investigated.

A prerequisite of the production of significant data in a chemical experiment is the planning of its conduct so that data collected therefrom are both sensitive -- to variables expected to be included in an objective formula or function that might serve to fit such data, and insensitive -- to extraneous factors, so that those data are optimally representative of a particular phenomenon that is being tested. An ensuing treatment typically involves a description of those data according to their fit to a functional form, or formula, based either on a pre-existing theoretical model or purely empirically in a convenient form, and eventually drawing inferences from those data. Measurements within a chemical experiment suffer inevitably from inaccuracy or error: a meaningful representation of those measurements must convey to a reader an indication, according to a standard criterion, of an extent of that inaccuracy and its propagation into derived parameters in some functional form, implying statistical considerations; a repeated sampling of observations on a continuous scale that yields results with some scatter of values implies an inevitable lack of precision such that a result is not exactly reproducible. Of errors of three types, gross error might result from a failure of an instrument or a blunder in its use, systematic error arises from imperfection in a procedure that yields a bias in the data from either a constant offset or proportionality, and random error that causes results to be spread from, or scattered about, a mean value; an experimental procedure must be designed to be accurate through the avoidance of systematic error and precise through the avoidance of random error. The repeatability of a result is a measure of the precision of measurements in a set made in the same laboratory with the same procedure; the reproducibility of a result is a measure of the precision of measurements in sets made in separate laboratories or with separate methods.

Experimental measurements suffer inevitably also from their necessarily limited number: for this reason, a conclusion from such evidence is hence an inference -- a generalisation from incomplete information -- rather than a deduction. As a result of a sampling error, a sample of measurements might exhibit a bias that constitutes a systematic tendency to misrepresent the entire population, as a sample is a subset of that population that has at least one common characteristic or attribute; a non-sampling error results solely from the manner of undertaking an observation -- a recording of information about some characteristic of an object, such as with a systematically malfunctioning mass balance. A measurement of an extent, intensity or size of a characteristic of an object constitutes such an observation. A sample space includes as a set all possible outcomes of an experiment, or all measurements in a process that generates a datum. A random variable from that sample space assumes either only discrete -- isolated or in a finite set resulting from counting -- or continuous values resulting from measurements necessarily imperfect, but in either case functions to assign a real value to each outcome in a sample space. A discrete variable has countable values, in a fixed set, and measurements involving the values of that variable between those countable values are impracticable; a continuous variable is measurable on a continuous scale, of which a result depends on the precision of the measuring instrument or on the accuracy of the observer. The four stages of a statistical exercise comprise a collection of data by means of

counting or measurement, an ordering, presentation and classification of those data in a convenient form, an analysis of those data according to statistical methods, and an interpretation of the results of that analysis and a formulation of conclusions. A random sample is designed to avoid an interference of shared properties, and thus allows for an equal probability of each subject of an observation to achieve a freedom from a sampling bias; a random sample might be *simple*, on choosing subjects such that each unit in a population might equally be selected, or *systematic*, on choosing a subject on a random basis and then further subjects at evenly spaced intervals, or *stratified*, on selecting independently a separate simple random sample from each stratum of population. The cumulative frequency is a sum of frequencies of all values less than, or equal to, a particular value.

Of an active group of British statisticians -- with Pearson and Galton, Fisher defined statistics to imply *reduction of data*; a broader definition entails quantitative data on any subject, the classification and interpretation of those data in accordance with a theory of probability and an application of methods to test hypotheses -- hence a mathematical treatment of a theory of such distributions and tests. The concerns of statistics are hence the collection, ordering and analysis of data, which comprise recorded observations or values in ordered sets from counts or measurements, and a consequent interpretation of those data.   A quantity that can assume diverse values is a variable, which adopts either discrete or continuous values. A discrete value arises from a count, or as one value in a fixed and finite set, whereas, for a variable that is measured on a continuous or pseudo-continuous scale, the result depends on either the precision of a measuring instrument or the accuracy of an observer. Many instruments that were formerly common in chemical laboratories presented abundant data in analogue form as a continuous curve, such as a spectrum printed on a paper chart, whether from nuclear magnetic resonance at radio frequencies or from optical absorption in the ultraviolet region, or from a chronopotentiometer or gas chromatograph, among many other possibilities; an analysis of such data with a computer requires prior conversion of those curves to digital form. Most contemporarily manufactured instruments for chemical analysis and other measurements transfer data directly to a computer, even if those instruments are not controlled directly through a computer. Just as *econometrics* evolved from various attempts to describe economic phenomena in a quantitative manner, *chemometrics* has evolved for an analogous purpose in relation to chemistry, with three principal thrusts:

- the control of instruments and experimental measurements thereon,

- a transfer and analysis, especially including statistical aspects, of data from those instruments and measurements, and

- a simulation of experimental data in the form of curves representing spectra or other instrumental observations in analogue form.

In this chapter, our concern is the second aspect, namely quantitative analysis of chemical data according to statistical concepts to yield conclusions of maximal significance; in chapters in part II we present examples of simulation of chemical and physical measurements and phenomena.

The reproducibility or repeatability of observations is an essential property of scientific research and chemical practice, whether in a teaching laboratory or for industrial testing. That repeatability implies that a particular experiment performed on similar systems should yield the

same result, apart from chaotic systems that anyhow have consistent aspects. A measurement of a chemical or physical quantity typically yields slightly varied values for similar systems; only the consistent part of those values is significant. For several measurements that appear, a priori, to be equally valid, the best estimate of the result is the mean value; the extent of the deviations from that mean value, as a distribution of a particular width, and the number of measurements upon which that mean is based enable an estimate of the precision of the result. *Accuracy* is a qualitative concept that denotes the proximity of agreement between a result of a measurement and a true value of a measurand, or particular quantity subject to measurement; a measure of accuracy might be a number of significant digits or decimal places, or a range of possible error stated in absolute or relative terms. In contrast, *precision* implies the proximity of an agreement between the results of independent tests obtained under stipulated conditions; a precision depends on only a distribution of random errors, and does not relate to a true or specified value; such random errors tend to have a nearly symmetric distribution, such that approximately as many results exceed the mean as are less than the mean. Systematic error, which arises from a miscalibration of an instrument involved in a measurement or from an ignored influence, affects not the precision but the accuracy, and tends to bias the result toward a mean of measurements that is greater or smaller than the true value. Whereas random errors determine the precision of a result, systematic errors determine its accuracy; to recognise and to control systematic error require ingenuity and diligence.

Precision is commonly expressed in terms of imprecision or uncertainty, calculated as a standard deviation of results of multiple tests. The results of tests are independent when they are obtained in a manner free from an influence of other results on the same or similar object of test; quantitative measures of a precision depend critically on the stipulated conditions, of which repeatability, when measurements are repeated under the same conditions of apparatus and operator, and reproducibility, when measurements are repeated in other laboratories and by other operators, are particular sets; reproducibility is typically poorer than repeatability. By statistical methods only deviations are demonstrable: similarities must be inferred from their absence. In a context of numerical calculations on a computer, precision indicates a number of decimal digits with which a calculation is performed, depending ultimately on the computer's processor, such as *single precision* typically implying 6 -- 8 decimal digits or double precision implying 16 -- 18 decimal digits, which obviously differs from precision in a context of uncertainty attached to a particular datum or result of an experiment or a theoretical calculation; in a context of working with *Maple*, the default setting `Digits := 10;` might be considered *single precision*, or somewhat better than single precision, whereas use of `evalhf` yielding 16 -- 18 decimal digits might be considered *double precision*, and an appreciably greater setting of `Digits` would achieve multiple precision to a desired extent.

A general dictum of science is that

a quantity specified without an estimate of its reliability, or its uncertainty, is worthless. In this context a practical definition of uncertainty is a parameter associated with a result of a measurement that characterizes a dispersion, or spread or scatter, of values that one can reasonably attribute to a quantity being measured, or *measurand*. Rather than the uncertainty of a measurement implying a doubt about its validity, a knowledge of that uncertainty implies an increased confidence in such validity. If one make only a single measurement of some quantity,

one must apply one's knowledge of experimental conditions as a basis of estimating an uncertainty; one's trust of such an uncertainty is greatly enhanced when one can apply a standard statistical treatment to the results of multiple measurements, so as to derive a *standard deviation* that constitutes a *standard uncertainty*. Because probability as a notion logically underlies statistics, we begin with rudimentary definitions of probability of an event *a priori* in relation to permutations and combinations of possible outcomes.

probability

An outcome of an event in a chemical context, like any other event, is subject to chance, to some extent; an experiment repeated under essentially the same conditions yields not precisely, but merely approximately to a greater or lesser extent, the same or identical results. According to Laplace's theory of chance, one reduces all events of a particular kind to cases of a certain number equally possible, or such as about which we might be equally undecided with regard to their existence, and determines the number of cases favourable to the event of which one seeks a probability. The ratio of this number to that of all possible cases is a measure of this probability, which is thus simply a fraction of which the numerator is the number of favourable cases and the denominator is the number of all possible cases. A subjective element enters the assessment of equiprobability. A probability is hence an estimate or a measure of a degree of confidence that one might have in an occurrence of an event in a particular trial or experiment, measured on a scale from zero -- impossibility -- to unity -- certainty. Each separate possible result of a trial is an outcome. According to this classical or theoretical probability, a measure of whether a particular outcome of an event occurs on an occasion of a trial is hence a probability: probability p_1 of a particular and fully characterized outcome of a single trial one estimates according to a ratio of number of trials that favour that outcome, or successful outcomes, to a total number of trials,

$$p_1 = \frac{\textit{number of trials yielding a particular outcome}}{\textit{total number of trials}}$$

The use of probabilities to evaluate the chance of a particular result of a trial is known as deductive or frequentist statistics. A chemical application of the frequentist approach that is impracticable with the classical approach involves an estimate of the probability that a molecule of a particular compound under particular conditions of a sample of that compound decomposes within a given period; this estimate might be based on the known rate of decomposition under the same conditions. According to an empirical or frequentist or objective probability, for a random experiment performed on numerous occasions numbering n that yields a particular outcome on n_1 occasions, the proportion of occasions on which outcome n_1 occurs tends to the probability p_1 of that event consistent with the law of large numbers:

$$\frac{n_1}{n} \; \text{-->} \; p_1 \quad \text{as } n \to \infty$$

This formula is valid provided that each possible distinct outcome of a trial is equally likely, and that events are uncorrelated: an outcome of one trial has no effect on an outcome of another trial. Called a statistical probability, this ratio is, for each outcome, a positive number that lies necessarily in a domain [0, 1] with certain failure and certain success as respective limiting cases. An impossible outcome implies zero probability and a certain outcome unit probability. The total probability of all outcomes of an event is unity. A proportion of a particular condition to a total

number of items in a sample is an empirical probability, which is formally a limit of this proportion as the size of a sample becomes indefinitely large; a smaller sample is subject to bias. From a practical point of view, this frequentist approach provides no probability of a particular outcome when the repetition of the experiment many times is physically impossible; to decide how many trials are required to produce a satisfactory estimate of the probability of an outcome is also difficult. Moreover, a frequentist interpretation of the probability of an event that occurs only once under the same conditions, as commonly encountered in econometrics or for natural disasters, is impracticable. The prospective confusion of an abstract mathematical object -- a theoretical probability -- with frequencies of events obtained directly from experiments is a significant criticism of this frequentist definition of probability. We define a theoretical or classical probability p of a particular outcome or event as

$$p = \frac{\textit{number of ways in which a particular outcome can arise}}{\textit{total number of all possible outcomes}}$$

Two contrasting approaches to a treatment of statistical data -- which are truly any data obtained in a chemical context -- are deductive and inductive statistics. Deductive statistics is applicable when a theoretical basis is known or anticipated; a result of a particular event has a probability calculable on that theoretical basis. Inductive statistics is applicable to the use of information about a sample to estimate the probability of an event in a population. These two approaches are entirely complementary: an assessment of samples requires an investigation of its generation before a generalization is practicable, but inductive statistics requires an assumption of a theoretical basis from which deductions might be drawn.

A third approach is called a bayesian or subjective probability, according to which a probability is identified with a degree of a belief of a particular individual person; this approach is uncommonly applied in physical science.

A classical probability reflects a theoretically possible number of outcomes of a particular event. An empirical probability might also be based on previously obtained results; a relative frequency of a particular outcome for past events is taken as an indication of likely occurrences in future. According to that law of large numbers that is a fundamental statistical result, the mean of n independent, identically distributed random variables in a sequence tends to their mean as $n \to \infty$, or, rather, that the difference between theoretical and experimental probabilities tends to zero as the number of tests becomes indefinitely large; the relative frequency of occurrence of an event in n independent repetitions of a test tends to its probability as n increases without limit. π as the ratio of the circumference of a plane circle to its diameter is considered to be a normal number because its digits are equally probable.

A combined probability of a particular outcome of one trial and another specific outcome of a second uncorrelated or independent trial is a product of separate probabilities, because a probability with respect to separate trials is multiplicative. If we undertake $n - 2$ additional trials, probability p_n of that particular outcome of every uncorrelated trial is a probability p_1 in each trial raised to a power equal to number n of these trials;

$$p_n = p_1^{\,n}$$

If a probability of a successful outcome of an event be p_1, the probability of a corresponding

failure is $q = 1 - p_1$, because these outcomes are complementary. An expectation of a particular outcome of any one of n trials is the number of trials multiplied by both the probability of that outcome of a single trial and the probability of $n - 1$ failures.

$$expectation \ = \ n \, q^{(n-1)} \, p_1$$

In a case of a small probability of success, $p_1 \ll 1$, or $q \sim 1$, this expectation becomes approximately $n \, p_1$. An expectation of exactly r successes in n trials or independent events is

$$expectation \ = \ \frac{n!}{r! \, (n-r)!} \, p^r \, q^{(n-r)}$$

that includes a binomial coefficient, $_nC_r = \dfrac{n!}{r! \, (n-r)!}$. The probability of at least r successes in n trials is

$$p^n \ + \ _nC_1 \, p^{(n-1)} \, q + \ _nC_2 \, p^{(n-2)} \, q^2 + ... + \ _nC_r \, p^r \, q^{(n-r)}$$

which is a sum of the first $n - r + 1$ terms of the binomial expansion of $(p + q)^n$. In a Bernoulli population, each element has one of two possibilities, described as success and failure; the outcome of a Bernoulli trial is one element of a Bernoulli population. A *population proportion* is the proportion of successes, for which the sample proportion is a useful estimate if it be unbiased, if its standard error tend to zero with increasing size of population and if it be efficient through having a variance smaller than any other unbiased estimator of the population proportion.

Two events are independent if an outcome of one event has no effect on an outcome of a second event. For a particular experimental outcome x, an outcome other than that x, which might be denoted $\bar{x}$, is called the *complementary outcome*, or the *complement*, of x. Two or more outcomes are mutually exclusive if an occurrence of any one such outcome preclude the occurrence of each and any other outcome; mutually non-exclusive events occur simultaneously. The probability of occurrence of outcome x, given that outcome y has occurred, is called the conditional probability of outcome x given outcome y, and is denoted $P(x|y)$. An outcome x is independent of outcome y if a probability of occurrence of outcome x be unaffected by the prior occurrence of outcome y, and vice versa, so that $P(x|y) = P(x)$ and $P(y|x) = P(y)$. A collection of outcomes becomes exhaustive when that collection includes all possible outcomes of the experiment. For two independent outcomes of events, a multiplicative law is applicable,

$$p_{x \text{ and } y} = p_x \, p_y$$

with a correspondingly extended product for multiple independent events. If two events be mutually exclusive, such that an occurrence of one outcome of an event in a particular test precludes simultaneously an occurrence of another particular outcome, these probabilities are additive.

$$p_{x \text{ or } y} \ = \ p_x \ + \ p_y$$

For two events of which outcomes x and y are possible, the probability of *at least* one such outcome is

$$p_{x \text{ or } y} = p_x + p_y - p_{xy}$$

in which $p_{xy} = p_x \, p_y$. For three possible outcomes x, y, z of an event, the probability of at least one such outcome is

$$p_{x \text{ or } y \text{ or } z} = p_x + p_y + p_z - p_{xy} - p_{xz} - p_{yz} + p_{xyz}$$

The extents of dependence are variable, with mutually exclusive and certainly inclusive as limiting cases. A conditional probability applies to dependent events: for instance, if from n distinguishable objects one is randomly selected, the probability of a particular outcome to occur is $\dfrac{1}{n}$; the conditional probability for that same object to be selected from the remaining objects is 0. The conditional probability of an event x, given that event y has occurred, is calculated as the quotient of a probability that both events occur and a probabiltiy that y occurs. Events are independent when an occurrence of one event does not preclude another event, in which case the probability of two independent events is a product of the separate probabilities, whereas events are dependent when one event affects the probability of occurrence of a second event. For a Bernoulli trial, only two outcomes -- success and failure -- are possible, and they are complementary; the number of successes is a random variable conforming to a binomial distribution, according to the formula above for the expectation of a success. A random variable conforms to a rule that represents the prospective numerical values associated with the outcomes of an experiment; a list of those values constitutes the range of that variable. The expected value of a random variable is its average value, which balances the distribution.

For repeated trials, if the probability of an event in a particular trial be p and the probability of another and incompatible event be $q = 1 - p$, the probability of occurrence r times in n trials is

$$p_n(r) = \frac{n!}{r! \, (n - r)!} \, p^r \, q^{(n-r)}$$

as for the expectation above.

Although chemical transformations, from reactants to products, appear generally to proceed smoothly, they are actually stochastic events in which an event at each microscopic centre is subject to laws of chance, for instance within a specified temporal interval. In a stochastic process, an individual event might seem random, but in totality such events obey well defined probabilistic laws; a stochastic process thus involves the dynamics of probability functions, in which random variables enter probability distributions that depend on time. As a typical chemical sample contains numerous molecules, for instance about 6×10^{17} molecules even in a micromole, the fluctuations or deviations from a smooth course of reaction are difficultly detectable; the results of an exercise below confirm that such fluctuations from a mean are of order $\sqrt{n}$ when the expected number is n. When we can detect reactive events individually, as for instance hearing a click from a Geiger-Muller counter equipped with a loudspeaker in the case of a sample containing nuclides of a slowly decaying radioactive element, we should replace concentrations or numbers of reactants with their probabilities. A radioactive decay according to the first kinetic order, for which a relation

$$N = N_0 \, e^{(-k_1 t)}$$

with initial number N_0 of radioactive nuclei of half life $\dfrac{\ln(2)}{k_1}$ and number N remaining at duration t of reaction might be an adequate expression with N large and many reactive events during a

particular period of counting, becomes expressed, when N is small and few reactive events occur during a counting period, as

$$p = p_0 \, \mathrm{e}^{(-k_1 t)} \ ;$$

here appear probability p_0 of an event within a small interval at the initial stage of a reaction, or of observation of a reaction, and corresponding probability p within an equal interval at a subsequent stage after a duration t. Although N_0 and N must clearly be integers, no such restriction applies to p_0 and p. Stochastic laws apply to events with discrete outcomes, such as a making and breaking of chemical bonds; only when these events are rare, either because participating entities are few or because events are strongly inhibited, do stochastic effects become discernible. A stochastic process is describable with a random variable, called hence a stochastic variable, that depends on some parameter, which might be discrete or continuous, and which is commonly taken to be time.

combination and permutation

If n distinct possibilities occur for one condition and m distinct possibilities occur for a separate condition, the two conditions together have $m \, n$ possibilities; for instance, for a collection of three H atoms distinguished as H, D and T, and two Cl atoms, distinguished as ^{35}Cl and ^{37}Cl atoms, the number of possible HCl molecules distinguished according to the isotopic masses of the separate atoms is $3 \, (2) = 6$.

A combination or unordered arrangement implies selection of a subset of objects, of specified number, from a set of equal or greater number without regard to an order, whereas a permutation is an ordered arrangement of objects, of specified number, selected from a set of equal or greater number. A combination is an arrangement that contains elements of particular number r in an arbitrary order, chosen from n specified elements, but permutations are impermissible within this arrangement. A permutation of n elements is an arrangement in a possible ordering of those elements; an interchange of two of n distinct elements in a permutation effects a transposition, and any permutation of n distinct elements becomes transformed into another permutation. A number of ways to choose r distinguishable objects from n such objects in a set, thus with $r \leq n$ and with no repetition, is hence a number of distinct combinations, expressed with factorials,

$$\mathrm{C}(n, r) \ = \ \frac{n!}{r! \, (n - r)!} \ ,$$

which is a binomial coefficient; an alternative symbol is $_nC_r$; by definition, $\mathrm{C}(n, r) = 0$ for all $r > n$. The total number of combinations of n distinguishable objects taken 1, or 2, ... or n at a time is $2^n - 1$. The number of distinct permutations of r distinguishable objects is

$$\mathrm{P}(n, r) \ = \ \frac{n!}{(n - r)!} \ ,$$

which is just a ratio of factorials; an alternative symbol is $_nP_r$. For a permutation of n objects n at a time, the number is hence $n!$; any particular combination of r items is hence subject to rearrangement within itself to yield $n!$ distinct permutations. If a permutation be circular such that no particular initial point of an arrangement be recognisable, a number of circular permutations is equal to a number of linear permutations divided by a number of objects, or

$$\frac{P(n, r)}{n} = \frac{(n-1)!}{(n-r)!}$$

In a situation in which objects in k groups are involved, for which objects in each group are indistinguishable from each other but distinguishable from objects in any other group, the number of permutations of n objects, as a sum of n_1 objects in group 1, n_2 objects in group 2, ... n_k objects in group k, taken n at a time is

$$P(n; n_1, n_2, n_3, ..., n_k) = \frac{n!}{n_1! \, n_2! \, n_3! \, ... \, n_k!}$$

A permutation is even or odd according to a number of exchanges of elements that yield a particular result with reference to an original order.

The number of ways in which n distinguishable objects are divisible into k classes, with n_1 in class 1, n_2 in class 2, ..., with $\sum_i n_i = n$ is, like the number of permutations above,

$$\frac{n!}{n_1! \, n_2! \, ... \, n_k!}$$

The number of ways that m distinguishable objects might be placed into n identical compartments, with $n \geq m$, is $\dfrac{n!}{(n-m)!}$, because $n - m$ compartments, being empty, are indistinguishable among themselves. If these objects be not distinguishable, the number is $\dfrac{n!}{(n-m)! \, m!} = C(n, m)$, because the m occupied compartments are become indistinguishable among themselves; rearrangements of the order of occupied and empty compartments count as distinct ways. The number of ways of placing m distinguishable objects into n ordered compartments is n^m when any number from 0 to m might enter each compartment. The number of ways of placing m indistinguishable objects into n ordered compartments, with any number in each compartment, is

$$C(n + m - 1, m) = \frac{(n + m - 1)!}{m! \, (n-1)!} \ .$$

distributions and statistical indicators

Among many continuous distributions, we list the following [from Probability Distribution Relations, by Y. Abdelkader and Z. Al-Marzouk, *Statistica*, 70 (1), 41-51, 2010] with their parameters and domains, some of which are discussed further below.

$$f_\beta(x) := \frac{x^{(\alpha - 1)} (1 - x)^{(\beta - 1)}}{B(\alpha, \beta)} \ , \quad \text{parameters } 0 < \alpha, \ 0 < \beta, \text{ domain } [0,1]$$

$$f_{\beta'}(x) := \frac{x^{(\alpha - 1)} (1 + x)^{(-\alpha - \beta)}}{B(\alpha, \beta)} \ , \quad \text{parameters } \alpha > 0, \ \beta > 0, \text{ domain } [0, \infty]$$

$$f_{Cauchy}(x) := \frac{b}{\pi (x - \eta)^2 + b^2} \ , \quad \text{parameters } \eta \text{ real}, \ b > 0, \text{ domain } [-\infty, \infty]$$

$$f_\chi(x) = \frac{2^{\left(1 - \frac{1}{2}\, n\right)}\, x^{(n-1)}\, e^{\left(-\frac{1}{2}\, x^2\right)}}{\Gamma\!\left(\frac{1}{2}\, n\right)} \; , \quad \text{parameter } n > 0,\ \text{domain } [0,\infty]$$

$$f_{degenerate}(x, x_0) \; = \; \textbf{if } x = x_0 \textbf{ then } 1 \textbf{ else } 0 \textbf{ end if} \; , \quad \text{parameter } x_0 \text{ real, domain } [-\infty, \infty]$$

$$f_{Erland}(x) = \frac{x^{(n-1)}\, e^{\left(-\frac{x}{\alpha}\right)}}{(n-1)!\ \alpha^n} \; , \quad \text{parameters } n \text{ posint, } \alpha > 0,\ \text{domain } [0, \infty]$$

$$f_{exponential}(x) = \frac{e^{\left(-\frac{x}{\alpha}\right)}}{\alpha} \; , \quad \text{parameter } \alpha > 0,\ \text{domain } [0, \infty]$$

$$f_F(x) = \frac{n1^{\left(\frac{1}{2}\, n1\right)}\, n2^{\left(\frac{1}{2}\, n2\right)}\, x^{\left(\frac{1}{2}\, n2 - 1\right)}}{B\!\left(\frac{1}{2}\, n1, \frac{1}{2}\, n2\right)(n1 + n2\, x)^{\left(\frac{1}{2}\, n1 + \frac{1}{2}\, n2\right)}} \; , \quad \text{parameters } n1 > 0,\ n2 > 0,\ \text{domain } [0, \infty]$$

$$f_\Gamma(x) = \frac{x^{(\theta-1)}\, e^{\left(-\frac{x}{\alpha}\right)}}{\Gamma(\theta)\, \alpha^\theta} \; , \quad \text{parameters } \alpha > 0,\ \theta > 0,\ \text{domain } [0, \infty]$$

$$f_{Gibrat}(x) = \frac{e^{\left(-\frac{\ln(x)^2}{2}\right)}}{\sqrt{2\, \pi}\, x} \; , x > 0,\ \text{domain } [0, \infty]$$

$$f_{Gumbel}(x) = \frac{e^{\left(\frac{\alpha - x}{\beta} - e^{\left(\frac{\alpha - x}{\beta}\right)}\right)}}{\beta} \; , \quad \text{parameter } \beta > 0,\ \text{domain } [-\infty, \infty]$$

$$f_{half\,normal}(x) = \frac{2\, b\, e^{\left(-\frac{b^2\, x^2}{\pi}\right)}}{\pi} \; , \quad \text{parameter } b > 0,\ \text{domain } [0, \infty]$$

$$f_{inverse\,\chi^2}(x) = \frac{2^{\left(-\frac{1}{2}\, v\right)}\, x^{\left(-\frac{1}{2}\, v - 1\right)}\, e^{\left(-\frac{1}{2}\frac{(1)}{x}\right)}}{\Gamma\!\left(\frac{1}{2}\, v\right)} \; , \quad \text{parameter } v > 0,\ \text{domain } [0, \infty]$$

$$f_{inverse\,\Gamma}(x) = \frac{\theta^\alpha\, x^{(-\alpha-1)}\, e^{\left(-\frac{\theta}{x}\right)}}{\Gamma(\alpha)} \; , \quad \text{parameters } \alpha > 0,\ \theta > 0,\ \text{domain } [0, \infty]$$

$$f_{Kumaraswamy}(x) = a\, b\, x^{(a-1)}\, (1 - x^a)^{(b-1)} \; , \quad \text{parameters } a > 0,\ b > 0,\ \text{domain } [0, 1]$$

$$f_{Laplace}(x) = \frac{1 \, e^{\left(-\frac{|x-\mu|}{b}\right)}}{2 \, b} \quad , \text{ parameter } b > 0, \text{ domain } [-\infty, \infty]$$

$$f_{Levy}(x) = \frac{\sqrt{\frac{1 \, c}{2 \, \pi}} \, e^{\left(-\frac{1 \, c}{2 \, x}\right)}}{x^{\left(\frac{3}{2}\right)}} \quad , \text{ parameter } c > 0, \text{ domain } [0, \infty]$$

$$f_{logistic}(x) = \frac{e^{\left(-\frac{x-\mu}{b}\right)}}{b \left(1 + e^{\left(-\frac{x-\mu}{b}\right)}\right)^2} \quad , \text{ parameter } b > 0, \text{ domain } [-\infty, \infty]$$

$$f_{lognormal}(x) = \frac{e^{\left(-\frac{1 \, \ln(x-\mu)^2}{2 \, \sigma^2}\right)}}{\sqrt{2 \, \pi} \, \sigma \, x} \quad , \text{ parameter } \sigma > 0, \text{ domain } [0, \infty]$$

$$f_{Maxwell}(x) = \frac{\sqrt{\frac{2}{\pi}} \, x^2 \, e^{\left(-\frac{x^2}{2 \, a^2}\right)}}{a^3} \quad , \text{ parameter } a > 0, \text{ domain } [0, \infty]$$

$$f_{noncentral \, \chi^2}(x) = \frac{x^{\left(\frac{1 \, n}{2} - 1\right)} \, e^{\left(-\frac{1 \, x}{2} - \frac{1 \, \lambda}{2}\right)} \left(\sum_{k=0}^{\infty} \frac{(\lambda \, x)^k}{2^{(2 \, k)} \, k! \, \Gamma\left(k + \frac{1 \, n}{2}\right)}\right)}{2^{\left(\frac{1 \, n}{2}\right)}} \quad , \text{ parameters } n > 0, \lambda > 0,$$

$$\text{domain } [0, \infty]$$

$$f_{noncentral \, F}(x) =$$

$$\sum_{k=0}^{\infty} \left(\sum_{l=0}^{\infty} \frac{n1^{\left(\frac{1 \, k}{2} + \frac{1 \, n1}{2}\right)} n2^{\left(\frac{1 \, l}{2} + \frac{1 \, n2}{2}\right)} x^{\left(\frac{1 \, k}{2} + \frac{1 \, n1}{2} - 1\right)} \lambda1^k \, \lambda2^l \, (n2 + n1 \, x)^{\left(k + l - \frac{1 \, n1}{2} - \frac{1 \, n2}{2}\right)}}{2^{(k+l)} \, e^{\left(\frac{1 \, \lambda1}{2} + \frac{1 \, \lambda2}{2}\right)} B\left(k + \frac{1 \, n1}{2}, l + \frac{1 \, n2}{2}\right)}\right) , \text{ parameters}$$

$$n1 > 0, n2 > 0, \lambda1 > 0, \lambda2 > 0, \text{ domain } [0, \infty]$$

$$f_{noncentral \, Student's \, t}(x) =$$

$$\frac{n^{\left(\frac{1 \, n}{2}\right)} n! \, \sqrt{2} \, \lambda \, x \, F\left(1, 1, \frac{1 \, n}{2} + 1, \frac{3}{2}, \frac{\lambda^2 \, x^2}{2 \, n + 2 \, x^2}\right) F\left(1, 1, \frac{1 \, n}{2} + 1, \frac{1}{2}, \frac{\lambda^2 \, x^2}{2 \, n + 2 \, x^2}\right)}{2^n \, e^{\left(\frac{1 \, \lambda^2}{2}\right)} (n + x^2)^{\left(\frac{1 \, n}{2}\right)} \Gamma\left(\frac{1 \, n}{2}\right) (n + x^2) \, \Gamma\left(\frac{1 \, n}{2} + \frac{1}{2}\right) \sqrt{n + x^2} \, \Gamma\left(\frac{1 \, n}{2} + 1\right)} , \text{ parameters } n >$$

$$0, \lambda > 0, \text{ domain } [-\infty, \infty]$$

$$f_{normal}(x) = \frac{e^{\left(\frac{1\,(x-\mu)}{2\,\sigma^2}\right)}}{\sqrt{2\,\pi}\,\sigma}, \quad \text{parameter } \sigma > 0, \text{ domain } [-\infty, \infty]$$

$$f_{Pareto}(x) = \frac{k\,x_0^{\,k}}{x^{(k+1)}}, \quad \text{parameters } x_0 > 0, k > 0, \text{ domain } [x_0, \infty\,]$$

$$f_{PearsonIII}(x) = \frac{\left(\dfrac{x-\mu}{\theta}\right)^{(\alpha-1)} e^{\left(-\frac{x-\mu}{\theta}\right)}}{\Gamma(\alpha)\,\theta}, \quad \text{parameters } \alpha > 0, \theta > 0, \mu > 0, \text{ domain } [0, \infty]$$

$$f_{Rayleigh}(x) = \frac{x\,e^{\left(-\frac{1}{2}\frac{x^2}{\sigma^2}\right)}}{\sigma^2}, \quad \text{parameter } \sigma > 0, \text{ domain } [0, \infty]$$

$$f_{Rice}(x) = \frac{x\,e^{\left(-\frac{1\,(x^2+v^2)}{2\,\sigma^2}\right)} I_0\left(\dfrac{x\,v}{\sigma^2}\right)}{\sigma^2}, \quad \text{parameters } v > 0, \sigma > 0, \text{ domain } [0, \infty]$$

I_0 is a modified [Bessel](#) function of the first kind and order 0.

$$f_{standard\_normal}(x) = \frac{e^{\left(-\frac{1}{2}\frac{x^2}{2}\right)}}{\sqrt{2\,\pi}}, \quad \text{domain } [-\infty, \infty]$$

$$f_{Student's\;t}(x) = \frac{\Gamma\left(\dfrac{1\,n}{2} + 1\right)}{\sqrt{n\,\pi}\,\Gamma\left(\dfrac{1\,n}{2}\right)\left(1 + \dfrac{x^2}{n}\right)^{\left(\frac{1\,n}{2}+\frac{1}{2}\right)}}, \quad \text{parameter } n > 0, \text{ domain } [-\infty, \infty]$$

$$f_{triangular}(x) = \text{piecewise}\left(x < a,\, 0,\, x < c,\, \frac{2\,(x-a)}{(b-a)\,(c-a)},\, x < b,\, \frac{2\,(b-x)}{(b-a)\,(b-c)},\, 0\right)$$

$$f_{triangular} := \begin{cases} 0 & x < a \\ \dfrac{2\,(x-a)}{(b-a)\,(c-a)} & x < c \\ \dfrac{2\,(b-x)}{(b-a)\,(b-c)} & x < b \\ 0 & otherwise \end{cases}, \quad \text{parameters } a < c, c < b$$

$$f_{uniform} = \frac{1}{b-a}, \quad \text{parameters } a < b$$

$$f_{Wald}(x) = \sqrt{\frac{1\,\lambda}{2\,\pi\,x^3}}\;\mathrm{e}^{\left(-\frac{1\,\lambda\,(x-\mu)^2}{2\,\mu^2\,x}\right)}\,, \quad \text{parameters } \lambda > 0,\ \mu > 0,\ \text{domain } [0, \infty]$$

$$f_{Weibull}(x) = \frac{\beta\,x^{(\beta-1)}\;\mathrm{e}^{\left(-\left(\frac{x}{\alpha}\right)^\beta\right)}}{\alpha^\beta}\,, \quad \text{parameters } \alpha > 0,\ \beta > 0,\ \text{domain } [0, \infty]$$

Among many discrete distributions, we list the following with their parameters, domains and probability mass functions; some distributions are discussed further below. In probability theory and statistics, a probability mass function indicates a probability that a discrete random variable is exactly equal to some value; this probability mass function is commonly the primary means to define a discrete probability distribution, and such functions exist for either scalar or multivariate random variables, given that the distribution be discrete.

The Bernoulli distribution is a special case of a binomial distribution

$$P_{Bernoulli}(x) = p \text{ if } x = 1 \text{ or } q = 1 - p \text{ if } x = 0\,, \quad \text{parameters } 0 \le p \le 1,\ \text{domain } \{0,1\}$$

A variable with a β binomial distribution is distributed as a binomial distribution with parameter p, such that p is a distribution according to a B distribution with parameters α and β. For n trials, it has probability density function

$$P_{\beta\ binomial}(x) = \frac{B(x+\alpha,\,n-x+\beta)\,n!}{B(\alpha,\beta)\,x!\,(n-x)!}\,, \quad \text{parameters } \alpha > 0,\ \beta > 0,\ n \text{ posint, domain } \{0,1,2,...n\}$$

The binomial distribution is a discrete probability distribution of the number of successes in a sequence of n independent trials yielding either *success* or *failure*, each of which yields a success with probability p. Such a trial is also called a Bernoulli experiment; when $n = 1$, the binomial distribution is a Bernoulli distribution.

$$P_{binomial}(x) = \frac{n!\,p^x\,(1-p)^{(n-x)}}{x!\,(n-x)!}\,, \quad \text{parameter } n \text{ posint, domain } \{0,1,2,\,...\,n\}$$

The discrete uniform distribution is a probability distribution whereby equally spaced values of finite number are equally likely to be observed; each value of n has equal probability $\frac{1}{n}$.

$$P_{discrete\ uniform}(x) = \frac{1}{n}\,, \quad \text{parameter } n \text{ posint, domain } \{0,1,2,...n\}$$

The geometric distribution is the probability distribution of number $y = x - 1$ of failures before the *first success*, supported on the set $\{0, 1, 2, 3, ...\}$.

$$P_{geometric}(x) = p\,(1-p)^x, \quad \text{parameter } 0 \le p \le 1 \text{ domain } \{0, 1, 2, \,...\,\}$$

The hypergeometric distribution is a discrete probability distribution that describes the probability of x successes in n draws from a finite population of size N containing m successes without replacement.

$$P_{hypergeometric}(x) = \frac{(Np)!}{x!\,(Np-x)!}\,\frac{(Nq)!}{(n-x)!\,(Nq-n+x)!\,\frac{N!}{n!\,(N-n)!}}\,, \quad \text{parameters } k = 0, 1, 2, \,...\, N$$

$$, n = 0, 1, 2, ...N, p = \frac{k}{N}, q = 1 - p, \text{domain } \{0, 1, ..., n\}$$

The log-series distribution is a discrete probability distribution derived from the Maclaurin series

expansion $-\ln(1 - p) = p + \dfrac{p^2}{2} + \dfrac{p^3}{3} + ...$

$$P_{log\text{-}series}(x) = -\frac{\theta^x}{x \ln(1 - \theta)}, \quad \text{parameter } 0 < p < 1, \text{domain } \{0, 1, 2, ...\}$$

The Pascal distribution with parameters k and p arises in the scheme of the Bernoulli trial with probabilities p of *success* and $1 - p$ of *failure*, as the distribution of the number of failures up to the occurrence of success k.

$$P_{Pascal}(x) = \frac{(x + k - 1)!}{(k - 1)! \, x!} p^k (1 - p)^x, \quad \text{parameters } 0 \leq p \leq 1, k = \text{posint, domain } \{0, 1, 2, ...\}$$

A discrete stochastic variable x has a Poisson distribution with parameter $\lambda > 0$ if, for $x = 0, 1, 2,$
... the probability mass function of x is

$$P_{Poisson}(x) = \frac{\lambda^x \, e^{(-\lambda)}}{x!}, \quad \text{parameter } \lambda > 0, \text{domain } \{0, 1, 2, ...\}$$

In probability theory and statistics, the Rademacher distribution is a discrete probability

distribution that has chance $\dfrac{1}{2}$ for either 1 or −1.

$$P_{Rademacher}(x) = \textbf{if } x = -1 \textbf{ then } \frac{1}{2} \textbf{ else if } x = 1 \textbf{ then } \frac{1}{2} \textbf{ else } 0 \textbf{ end if}, \text{ domain } \{-1, \ 1\}$$

A Skellam distribution is the discrete probability distribution of difference $n_1 - n_2$ of two statistically independent random variables N_1 and N_2, each having Poisson distributions with distinct expected values μ_1 and μ_2; $I_{|x|}(z)$ is a modified [Bessel](#) function of the first kind.

$$P_{Skellam}(x) = e^{(-(\lambda_1 + \lambda_2))} \left(\frac{\lambda_1}{\lambda_2}\right)^{\left(\frac{x}{2}\right)} I_{|x|}(2\sqrt{\lambda_1 \lambda_2}), \quad \text{parameters } \lambda_1 > 0, \lambda_2 > 0, \text{domain } \{...,-2, -1$$
$$,0,1,2,...\}$$

  An application of statistical methods to analysis of experimental data requires an understanding of these data and their characteristics. The descriptive statistics of the most common types applied to a data set are a measure of its central tendency and a measure of its variability or [dispersion](#); the central tendency of data in a sample is an [average](#) value of a [variable](#) being observed, which might be taken to be typical of that variable, and the variability is a measure of the extent to which all data are near that typical value. Among a few measures of the location of a data set, the most meaningful characteristics are, for a particular quantity, the [mean](#); for the variation from that mean, a [variance](#) or [standard deviation](#) is a common statistic. According to statistical convention, greek letters denote parameters of a parent distribution, and latin letters denote their estimates; because of limitations of type and fonts available for a *Maple* worksheet, we refrain from rigorous conformity to this convention.

  Statistical methods are generally based on a theoretical [distribution](#) that approximates,

according to an expectation of an applicability of the law of large numbers, an actual distribution. We treat here first a univariate distribution, applicable to a single random variable, directly measured on multiple occasions under otherwise nominally similar conditions; below, we discuss bivariate systems in which a response variable variable depends on a single factor as independent variable. A distribution resulting from the measurements of an original population is called an underlying or parent distribution. A distribution that is not symmetric with respect to its mode is skewed, to greater values than that mode in the distribution if the tail in that direction be greater than in the other direction from the mode, or positively skewed if frequencies greater than at the mode be favoured over frequencies less than at the mode, and negatively skewed otherwise; Pearson's measure of the skewness is $\dfrac{mean - mode}{standard\ deviation}$. Whether a mean lie to the left or right of a median for a positively skewed distribution depends on the particular characteristics of that distribution, such as whether the areas to the right and left of the median are equal. A distribution is bimodal if it have two maxima; a single maxima is called a mode; the existence of a bimodal distribution, if the difference between the modes be significant, might indicate some inhomogeneous factor to be present in the population.

No measurement of a physical property of a chemical system on a continuous scale is perfect or exact, unless fortuitously so; apart from gross blunder, error random and systematic enters inevitably to some extent a measurement of a nominally continuously variable quantity. Error is defined as a difference or deviation between a result of an individual measurement, or observation, and the true value of a measurand that might be obtained from a perfect measurement; because a true value is indeterminate in a practical case, error is an idealized concept, estimated according to a difference, called a residual, between an individual measurement and an expected result based on numerous tests under repeatable conditions. A discrepancy is a difference betwee two measured values of the same quantity. Apart from a measurement, an observation might also be an answer to a question -- affirmative or negative -- or a classification -- acceptable or unacceptable. A known error is applicable to a result as a correction. According to convention, an assigned or reference value is accepted as a true value. For instance, with spectral measurements of rotational parameters of simple gaseous compounds based on both wave lengths and frequencies, one could formerly estimate the speed of light in vacuo through $c = \lambda \nu$, with an associated uncertainty comparable to that attached to another fundamental physical or chemical constant such as that of Planck h or Avogadro N_A, but to this quantity is assigned a value $c = 299792458$ m s$^{(-1)}$, without uncertainty. Before that assignment, measurements of that speed in diverse experiments exhibited a central tendency, i.e. to occur within a small domain about some central value that might or might not have been an explicit experimental result; during the nineteenth and twentieth centuries, the range of values of c became progressively more narrow, as the dispersion of values about some central value decreased. An eventual assignment of that stated value of c evades uncertainty, but that situation is rare: measurements of other quantities, and thereby their consequent values, are inevitably subject to uncertainty. Although an uncertainty might have a form of a range, such a value can not serve to correct a result of a measurement. To estimate an uncertainty of a measurement, one must

- state clearly what is being measured, including a relation between a measurand and input

quantities,

- identify sources of uncertainty,

- measure or estimate the magnitude of each identified prospective component of uncertainty, and

- calculate a combined uncertainty, which remains essentially qualitative even if expressed in numerical form.

The fractional uncertainty of a measurement is the ratio of the uncertainty δx to the best estimate x of the value of the quantity, i.e. $\left|\dfrac{\delta x}{x}\right|$, in which the absolute value ensures a positive result. For a result of a counting experiment, such as the number of radioactive decays per counting interval indicated with a Geiger-Mueller counter, the average number of events in that interval might be expressed as $n \pm \sqrt{n}$. The uncertainty δx in a quantity x taken to a power n, for $f = x^n$, produces fractional uncertainty $\dfrac{\delta f}{|f|} = |n| \dfrac{\delta x}{|x|}$. The uncertainty in a formula f(x) involving one variable is δf $= \left|\dfrac{\partial}{\partial x} f\right| \delta x$, whereas the uncertainty in a formula f(x, ..., z) involving multiple uncorrelated variables is

$$df = \sqrt{\left(\left(\frac{\partial}{\partial x}f\right)\delta x\right)^2 + ... + \left(\left(\frac{\partial}{\partial z}f\right)\delta z\right)^2}$$

To extract quantitative information about a chemical system that transcends random effects, one must apply statistical methods. Summary measures of data correspond to either populations or samples; a summary measure from a sample constitutes a statistic. In either case, one's objective is to express a result as ($x \pm u$) units, in which x denotes the best or most representative value of a measured quantity and u denotes an uncertainty attributed to that measurement; a standard deviation u is a standard uncertainty and a measure of the dispersion of measurements of a property of a sample, and serves as a practical statistical indicator. Although a variance has a greater significance in statistical theory, the standard deviation, or standard error in relation to a value of a parameter, has a direct relevance through its magnitude in relation to the mean, and has the same units as the quantity being measured. Three criteria of the worth of a statistic to estimate a parameter are

- that there is a tendency of the statistic to assume values near that parameter,

- that the estimator is reliable, and

- that this reliability improves with increasing size of the sample.

A statistic becomes a consistent estimator of a parameter when the probability that its value is near the true value approaches unity with increasing size of the sample.

Two measures of the spread of a distribution of observations are a Q spread or the interquartile range, which is the distance between the first and third quartiles, and an E spread, which is the distance between the first and seventh eighths of samples.

Moment k of a set of values is a mean of those values each taken to power k:

$$\mu_k(x) \; = \; <x^k> \; = \; \frac{\sum\limits_{j=1}^{n} x_j^{\;k}}{n}$$

The first moment about zero is thus the arithmetic mean of a sample, but the first moment about the mean is zero as the average deviation from the mean is zero. The mean or expected value $\overline{x} = \mu_1(x) = <x>$ of a discrete random variable x is

$$\overline{x} = \sum\limits_{j=1}^{n} x_j \, p_j \, ,$$

in which p_j is the probability of occurrence of a particular value x_j or the sample proportion, or

$$\overline{x} = \frac{\sum\limits_{j=1}^{n} x_j}{n} \, ,$$

in which n denotes the number of observations or sample size, or

$$\overline{x} = \frac{\sum\limits_{j=1}^{k} f_j \, x_j}{n}$$

in which f_j is the frequency of particular class j in an interval represented with its midpoint x_j for classes numbering k ; for a continuous random variable represented according to a formula $\mathrm{f}(x)$, the mean is

$$\overline{x} = \int x \, \mathrm{f}(x) \, dx.$$

The difference between the second moment, $\mu_2(x)$, and the first moment squared, $\mu_1(x)^2$, is the variance; its positive square root is called standard deviation σ_x, for the sample:

$$\mu_2(x) - \mu_1(x)^2 = \sigma_x^{\;2}$$

The variance of the mean is $\dfrac{\sigma_x^{\;2}}{n}$. That standard deviation is a measure of the likely random error of any single measurement. A function $<\mathbf{e}^{(tx)}>$ generates moments for a continuous random variable x as a weighted sum of the moments upon substitution of a Taylor expansion of this exponential function:

$$<\mathbf{e}^{(tx)}> \; = \; \int \mathbf{e}^{(tx)} \, \mathrm{f}(x) \, dx = \; 1 + \; t <x> \; + \; \frac{t^2}{2!} \; <x^2> \; + \; \dots$$

Because the sum of all deviations equals zero, we define the mean deviation in terms of the absolute values, or magnitudes, of the deviations $\overline{(\delta x)} = \sum\limits_{j=1}^{n} \dfrac{\left| x_j - \mu_1(x) \right|}{n}$. This mean deviation is a measure of the dispersion of the observations about the mean, but its definition in terms of absolute values causes inconvenience in statistical analysis; for this reason the variance is a

superior indicator of the dispersion.

For a weighted mean,

$$\bar{x} = \frac{\displaystyle\sum_{j=1}^{n} w_j\, x_j}{\displaystyle\sum_{j=1}^{n} w_j}$$

with the weight typically taken as the reciprocal of the variance of the measurement,

$$w_j = \frac{1}{\sigma_j^2}$$

for each measurement j in a set numbering n, such that the uncertainty in $\bar{x}$ is the reciprocal square root of the sum of individual weights. The variance of the mean is accordingly $\dfrac{1}{\displaystyle\sum_{j=1}^{n} w_j}$.

Other than a mean, measures of the central tendency comprise a median, which is the central value of ordered data so with equal numbers of observations of greater or lesser values, and a mode, which is the most common value of a datum in a set of observations. A median provides a superior measure of central tendency when data exhibit atypically large or small observations, or when the data exhibit an asymmetry between small and large values; the mean is much more strongly influenced by an outlying datum than the median. When data are grouped into classes, the midpoint of the interval containing the greatest class frequency represents the mode. For qualitative data, the proportion of a subset is the only available measure.

Apart from the central tendency described with a mean, median or mode, for a set of observations we require a measure of the dispersion of those data; as the range of the data, from the smallest to the largest values in the set, is sensitive to the extent of the data -- as a further point might lie outside the preceding range, other measures of dispersion are generally applicable, namely variance and its square root as a standard deviation. Variances of a sum or difference of quantities x_1, x_2, x_3, ... are additive: for this sum or difference,

$$y = x_1 \pm x_2 \pm x_3 \pm \ \dots$$

providing that there be no correlation between x_1 and x_2, between x_1 and x_3 et cetera, with σ_1^2 being the variance of x_1, σ_2^2 being the variance of x_2 and so forth, the variance of y is

$$\sigma_y^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ \dots$$

For a product or quotient, squares of relative errors are additive: for this product or quotient,

$$z = \frac{x\,y}{w}$$

providing that no correlation exist between x and y, between x and w, et cetera, the variance of z is calculated as its ratio with z^2:

$$\frac{\sigma_z^2}{z^2} = \frac{\sigma_x^2}{x^2} + \frac{\sigma_y^2}{y^2} + \frac{\sigma_w^2}{w^2}$$

and analogously for other expressions involving multiplicands and divisors. Tests for outliers applicable to data in a univariate set are attributed to Dixon and to Grubbs. An estimate of a standard deviation based on a range is a quarter of that range.

Through an analysis of variance, one tests, for univariate data, whether samples in one group, such as from a batch, a method or a laboratory, differ from the population of subjects investigated, such as several batches of one product or varied methods for the same parameter or multiple laboratories participating in the testing. To problems of the following type, testing according to analysis of variance is amenable: measurements in several groups are available for a particular product, several repeated measurements were conducted on each batch, and the same analytical method was applied for all testing. An analysis of variance according to a matrix generated to represent a table for a classification has a purpose, to test the null hypothesis, which is either true or false, either a positive or a negative statement, with no ambiguity: for a null hypothesis that the sample means of many populations are all equal, the alternative hypothesis is that they are not equal. A null hypothesis summarizes a situation in such a manner that any non-random variation in the situation disproves the hypothesis. In general a test statistic must be chosen, and a level of significance must be specified, according to both of which a decision rule must be formulated. On the basis of a calculated test statistic for the sample, a decision is made to accept or to reject the null hypothesis; in the latter case the alternative hypothesis becomes applicable. A null hypothesis can not be proved -- it can be only disproved, but with some probability of being incorrect. Tests of hypotheses are based on such parameters as Student's t or other statistic. Such testing is susceptible to error of two types, I and II: error of type I occurs when a rejected hypothesis is true, whereas error of type II occurs with the acceptance of an hypothesis that is false. A null hypothesis, and an alternative hypothesis, should be formulated before a collection of data, and hence become a part of the experimental design.

A statistical population is a collection of all possible observations of a specific characteristic or property of interest, whereas a sample contains only some observations. When a characteristic is not numerical, such as a composition in terms of chemical elements, the population is qualitative, whereas a quantitative population is expressible numerically. One might distinguish formally between estimates of characteristics for a total quantitative population and for a sample: in statistical contexts, greek letters conventionally denote characteristics of a population, such as its mean μ and its standard deviation σ, and roman letters denote characteristics of a sample taken from that total population, such as estimate m or $\bar{x}$ of its mean and estimate s of its standard deviation. Because in conditions typical of chemical experiments one makes finite measurements on only a sample rather than a total population, quantities derived from those measurements are merely estimates of what might be applicable for a total population; for such estimates roman letters are accordingly preferable. A theoretical population is generally an idealization of an actual population from which one draws a sample. As symbols for all chemical uses number finitely, we tend here to neglect that formal distinction between roman and greek letters, but endeavour to maintain consistency of usage with whatever letters seem appropriate in particular circumstances.

For a random variable, the numerical value assigned to it or its symbol is determined

stochastically; whereas a continuous random variable might assume any numerical value on a continuous scale, a discrete random variable can assume a value among a countable number of such values. The expected value of a discrete random variable is a mean of possible values of that variable weighted according to the respective probabilities. The variance of a random variable is a mean of the squared deviations from the expected value calculated with probability weights. A distribution is a set of possible values of a random variable, or a set of points in a sample space, considered in terms of their theoretical or observed frequency. A probability distribution provides a probability for each possible value of a random variable. Raw data become converted into values clustered into class intervals, each of which has a lower and an upper bound to demarcate the adjacent intervals; the width of a uniform class interval equals the difference between largest and smallest values in the sample divided by the number of class intervals. The ratio of a number of observations in a particular class to the total number of observations becomes the relative frequency; a sum of frequencies for successively greater class intervals yields a cumulative frequency, a curve to depict which has typically the shape of an ogive. The subpopulations of the class intevals in an ordered manner generate a frequency distribution of a sample. Such a distribution might be depicted as an histogram or bar chart, or as a frequency polygon or curve. To plot an histogram for a single experimental factor, we partition the factor space into bins, so that a value of factors in any particular combination corresponds to a position in factor space and falls into only one bin.

For a distribution of arbitrary shape, Chebyshev's rule states that at least $\frac{3}{4}$ of the observations fall within two standard deviations of the mean, or at least $\frac{8}{9}$ within three standard deviations, but that rule is too general for practical application despite its implication that much information is imparted in that descriptor of a population. To achieve a more concrete criterion, we require a knowledge of the distribution of frequencies associated with a measured quantity.

For any continuous distribution f(x) that might be considered to represent the limit of a histogram for many measurements of a continuous variable x and narrow class intervals, the probability that a single measurement yields a value between x and $x + dx$ is f(x) dx; the probability that a single measurement yields a value between $x = a$ and $x = b$ is $\int_{a}^{b} f(x)\,dx$, and the total probability of any value corresponds to the normalization condition, $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

A gaussian or normal distribution is important because, in statistical treatments of error associated with physical measurement in chemical or other experiments, one commonly assumes that such random error is normally distributed; many distributions of measured characteristics display such a form, but by no means all distributions have such a symmetric shape, others showing a skewed form. Such an assumption of gaussian form might be justified upon neglect of systematic error, which is taken into account separately -- when detected, and in an absence of other bias of a measuring device; a systematic error can arise from a common offset of all measurements, whereas bias arises if large values of a measured quantity be significantly favoured

or disfavoured relative to small values, i.e. on either side of an arithmetic mean, thus skewing a distribution. A gaussian or normal distribution of x is unimodal, continuous and symmetric about its mean μ, which is also its median and its mode, according to this formula for the probability density with standard deviation σ:

$$\frac{e^{\left(-\frac{(x-\mu)^2}{2\,\sigma^2}\right)}}{\sigma\sqrt{2\,\pi}}$$

According to the central-limit theorem, for a sum Y of n independent variables x_i, $i = 1 .. n$, each with its own mean and variance, a distribution for Y becomes *normally distributed* as $n \to \infty$, because random errors from varied sources tend to compensate for one another; because the practical number of measurements of a quantity is limited, one must rely on this theorem in expecting that a few actual data behave in the same manner as many prospective but inaccessible data. For a random sample with mean $\bar{x}$ and of size n taken from a normally distributed and large population with mean μ and standard deviation σ, the sampling distribution for mean $\bar{x}$ of that sample is also gaussian; its standard deviation is $\dfrac{\sigma}{\sqrt{n}}$. When the population N is small relative to the sample n, this quantity becomes xxx

$$\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{n-1}}\ .$$

According to Winsor's principle, the variation of frequency near the centre of any distribution is typically approximated closely with a gaussian distribution, but one might argue that this approach by devoting too much attention to the centre of a distribution, becomes misleading. For practical purposes, unless there be shown a manifest asymmetry or there is a knowledge a priori that another distribution is applicable, a distribution over less than 30 measurements must be regarded as belonging to a gaussian kind, even if small deviations be observed, because the contrary can not be proved. In one sense, parametric statistics, for instance in terms of a mean and a standard deviation, are concerned with measurements that conform to a gaussian distribution, and non-parametric statistics to other distributions.

   The important advantages of this gaussian distribution are its efficiency, its lack of bias, its wide acceptance and its incorporation into many tests and much software for analytical chemistry and elsewhere. A gaussian or normal distribution serves as a basis upon which one compares other distributions. When a plot of a distribution fails to exhibit a symmetric shape that might be approximately normal, a variant of a gaussian distribution, called a log-normal distribution and obtained on plotting the abscissal quantity on a logarithmic scale, might be appropriate; *Maple* provides a plot for this distribution in both old package stats and package Statistics that supersedes the former. A highly skewed distribution, exponential, is typical of populations of which observations vary over time, such as the temporally decreasing concentration of a chemical reactant for which, for $t > 0$, $c = c_0\,e^{(-k\,t)}$, or the radioactive decay of unstable elements. Other distributions might be positively or negatively skewed, in which case the median is a realistic measure of the location of data. Distributions of still other shapes are encounted in chemical

measurements, such as a [bimodal](#) distribution that exhibits two maxima, implying a non-homogeneous factor within the sample, and thereby defies direct description with standard measures.   For a normalized probability distribution or normalized probability density $p(x)$ governing a value of continuous variable x between a smallest value a and a largest value b, the probability of values between x and $x + dx$ is $p(x)\,dx$, and

$$\int_a^b p(x)\,dx = 1.$$

With such a normalized distribution, the population mean is calculated as

$$\mu = \int_a^b x\,p(x)\,dx$$

One might generally take $a = -\infty$ and $b = \infty$ without introducing significant error.   The variance of a probability distribution of that variable x is $\overline{(x^2)} - \overline{\mu^2}$ and the standard deviation is thus

$$\sigma = \sqrt{\overline{(x^2)} - \overline{\mu^2}} \; = \; \sqrt{\int_a^b (x - \mu)^2\,p(x)\,dx} \; = \; \sqrt{\int_a^b x^2\,p(x)\,dx - \left(\int_a^b x\,p(x)\,dx\right)^2}$$

If a raw distribution of numerous data have, like a Laplace distribution also known as a double exponential distribution, a symmetric shape and long tails, relative to a gaussian distribution, the mean is a poor estimator because it is sensitive to outliers; in this condition a trimmed mean is preferable, such that trimming, by 10 or 20 per cent of the data, eliminates the outliers.   For large samples the sample median has a standard error $\dfrac{\sigma}{\sqrt{2}\,\sqrt{n}}$, whereas the standard error of the mean is $\dfrac{\sigma}{\sqrt{n}}$.

Whereas that gaussian distribution is a continuous function, [Poisson's distribution](#),

$$\frac{e^{(-\mu)}\,\mu^r}{r!}$$

is discrete in describing the probability of the number r of successes of an event according to a distribution with mean μ.   Inserting r with values of successive integers $r = 0, 1, 2, ...$, we obtain a total probability,

$$\frac{e^{(-\mu)}\,\mu^0}{0!} + \frac{e^{(-\mu)}\,\mu^1}{1!} + \frac{e^{(-\mu)}\,\mu^2}{2!} + \frac{e^{(-\mu)}\,\mu^3}{3!} + ... = e^{(-\mu)}\left(1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + ...\right) = e^{(-\mu)}\,e^{\mu} = 1$$

which must be unity by definition.   This distribution is useful when a probability p of any particular outcome of a trial is small and the number n of trials is large; in that case, $\dfrac{e^{(-\mu)}\,\mu^r}{r!}$ approaches closely the value of $C(n, r)\,q^{(n-r)}\,p^r$, so that this poissonian distribution approximates closely a binomial distribution, for instance for $n > 50$ and $p < \dfrac{1}{10}$.

The [binomial distribution](#) is typically applicable to experiments in which a result is one final

state among a small number thereof. Both the gaussian and Poisson's distribution might be considered to be limiting cases of a binomial distribution.

For an approximately gaussian or normal distribution, approximately 68 per cent of observations fall within one standard deviation of the mean, and 95 per cent within two standard deviations.

As a robust alternative to regression based on a criterion of least squares of residuals, on which outliers tend to exert an undue influence, a *resistant line* might be formed from the use of medians of data divided into three or more sets; residuals should be investigated to detect patterns unexplained by the resistant line. This use of medians should occur only when weighting of data, as explained below, is impracticable. A plot of residuals shows whether there is a systematic trend of the scatter, an increasing or decreasing scatter with increasing independent variable -- which might indicate the necessity of a transformation of data, or a curvilinear pattern -- which indicates that an alternative model might be preferable. For data that are collected sequentially during some temporal period, if these data or their residuals exhibit a dependence on time, a time factor should be included in the model.

For practical calculations involving statistical treatment of experimental results of univariate or multivariate systems, a spreadsheet is a useful tool.

regression

Suppose that, in some experiment, we collect ten data points with one independent variable x and one dependent variable y, so (x_i, y_i) for $i = 1 .. 10$; when we plot those points as cartesian coordinates, their locations appear to lie near a straight line, so to indicate a linear dependence of y on x. We choose hence to represent these data compactly with a formula $y = m x + b$; the best linear approximation of the data in an absolute sense becomes the problem of minimizing $\max(|y_i - (m x_i + b)|, i = 1 .. 10)$, which is called a minimax problem, but for which there is no solution according to elementary methods. A second approach to determine the best linear approximation requires an evaluation of m and b to minimize $\sum_{i=1}^{10} |y_i - (m x_i + b)|$ that corresponds to the sum of the absolute deviations from the linear relation; such an approach entails the difficulties that the absolute-value function is not differentiable at zero and that solutions to the equations $\frac{\partial}{\partial m}\left(\sum_i |y_i - (m x_i + b)|\right) = 0$ and $\frac{\partial}{\partial b}\left(\sum_i |y_i - (m x_i + b)|\right) = 0$ cannot necessarily be obtained. A third approach for this purpose requires an evaluation of m and b to minimize $\chi^2 = \sum_{i=1}^{10} (y_i - (m x_i + b))^2$ that corresponds to the sum of the squared deviations from the linear relation. The latter method of least squares is not only convenient but has advantages over the preceding methods as follows: the minimax method typically assigns excessive weight to an item of data that is seriously erroneous; the method of absolute devation averages the error of each point and lacks sufficient weight for a point that deviates much from the apparent linear relation, whereas the method of least squares places much weight on a point that seems inconsistent with the other data, but prevents that point from dominating the approximation completely. The

method of least squares is not only computationally convenient but also favoured by theoretical considerations in relation to the statistical distribution of error. If standard deviation σ_j be known for each value of dependent variable, the corresponding expression to minimize is

$$\chi^2 = \sum_{i=1}^{10} \left(\frac{y_i - (m\,x_i + b)}{\sigma_j} \right)^2$$

so that $\dfrac{1}{\sigma_j^2}$ serves as a weight of each term in the sum. In case of either weighted or unweighted data, the method of least squares involves minimizing simultaneously χ^2 with respect to parameters m and b.

One example of the application of an analysis with linear regression and least squares of residuals is to ascertain whether a non-zero intercept on the ordinate axis be significant, which would imply the necessity of a 'blank' in that analysis. A null hypothesis would be that this intercept equals zero; as a test, Student's t statistic at confidence level 95 per cent might be applied. Analogously, when a theoretical slope of a regression line is known, an estimated slope can be tested for error. The null hypothesis is that the theoretical slope is statistically equivalent to the estimated slope; Student's t statistic provides a test of this hypothesis.

The fitting of such a collection of data to a relation, linear or non-linear, is called regression if there be no constraint on the values of the fitting parameters, such as m and b in the preceding examples, or optimization if there exist one or more such constraints. Regression is a mathematical method to relate one or more variables to another by means of a functional relation or formula, which is then amenable for a use to predict an unknown value of a variable from a known or given value, or multiple values, of other variables. In an analysis of data to be subjected to regression, one must first select a model or objective function, which is really a formula represented in a geometric construction as a straight line, a parabola or an exponential decay or other form; a plot of the data, if practicable in two or three dimensions, might provide an indication of a suitable form. Any such form must contain variables -- indpendent and dependent -- that correspond to measurable quantities and are consequently known rather than unknown, and parameters as coefficients or addends of such variables; each such form has intrinsically a maximum number of such parameters, and the purpose of regression is to evaluate the extent of association, or correlation between or among these variables. For data in a particular set and a selected model, the evaluation of the best fit requires a criterion. If a selected model seem to be amenable to improvement, in that its parameters seem inadequate either to represent the data adequately or to be poorly defined, either a forward strategy, in which the model is expanded to encompass additional parameters, or a backward strategy, in which poorly defined parameters are eliminated, is commonly implemented. For chemical and physical phenomena, a theory is commonly available to indicate a preliminary model that is subject to test for the particular data set; the failure of such a theoretical model to represent adequately -- i.e. within the expected error of measurement, according to measures of goodness of fit -- these experimental data might imply the development of a further theory. Although, as a model, a straight line is prototypical for linear regression, the descriptor applies to the parameters, not to the variables; that a model is linear in parameters implies that a partial derivative of a dependent variable, or of a difference between

expected and measured values of a dependent variable, with respect to any parameter in the model contains no parameter, whether the same or other. For instance, for an electric circuit containing a resistor of known resistance R with the potential difference across the resistor depending on time according to $V = V_0 \cos(\omega t)$ with *known* frequency ω, for the measurement of instantaneous current as a function of time parameter V_0 occurs linearly, because $\dfrac{\partial}{\partial V_0} V = \cos(\omega t)$, as the right side contains no parameter; in contrast, with known V_0 for the measurement of instantaneous potential difference V as a function of time to evaluate parameter ω, that parameter occurs non-linearly because $\dfrac{\partial}{\partial \omega} V = -t\, V_0 \sin(\omega t)$ and the right side contains that same parameter ω. In the former case, one can apply methods or algorithms of linear regression to evaluate the parameters, whereas in the latter case one should apply directly methods or algorithms of non-linear regression, rather than to reformulate the model into a pseudo-linear form; the latter procedure would yield a biased estimate of parameters unless the nature of that pseudo-linear transformation be taken into account through appropriate weighting of the data. Such pseudo-linear regression might serve, however, to produce initial estimates of parameters for use in non-linear regression, for which purpose such estimates are typically required.

    The fitting of experimental data to an empirical model or a theoretical formula is a common computational operation in experimental science. A primary objective of analysis with methods of linear or non-linear regression is to obtain predictions of one variable using known or set values of others. In a case of a single independent variable for which a fit of experimental data (x_j, y_j) to a straight line is a typical exercise, a graphical fit that relies on visual judgment for an optimal quality might suffice, but any deductions about the precision of the parameters -- slope and either intercept -- must then be subjective. When the variables or fitting parameters become numerous, numerical means of fitting are inescapable. Although computational procedures based on a criterion of the least squares of residuals, such as those in this chapter, available in spreadsheets or even those on a pocket calculator, relieve the burden of manual calculations, one must maintain a critical sense about the validity of the results obtained, for which statistical indicators provide essential tests. The criterion of the least squares of residuals in a fit is based on a concept maximum likelihood: the parameters in an optimal set yield a maximum probability function for all measurements. For a single measurement of which the error is subject to a gaussian distribution, a probability of making at $x = x_j$ a single measurement of y equal to y_j is

$$P_j = \frac{e^{\left(-\dfrac{(y_j - f(\beta_1, \beta_2, \ldots, \beta_n, x_j))^2}{2\,\sigma_j^2} \right)}}{\sqrt{2\,\pi}\,\sigma_j}$$

For all measurements, the total probability is a product of these individual probabilities,

$$P = \prod_j P_j = \prod_j \frac{e^{\left(-\left(\sum_j \frac{(y_j - f(\beta_1, \beta_2, ..., \beta_n, x_j))^2}{2\,\sigma_j^2}\right)\right)}}{\sqrt{2\,\pi}\,\sigma_j}$$

The variation of P with respect to a parameter β_j involves only the sum in the exponent, so that a maximum of P corresponds to a minimum sum there. The maximum likelihood of P becomes the principle of least squares according to the variation of the parameters,

$$\delta_{\beta_1}\,\chi^2 = \delta_{\beta_2}\,\chi^2 = ... = \delta_{\beta_n}\,\chi^2 = 0$$

in which $\delta\,\beta_j$ signifies a variation of χ^2,

$$\chi^2 = \sum_j \left(\frac{y_j - f(\beta_1, \beta_2, ..., \beta_n, x_j)}{\sigma_j}\right)^2$$

with respect to an infinitesimal and independent variation of parameter β_j, provided that $f(\beta_1, \beta_2, ..., \beta_n, x_j)$ possesses no discontinuity in either itself or its first derivative with respect to any β_j. The practice of regression according to a criterion of the least squares of residuals is applicable not only to cases in which the measurements of y_j are known to conform to a gaussian distribution but even when these might be known not so to conform, because the central-limit theorem states that a sum or mean of many measurements approaches that of a gaussian distribution, irrespective of the distributions of individual measurements, unless a particular measurement contributes a large fraction of the sum or mean or there are excessive variations of the widths of individual distributions.

An alternative definition of χ^2 has as basis

$$\chi^2 = \sum_j \frac{(f_o - f_e)^2}{f_e},$$

in which f_o denotes the observed frequency of an event and f_e denotes its expected frequency. A large ratio $\dfrac{(f_o - f_e)^2}{f_e}$ implies a large value of χ^2 and hence that the deviation from an expected result is large. This statistic has its associated distribution,

$$\chi(\nu)^2 = \frac{x^{\left(\frac{\nu}{2} - 1\right)}\,e^{\left(-\frac{x}{2}\right)}}{2^{\left(\frac{\nu}{2}\right)}\,\Gamma\left(\frac{\nu}{2}\right)},$$

in which ν denotes the number of degrees of freedom, which is typically defined as the number of variables minus the number of parameters; for this χ^2 distribution and others, ν is hence the only parameter.

For a bivariate system, a typical experiment in a chemical or physical laboratory involves measurement of some quantity according to set or accepted values of another quantity; for

instance, one might measure the concentration of a product of a chemical reaction as a function of time or temperature or initial concentrations of reactants, which also imply the corresponding measurements or settings. The former quantity is deemed a response variable, which might vary continuously, or might be a count or a proportion or a 'time at death' or a category; the latter quantity, known as a factor or *explanatory variable* because it serves to *explain* the response, might vary continuously, or be categorical, or both. A common objective of such chemical or physical experiments involves seeking a quantitative relation between values of explanatory variables and the resulting values of response variables; the latter terms amount to statistical jargon that imply independent and dependent variables respectively. An analysis of regression indicates how one variable, such as a factor, is related to another, such as a response, by yielding an equation according to which a known value of a factor serves to estimate the unknown value of the response variable. In our discussion and statistical treatment of regression, we assume implicitly that a quantity that acts as a dependent variable varies continuously, although values of independent variables might vary continuously or adopt only discrete values, such as integer values of quantum numbers; other forms of variation, such as count or proportion, are less common in a chemical context.

A functional relation, such as $y = f(x)$ between two variables x and y and function f, either is based on theory or might be expected to be amenable to such a theoretical basis; such a relation might be perfect if that theory be valid. For a statistical relation, a theoretical justification is lacking; moreover, because there is only a partial dependence of y on x, an analysis can indicate an association or a correlation between independent variables, known as predictor *variables* or *regressors*, and a dependent variable or response variable. A linear relation between two variables x and y is represented by a straight line in a plot in two dimensions, about which particular data points might be somewhat scattered. One deduces a statistical relation by means of analysis of a kind called regression, which might be useful to reproduce data in a compact form, but such a statistical relation by no means implies a causal relation; in forming such a relation the values of independent and dependent variables are known from measurement, and other symbolic quantities, called parameters, within a relation are unknown but become subject to fitting as a result of that regression. Before one undertakes such an analysis, plotting the data points in two spatial dimensions is helpful to ensure that a trend of correlation exists between values of the various variables; a large scatter of points about any possible direction might indicate either a weak correlation or a lack of correlation between variables. If a theoretical relation be unknown or inapplicable to these data that still exhibit a correlating trend, the next step is to postulate some functional relation between the variables, of which a line be not necessarily linear but possibly curvilinear. This regression might show either a direct or inverse relation between factors and response, depending whether the response increase or decrease when the factor increases.

For chemical and physical experiments involving numerous molecules in each trial and conditions that might be well defined or controlled, a causal relation, proposed as a working hypothesis, might be established more convincingly than in a biological or sociological system, according to these criteria:

- strength of association -- the stronger is the association, as indicated by coefficients of correlation of variables with magnitudes near unity and by other applicable statistical indicators,

the more likely is the avoidance of a spurious association because of a bias;

- effect of predictor and response variables -- the value of the response variable alters in a meaningful manner with the predictor or causal agent being tested;

- temporal order -- the hypothetical cause precedes the occurrence of the effect;

- consistency of the findings -- tests repeated yield reproducible results;

- plausibility of the hypothesis -- the hypothetical causal relation is consistent with current theoretical knowledge, although the latter might be insufficient to explain further findings;

- coherence of the evidence -- there is no serious conflict of the findings with accepted knowledge of the response variable under test, and

- specificity of the association -- the suspected predictor variables are associated with only one response variable.

Even if all these criteria be satisfied, a causal relation can not be claimed with complete certainty, because other pertinent factors or predictors might have been neglected.

simple linear regression with weighting of data

In an experiment in which are collected n data points in a set that is randomly selected from the sets of infinite number from the parent population, and which data points are distributed according to that parent population, for a gaussian distribution of mean μ and variance σ^2, the probability dP_j of making a single observation x_j within interval dx is $dP_j = p_j\,dx$ according to probability function $p_j = p_g(x_j, \mu, \sigma)$. For all n observations of a trial distribution of mean μ' and variance σ^2, the probability of observing that particular set is calculated as the product of the individual probability functions,

$$P(\mu') = \prod_{j=1}^{n} P_j(\mu')$$

According to the principle of maximum likelihood, on comparison of probabilities $P(\mu')$ of obtaining that set of data from various parent populations with different mean μ' but the same variance σ^2, the probability is greatest that the data were derived from a population with $\mu' = \mu$ -- i.e. the most likely population as a source of these data is assumed to be the correct one. If the probability of measuring a value x_j conforms to the gaussian distribution,

$$P_j(\mu') = \frac{1}{\sigma\sqrt{2\pi}}\, e^{\left(-\frac{(x_j - \mu')^2}{2\sigma^2}\right)}$$

the product of the probabilities becomes expressible as

$$P(\mu') = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\left(-\frac{\sum_{j=1}^{n}\left(\frac{x_j - \mu'}{\sigma}\right)^2}{2}\right)}$$

A maximum of that probability corresponds to a minimum of the expression in the exponent, which yields $\mu' = \dfrac{1}{n}\sum\limits_{j=1}^{n} x_j$, so that, for this gaussian distribution, the most probable value of the mean is just the average as established above. The method of <u>least squares</u> is a special case, highly practical and well established experimentally, of the more general method of maximum likelihood; we explain the former method in what follows, first for a simple case of one independent variable and then for multiple independent variables.

For two variables x, y related according to $y = m\,x + b$, in which parameters m and b are constant in all experiments, the mean $\overline{y}$ of y is related to the mean of $m\,x + b$, expressed as $\overline{(m\,x + b)}$, but, because evaluating a mean is a linear operation, the latter expression becomes $m\,\overline{x} + b$, so $\overline{y} = m\,\overline{x} + b$.

For analysis of data by means of methods of *simple* <u>linear regression</u> according to a criterion of <u>least sum of squares</u> of residuals to be valid, four criteria must apply.

- Each value of dependent variable y_i is related to an associated value of independent variable x_i according to a simple linear relation of form

$$y_i = \eta(\beta_j, x_i) + \varepsilon_i = \beta_0 + \beta_1\,x_i + \varepsilon_i$$

in which <u>expectation</u> function $\eta(\beta_j, x_i) = \eta_i$ of an independent variable x that takes a measured value x_i yields an expected value of dependent variable y that differs from a measured value y_i by ε_i; for linear regression, the derivative of dependent variable y with respect to any parameter β_i must contain neither other parameter β_j nor β_i itself.

- A measurement of y_i yields a response to an expectation function $\eta(\beta_j, x_i) = \beta_0 + \beta_1\,x_i$ plus a disturbance ε_i. All uncertainty ε_i of a measurement is associated with dependent variable y_i: each value x_i of independent variable is known exactly. For any fixed value of that independent variable x_i, there is a random component ε_i contributing to a value of dependent variable y_i.

- Each measurement of dependent variable y_i has the same <u>variance</u> σ^2 or <u>standard deviation</u> σ, associated with ε_i; these disturbances have zero mean and a common standard deviation σ. If ε_i conform to a <u>gaussian or normal distribution</u>, various statistical treatments become applicable, specifically according to a criterion of least sum of their squares. In practice, this criterion becomes relaxed upon inclusion of individual weighting of each measurement y_i. The method of least squares does not, however, require that the data have normally distributed (gaussian) error for its validity.

- Each error term ε_i is statistically independent of another error term ε_j; each measurement y_i is hence independent of, and uncorrelated with, any other measurement y_j.

The method of least squares hence provides an estimate of <u>maximum likelihood</u> when one might reasonably assume errors to be both independent and normally distributed with the same variance for each point; the latter criterion is relaxed with weighting methods.

A criterion of applicability of methods of linear and non-linear regression as we practise them here is either that all error is associated with a dependent variable or that a contribution to error of

a particular measurement from an independent variable is negligible. Although some data sets adhere rigorously to the former criterion, in general some error, apart from gross blunder, might be associated with a controlled variable. For *total least squares* or *orthogonal least squares* or *generalized least squares* to treat data in sets for which appreciable error is associated with both independent and dependent variables, methods are less well developed than conventional methods according to the above criteria, and generally involve iterative schemes that require initial estimates of parameters, but we outline one approach below. For actual data for which an error of measurement of independent variables be non-negligible, known and appreciably inconstant among data sets, one might to some extent take account of this condition through modification of error formally associated with a dependent variable according to weighting of data sets included in an analysis; such a procedure introduces a bias into estimates of parameters of regression.

   A minimum sum of squared residuals is an arbitrary criterion for use in fitting data: a general criterion is a minimum sum of magnitudes of residuals to power p, as in

$$\min \sum_{i=1}^{n} \left(\, | y_i - \eta(\beta_j, x_i) |^p \right)$$

in which y_i is measured value i of dependent variable y and $\eta(\beta_j, x_i) = \eta_i$ is a corresponding expected value calculated on a basis of measured values of independent variables x_i as regressors and fitting parameters β_j as their regression coefficients. Special cases include

- a method of *least* absolute values, for which $p = 1$,

- a method of *least squares*, for which $p = 2$, and

- a *minimax* method, for which $p \rightarrow \infty$.

The mode of a distribution corresponds to a method of least number or least sum of zero powers, for which $p = 0$, or the mode constitutes the most frequent result; the median corresponds to a method of least sum of magnitudes, with $p = 1$; the mean corresponds to a method of least squares, so $p = 2$, and the *midrange* to a least maximum or least sum of infinite powers, with $p \rightarrow \infty$. When errors conform to a symmetric exponential distribution of form

$$f(\varepsilon_i) = \frac{e^{\left(-\frac{|\varepsilon_i|}{\sigma} \right)}}{2\,\sigma}$$

the appropriate method is that of least sum of absolute values, according to this minimization,

$$\min \left(\sum_{i=1}^{n} |\varepsilon_i| \right)$$

For a normal distribution of errors, or double exponential distribution, according to

$$f(\varepsilon_i) = \frac{e^{\left(-\frac{\varepsilon_i^2}{2\sigma^2} \right)}}{2\,\pi\,\sigma^2}$$

the method of least sum of squares of errors, according to

$$\min\left(\sum_{i=1}^{n} \varepsilon_i^2\right)$$

is appropriate. For a uniform or rectangular distribution of errors, the minimax method is appropriate: according to the minimax method, the values of regression parameters are evaluated so as to minimize the largest deviation from regression, or the largest residual; the values of p practised for a minimax method lie typically in a range [6, 10]. One should not assume in general that the method of least squares, corresponding to $p = 2$ in the above general criterion for a minimum, is the only, or even the correct, approach to a reduction of particular data according to variables with continuous distributions, but, when errors associated with those variables are known to be normally distributed, this method is an appropriate choice.

To undertake linear regression according to a criterion of a least sum of squares of residuals in terms of constructs of linear algebra for a model that comprises a single independent variable and a single dependent variable with two parameters as coefficients of x^0 and x^1, we express the first of two normal equations (this name implies no connexion to a gaussian or normal distribution, but instead relates to a mathematic property of linear algebra),

$$\sum_j y_j = n\,\beta_0 + \beta_1\left(\sum_j x_j\right)$$

$$\sum_j x_j y_j = \beta_0\left(\sum_j x_j\right) + \beta_1\left(\sum_j x_j^2\right)$$

in compact form with matrices as

$$\mathbf{y} = \mathbf{x}\,\beta + \varepsilon$$

in which

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ ... \\ \varepsilon_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

here β_0 as a coefficient of x^0 corresponds to b as intercept of a fitted straight line on the ordinate axis, and β_1 as a coefficient of x corresponds to m as a slope of that fitted straight line; ε_i corresponds to a residual, that is a difference $y_i - \eta_i$, between y_i, measured value, and expectation η_i, calculated through fitted values of parameters β_i through $\eta_i = \beta_0 + \beta_1 x_i$. One can equally well consider $\mathbf{y}$, β and ε to be matrices having only one column or column vectors. This notation facilitates extension to treatment of data for multiple independent variables with, accordingly, coefficients β_i as parameters to be fitted numbering more than two; in that case, each set among n data includes values of k independent variables, requiring correspondingly fits to parameters up to β_k. If a fitted line be constrained to pass the origin, parameter β_0 for an ordinate intercept is correspondingly absent; parameters are then numbered in a range $1 \leq i \leq k$. For that fitted line for which the sum of squares of residuals ε_i is made minimum, those residuals represent the vertical distance between a fitted point y_i and its expectation $\eta(\beta_j, x_i)$ on that line.

In the case of fitting a quadratic model instead of a linear model, the normal equations become

$$\sum_j y_j = n\,\beta_0 + \beta_1\left(\sum_j x_j\right) + \beta_2\left(\sum_j x_j^2\right)$$

$$\sum_j x_j\, y_j = \beta_0\left(\sum_j x_j\right) + \beta_1\left(\sum_j x_j^2\right) + \beta_2\left(\sum_j x_j^3\right)$$

$$\sum_j x_j^2\, y_j = \beta_0\left(\sum_j x_j^2\right) + \beta_1\left(\sum_j x_j^3\right) + \beta_2\left(\sum_j x_j^4\right)$$

If the values of x_j span an interval $a \,..\, b$, the normal equations must be solved for parameters β_j, $j = 0 \,..\, 2$, in this linear system. The coefficients of these parameters have a form

$$\int_a^b x^{(j+k)}\,dx = \frac{b^j + k + 1 - a^{(j+k+1)}}{j+k+1}$$

that resembles the elements in a Hilbert matrix, discussed in section 6.116; the latter is notoriously ill conditioned, causing severe difficulties with rounding error in calculations involving real numbers.

We apologize in advance for prospective confusion relating to notation involving these coefficients: although a natural notation has numbering of coefficients β beginning at zero, so that β_0 is the coefficient of $x^0 = 1$ and corresponds simply to a quantity denoted b in a standard linear representation as in an equation $y = m\,x + b$ and as the ordinate intercept in a plot of a straight line, eventually in performing calculations involving linear-algebraic quantities matrix and vector we employ name β_1 for this purpose, because in *Maple* numbering of elements in a matrix or components of a vector must begin at unity not zero.

In rigorous statistical treatments many authors distinguish between *true* -- but unknowable -- quantities, denoted with greek letters such as β_0 for ordinate intercept and applicable to an entire population, and their *estimates*, denoted with roman letters such as b_0 correspondingly and applicable for a sample of a population; for practical purposes we try to employ symbols in a consistent manner but fail to conform totally to that convention. The values of parameters deduced from regression according to a criterion of least squares of residuals are the most reliable, consistent and least biased estimators of the unknowable true values of these parameters; this method of least squares of residuals is hence the most efficient unbiased estimator of regression coefficients.

In terms of these quantities as matrix or vector, our objective is to evaluate a vector β of estimators to minimize a sum of n squared residuals named χ^2 that is a function of the parameters β_j,

$$\chi^2 = SSE = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^T \varepsilon = (\mathbf{y} - \mathbf{x}\,\beta)^T\,(\mathbf{y} - \mathbf{x}\,\beta)$$

in which appended superscript T, as ^{T}, denotes a transpose of a matrix quantity. We employ χ^2 as synonymous with a sum of squared errors, which in practice become residuals, but χ^2 has another but related technical meaning in a statistical context. Differentiating this expression with respect to β, we obtain a condition for a minimum that

$$\mathbf{x}^T \mathbf{x}\,\beta - \mathbf{x}^T \mathbf{y} = 0$$

or, providing that matrix product $\mathbf{x}^T\mathbf{x}$ be not singular,

$$\beta = (\mathbf{x}^T\mathbf{x})^{(-1)}\mathbf{x}^T\mathbf{y}$$

in which the first factor corresponds to an inverse of a product of matrices; $\mathbf{x}^T$ is called a *design matrix*, and $\dfrac{1}{n}\mathbf{x}^T\mathbf{x}$ is called an *information matrix*. This formula provides a valid and practical means to evaluate estimators β as parameters of an optimal fit of a straight line to a given set of points, although practical methods avoid direct use of matrix inversion in favour of a numerically stable decomposition.

For a line constrained to pass the origin, such that intercepts on both abscissal and ordinate axes are zero, the uncertainty σ_y in a measurement of y is

$$\sigma_y = \sqrt{\dfrac{\displaystyle\sum_{j=1}^{n}(y_j - \beta_1 x_j)^2}{n-1}}$$

and the corresponding uncertainty in parameter β_1 as the only fitted parameter is

$$\sigma_1 = \dfrac{\sigma_y}{\sqrt{\displaystyle\sum_{j=1}^{n}x_j^2}}$$

A *robust* method for regression is insensitive to extraneous factors not under test, but still powerful through being sensitive to factors under test. A criterion involving a least sum of absolute values provides a more robust method than a least sum of squares of errors. Whereas a minimax method is thus not robust, if a distribution of error be other than normal, a minimum sum of magnitudes of residuals to power p with $1 < p < 2$ might provide an optimally robust line of regression to represent discrete data of a finite number. An alternative approach involves interactive weighting of data in an iterative process with $p = 2$ to yield estimates of the parameters; the customary indicators of goodness of fit might become unrealistic under these conditions. Another method of robust regression employs as criterion the least median of squares, but computations are intricate. As a theory of regression with $p = 2$ is well developed since the time of Legendre, and of Gauss who proposed a normal distribution as a means to justify this criterion and the method of least squares, this approach is conventional; with careful individual weighting of each data set, such as with $w_i = \dfrac{1}{\sigma_{y,i}^2}$ so that the relative weight of a particular datum is the reciprocal square of its standard deviation, one might improve the robustness of conventional fitting with retention of a scheme involving $p = 2$, so to diminish the sensitivity of results to possible outliers. According to a theorem attributed to Gauss and Markov, an unbiased estimator of a population parameter, based on sample observations in a linear combination, has minimum variance -- is thus best or optimal -- when the estimator is obtained according to a criterion of least sum of squares of errors or -- in practice -- residuals; this criterion hence yields the best linear unbiased estimates of parameters of a statistical population from ordered observations in a random

sample. Robust statistical procedures should be applied to only symmetrical distributions of ε_i or when severe outliers might occur infrequently, not to a bimodal or multimodal distribution of ε_i, or severely asymmetric such as log-normal. A robust estimate of a quantity and its uncertainty, as a variance, arises through a median absolute deviation, defined as the median of the magnitude of deviations from the median; a related criterion for a rejection of a prospective discordant datum or outlier x_o is a quantity $\dfrac{\left| x_o - \mathrm{median}(x_j) \right|}{M_{absdev}} > 5$, in which M_{absdev} is the median absolute deviation.

One condition of applicability of regression analysis is stated to be that an error of each measurement conforms to a common standard deviation or variance. As this condition is generally inapplicable to experimental measurements, we extend our analysis to take into account such variability of error by considering the significance of the input data. In particular, is each measurement of variables (x_i, y_i) equally reliable? We might acquire independent information on whether each measured value y_i is equally reliable, for instance by making multiple measurements of y_i for a particular and fixed value x_i; in that case, we might associate with each measurement of dependent variable y_i an uncertainty, best expressed in a form standard deviation σ_i, which can form a basis of an error bar for that point in a plot. When such information be known, we ought to employ it in our fit of all available data, so that less reliable data that thus weigh less heavily influence its parameters less adversely than other data included within that fit. The data in each set input into a procedure for regression should comprise accordingly three numbers (x_i, y_i, w_i) with the latter quantity being a *weight* of that measurement with respect to the entire set of data; in practice, this weight w_i is best taken simply as $\dfrac{1}{\sigma_i^2}$, or the reciprocal variance, of a particular measurement, as mentioned above. There is no necessity, for instance, to normalize a sum of such weights w_i to a number n of observations as we can compensate for an arbitrary value of this sum in subsequent formulae. In the lack of a precise measurement of σ_i, we might estimate it for each individual y_i; we otherwise take each weight w_i as unity if there be no reason to expect that such weights be not constant for an entire set of data.

We assume a square, diagonal weight matrix **w**,

$$
\begin{bmatrix}
w_1 & 0 & 0 & 0 & 0 \\
0 & w_2 & 0 & 0 & 0 \\
0 & 0 & w_3 & 0 & 0 \\
0 & 0 & 0 & w_{...} & 0 \\
0 & 0 & 0 & 0 & w_n
\end{bmatrix}
$$

of order n being the number of observations to be fitted, with non-zero entries accordingly along only the principal diagonal and numbering n, one such entry associated with each y_i. According to a criterion of least squares, on differentiation of

$$
\chi^2 = SSE = \sum_{i=1}^{n} \varepsilon_i^2 = \varepsilon^T \varepsilon = (\mathbf{y} - \mathbf{x}\beta)^T \mathbf{w} (\mathbf{y} - \mathbf{x}\beta)
$$

$$
= \mathbf{y}^T \mathbf{w}\, \mathbf{y} - \beta^T \mathbf{x}^T \mathbf{w}\, \mathbf{y}
$$

the normal equations in matrix form become

$$(\mathbf{x}^T \mathbf{w} \mathbf{x}) \beta - \mathbf{x}^T \mathbf{w} \mathbf{y} = 0$$

and weighted estimators become accordingly

$$\beta = (\mathbf{x}^T \mathbf{w} \mathbf{x})^{(-1)} \mathbf{x}^T \mathbf{w} \mathbf{y}$$

extending and replacing matrix relations lacking $\mathbf{w}$ presented above; if weight matrix $\mathbf{w}$ be a unit matrix, these formulae become equivalent to those derived in section 8.304.

Under these conditions, the matrices that yield *determinants* to generate values of parameters m and b estimated for a simple linear relation with one regressor x become,

$$m_d = \begin{vmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} w_i y_i \\ \sum_{i=1}^{n} w_i x_i & \sum_{i=1}^{n} w_i x_i y_i \end{vmatrix},$$

$$b_d = \begin{vmatrix} \sum_{i=1}^{n} w_i y_i & \sum_{i=1}^{n} w_i x_i \\ \sum_{i=1}^{n} w_i x_i y_i & \sum_{i=1}^{n} w_i x_i^2 \end{vmatrix},$$

$$\Delta_x = \begin{vmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} w_i x_i \\ \sum_{i=1}^{n} w_i x_i & \sum_{i=1}^{n} w_i x_i^2 \end{vmatrix}$$

with

$$m = \frac{m_d}{\Delta_x} \quad \text{and} \quad b = \frac{b_d}{\Delta_x} .$$

For such a weighted fit of a linear model, the uncertainties in the resulting parameters are, for m_d,

$$\beta_m = \sqrt{\frac{\sum_{i=1}^{n} w_i}{\Delta_x}},$$

and for b_d,

$$\beta_b = \sqrt{\frac{\sum_{i=1}^{n} w_i x_i^2}{\Delta_x}} .$$

For cases in which weights of values of dependent variable are lacking, all w_i become unity.

Some relations between the independent and dependent variables that are non-linear in parameters to be evaluated through regression involving numerous data might be converted into a pseudo-linear form. For instance, a formula that arises in chemical kinetics for exponential decay

of first kinetic order has a form

$$c = c_0 \, \mathrm{e}^{(-k\,t)}$$

that is readily transformed to

$$\ln(c) = \ln(c_0) - k\,t$$

By analogy with standard linear form $y = m\,x + b$, y corresponds to $\ln(c)$, x corresponds to t, b corresponds to $\ln(c_0)$ and m corresponds to $-k$; fitting $\ln(c)$ as a function of t thus yields a straight line with slope $-k$ and ordinate intercept $\ln(c_0)$: this process constitutes pseudo-linear regression. If data -- measured values of c -- be imperfect -- as is inevitably the case, and if standard deviations of those measured values be roughly independent of c, fitting according to a pseudo-linear model without taking into account the effect of transformation on those errors in c introduces a distortion or bias of evaluated parameters. For this reason, further weighting of data becomes necessary to compensate for this transformation.

Either as a result of such a transformation or when a linear model involves variables not in direct proportionality, such as y being proportional to x in a standard linear form $y = m\,x + b$, a proper weighting of data is necessary to achieve the utmost significance of evaluated parameters and an absence of bias in their evaluation. For instance, when parameters occur linearly whereas variables occur in other than simple proportional form, such as with a model $\dfrac{1}{y} = m\,x + b$, an error in y, measured according to a standard deviation of each point y_i, is not equivalent to the same standard deviation applied to $\dfrac{1}{y}$. If one take into account transformation of a formula from a form $f(Y) = m\,f(X) + b$ to a form pseudo-linear in variables x and y, as in $y = m\,x + b$, a proper weighting of each data item thus becomes its weight, according to the reciprocal variance of Y, divided by a derivative $\left(\dfrac{\partial}{\partial y} f(Y)\right)^2$ evaluated for each value of y, or

$$w_i = \frac{1}{\sigma_{Y,i}^2 \; \left(\dfrac{\partial}{\partial y} f(Y)\right)^2 \Big|_{Y = Y_i}}$$

No corresponding account of a relation between $f(X)$ and x arises because a fundamental assumption of regression, as customarily applied and as reflected in our treatment here, is that all error is associated with a measurement of dependent variable y or Y.

Such pseudo-linear, or transformable linear, regression is an alternative procedure to non-linear regression, but procedures present in section group 8.4 enable one to perform directly non-linear regression practically as readily as linear regression. A disadvantage of pseudo-linear regression is that a standard error in a parameter of a transformed fitting equation bears no simple relation to original parameters. For instance, a standard error of $\ln(c_0)$ in a case discussed above has no direct interpretation in terms of c_0, and becomes asymmetric for c_0.

Although a basic premise of regression is that the measurements of a dependent variable be entirely uncorrelated, a weight matrix can take into account such correlation if it exist; under these

conditions, the elements of this weight matrix off the principal diagonal might assume non-zero values, but the matrix remains symmetric. The planning and conduct of experiments to preclude such correlated measurements are recommended.

As a practical approach to the practice of *total least squares*, we consider fitting of n unweighted data in a set supposed to obey approximately a linear dependence of variable y on independent variable x of which both variables are subject to error of measurement; for this purpose we assume a fitting equation $y = m\,x + b$, as above, and calculate the following sums, first for the means of the measured values of x_j and y_j separately,

$$x_{mean} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

$$y_{mean} = \frac{1}{n} \sum_{j=1}^{n} y_j$$

then for their variances,

$$S_{xx} = \left(\sum_{j=1}^{n} x_j^2 \right) - \frac{\left(\sum_{j=1}^{n} x_j \right)^2}{n} = \sum_{j=1}^{n} x_j^2 - n\, x_{mean}^2$$

$$S_{yy} = \left(\sum_{j=1}^{n} y_j^2 \right) - \frac{\left(\sum_{j=1}^{n} y_j \right)^2}{n} = \sum_{j=1}^{n} y_j^2 - n\, y_{mean}^2$$

and their covariance,

$$S_{xy} = \left(\sum_{j=1}^{n} x_j\, y_j \right) - \frac{\left(\sum_{j=1}^{n} x_j \right)\left(\sum_{j=1}^{n} y_j \right)}{n} = \sum_{j=1}^{n} x_j\, y_j - n\, x_{mean}\, y_{mean}$$

and Pearson's coefficient of linear correlation for a sample,

$$r = \frac{S_{xy}}{\sqrt{S_{xx}\, S_{yy}}}$$

The coefficient of linear correlation for a population is denoted ρ, and the square of the correlation coefficient for a sample is called the coefficient of determination, so

$$r^2 = \frac{S_{xy}^2}{S_{xx}\, S_{yy}}.$$

For data with negligible error in measurement of independent variable x, these formulae are based on a minimum of a sum of squared residuals, $\sum_{j=1}^{n} \varepsilon_j^2$, in which $\varepsilon_j = y_j - (m\, x_j + b)$ with values for the slope of the best fitting straight line as simply

$$m = \frac{S_{xy}}{S_{xx}}$$

and the ordinate intercept as

$$b = y_{mean} - m\, x_{mean}$$

For variable w dependent on independent variables $x, y, z, ...$, the variance of w becomes approximately

$$\sigma_w^{\,2} = \sigma_x^{\,2}\left(\frac{\partial}{\partial x}w\right)^2 + \sigma_y^{\,2}\left(\frac{\partial}{\partial y}w\right)^2 + ... + 2\,\sigma_{xy}^{\,2}\left(\frac{\partial}{\partial x}w\right)\left(\frac{\partial}{\partial y}w\right) + ...$$

in which appear variances $\sigma_x^{\,2}, \sigma_y^{\,2}, ...$ and covariances $\sigma_{xy}^{\,2}, ...$ This relation is called the equation of propagation of error.

For data of which both variables are subject to experimental error, ε_j for y_j and δ_j for x_j, the criterion for a best fit according to the least squares of errors involves the square of the statistical distance of the point $(x_j + \delta_j, y_j + \varepsilon_j)$ from the point on the line associated with x_j. In that case but still on assumptions that the *population variances* of the errors $\sigma_{\delta\delta}$ and $\sigma_{\varepsilon\varepsilon}$ are equal and that their covariance $\sigma_{\delta\varepsilon} = 0$, we define an additional statistical quantity,

$$\tau = \frac{S_{xx} - S_{yy}}{2\,S_{xy}}$$

with which we calculate the slope as

$$m = -\tau + \sqrt{1 + \tau^2}.$$

The value of the ordinate intercept b remains according to the formula above to calculate this new value of m. The population variance is the mean squared distance of all measurements from the population mean, so

$$\sigma_{\delta\delta} = \frac{1}{n}\sum_{j=1}^{n}(x_j - x_{mean})^2$$

$$\sigma_{\varepsilon\varepsilon} = \frac{1}{n}\sum_{j=1}^{n}(y_j - y_{mean})^2$$

$$\sigma_{\delta\varepsilon} = \frac{1}{n}\sum_{j=1}^{n}(y_j - y_{mean})(x_j - x_{mean})$$

If $\sigma_{\delta\delta} \neq \sigma_{\varepsilon\varepsilon}$, so that $\dfrac{\sigma_{\delta\delta}}{\sigma_{\varepsilon\varepsilon}} = \upsilon$, but $\sigma_{\delta\varepsilon} = 0$, the slope becomes

$$m = \frac{-\tau + \sqrt{\upsilon + \tau^2}}{\upsilon}$$

with

$$\tau = \frac{S_{xx} - \upsilon\, S_{yy}}{2\,S_{xy}}$$

If in addition $\sigma_{\delta\varepsilon} \neq 0$, the slope becomes

$$m = \frac{S_{xy} - \lambda\,\sigma_{\delta\varepsilon}}{S_{xx} - \lambda\,\sigma_{\delta\delta}}$$

in which λ is the least root of the determinantal equation

$$\det\left(\begin{bmatrix} S_{xx} & S_{xy} \\ S_{xy} & S_{yy} \end{bmatrix} - \lambda \begin{bmatrix} \sigma_{\delta\delta} & \sigma_{\delta\varepsilon} \\ \sigma_{\delta\varepsilon} & \delta_{\varepsilon\varepsilon} \end{bmatrix}\right) = 0$$

For this purpose $\sigma_{\delta\delta}$, $\delta_{\varepsilon\varepsilon}$ and $\sigma_{\delta\varepsilon}$ for the population errors must be known.

Because the neglect of an error in the independent variable causes a significant distortion of coefficient *m* -- generally to increase its value -- and correspondingly *b*, one must establish, before undertaking a fit of experimental data, whether error is significant in both variables, so that a treatment such as that above is practised when required. The extent of distortion increases with increasing ratio $\dfrac{S_{yy}}{S_{xx}}$ and for *r* decreasing from unity.

multiple linear regression

To treat cases of multiple independent variables, or a single independent variable in a model containing distinguishable terms such as those with various exponents as in a polynomial, or a combination of these, in all cases we require that a model be linear in parameters to be fitted; an explicit specification of that linear condition is that a derivative of the dependent variable with respect to a parameter, or, more precisely, of a residual between a measured value and a calculated value of that dependent variable with respect to a parameter, involve no parameter to be fitted. In a matrix form, the regressor matrix **x** for multiple linear regression simply has multiple columns, one for each regressor, plus a column for unity if a constant term be required, and other matrices or vectors have corresponding dimensions, but the relations have the same form as those derived above for simple weighted linear regression with only one regressor; for this reason we refrain from repeating these relations here. In a case of two independent variables a plot of the derived regression equation exhibits a plane, the sum of squares of vertical deviations from the surface of which are minimized, but more numerous independent variables preclude such a plot or direct geometric interpretation; a plot of residuals versus some characteristic of a data set, such as the number of the data point in the list of input, is, however, invariably practicable.

Our task is to investigate a single functional dependence of the results of experiments on multiple controlled factors $x_1, x_2, .. x_k$ as *regressors*, *factors* or *predictor variables* of predetermined number that are adjustable at varied levels within an experimental region or sample space of an euclidean space of dimension $k + 1$; each x_i here denotes a separate quantity of a distinct kind, such as pressure or temperature, not a separate measurement of the same quantity. For each measurement, a result is a value of dependent variable $y = \eta(\beta_0 .. \beta_k, x_1 .. x_k) + \varepsilon$, or *response variable* or uncontrolled factor, containing a non-stochastic part $\eta(\beta, x)$ involving each independent variable x_i, $1 \leq i \leq k$, as a regressor and its coefficient β_i, $0 \leq i \leq k$, as a parameter, and a random or stochastic part ε that entails a disturbance to a response for each measurement, for measurements numbering $1 .. n$. Parameters β_i are coefficients of regressors, numbering *k*, in an objective function, or model, of form

$$y = \beta_0 + \left(\sum_{i=1}^{k} \beta_i\, x_i \right)$$

for each value of dependent variable y. Among $k + 1$ parameters, one parameter β_0 might be simply a coefficient of unity, corresponding in geometric terms to an intercept on the axis for the dependent variable in a two-dimensional plot in simple linear univariate regression; if the surface, or hypersurface, as a graph of the objective function be constrained to pass the origin, β_0 becomes eliminated from the set of parameters, leaving k regressors. With only a single variable, x_1, we revert to simple linear regression. With n experimental measurements, $n > k + 1$, there are thus $n - k - 1$ degrees of freedom. For a purpose of undertaking analysis of regression, the values of regressors are known and the values of regression coefficients or parameters are unknown; for that reason, we express an expectation function in terms of the parameters as $\eta(\beta_0 \,..\, \beta_k)$, with the regressors apparently neglected, or alternatively as $\eta(\beta_0 \,..\, \beta_k; x_1 \,..\, x_k)$ to indicate a parametric relation. The square of a difference between a measurement y_i and that value expected from the non-stochastic part $\eta(\beta_1 \,..\, \beta_k)$ is a square of a vertical distance between a data vector and a corresponding point on a response surface, or hypersurface which is a surface in hyperspace for $k > 2$; in this way linear regression becomes linked to euclidean geometry and to linear algebra. If for each regressor $x_i = x_1^{\,i}$ for $i \geq 1$, this multivariate linear regression becomes a special case -- polynomial regression; combinations of separate regressors x_i, x_j and these quantities to various powers are naturally practicable through an appropriate construction of a design matrix, but the coefficients of such regressors as fitting parameters become highly correlated unless orthogonal functions are constructed to serve as these polynomials.

For n measurements of which we observe value o_j with standard deviation σ_j and expect value e_j, we define a quantity χ^2,

$$\chi^2 = \sum_{j=1}^{n} \left(\frac{o_j - e_j}{\sigma_j} \right)^2 ;$$

for a satisfactory agreement between the observed and the expected values, $\chi^2 \sim n$. For f degrees of freedom and c constraints, $f = n - c$, and the expected order of χ^2 is f. If the measurements be the numbers of events in each of multiple bins, the standard deviation is just $\sqrt{e_j}$, and

$$\chi^2 = \sum_{j=1}^{n} \frac{(o_j - e_j)^2}{e_j} .$$

The reduced χ^2 for f degrees of freedom is $\chi_r^2 = \dfrac{\chi^2}{f}$.

The normal equations for multiple linear regression are generated in an obvious way from those for single linear regression, for instance with two independent variables similarly as indicated above for the fitting of a quadratic model,

$$\sum_j y_j = n\, \beta_0 + \beta_1 \left(\sum_j x_{1,j} \right) + \beta_2 \left(\sum_j x_{2,j} \right)$$

$$\sum_j x_{1,j}\, y_j = \beta_0 \left(\sum_j x_{1,j} \right) + \beta_1 \left(\sum_j x_{1,j}{}^2 \right) + \beta_2 \left(\sum_j x_{1,j}\, x_{2,j} \right)$$

$$\sum_j x_{2,j}\, y_j = \beta_0 \left(\sum_j x_{2,j} \right) + \beta_1 \left(\sum_j x_{1,j}\, x_{2,j} \right) + \beta_2 \left(\sum_j x_{2,j}{}^2 \right)$$

but a matrix formulation is preferable. According to a criterion of a least sum of squared weighted residuals, we form this quantity directly from its definition in which χ^2 is a function of all parameters β_j.

$$\chi(\beta_j)^2 \;=\; SSE = \sum_{j=1}^{n} w_j \left(y_j - \left(\beta_0 + \left(\sum_{i=1}^{k} \beta_i\, x_{i,j} \right) \right) \right)^2$$

Although, for a purpose of applying this criterion to data in one or other set, one might compose a procedure on a basis of exactly those formulae in terms of only matrix $\mathbf{x}$ and vectors $\underline{y}$ and $\underline{w}$ presented above, such a procedure can yield poor results. A source of this problem is explicit inversion of matrix product $\mathbf{x}^T \mathbf{w}\, \mathbf{x}$, which is proportional to the information matrix; in a case of a design matrix $\mathbf{x}$ for polynomial regression, the particular matrix to be inverted strongly resembles an Hilbert matrix with a large condition number, of which a determinant can evaluate to a minute magnitude, as explained in section 6.116: during such inversion of a matrix containing real numbers as floats as its elements, arithmetical operations produce a serious loss of precision. To circumvent such problems of an ill conditioned matrix or system, either an alternative mathematical approach might be sought, or the precision of the calculation might be increased through increasing the number of decimal digits.

An alternative approach to multiple linear regression that is amenable to implementation on a computer is stepwise multiple regression in which, from a set of data and the corresponding factors or regressors or predictor variables, a selection at each step or stage of the analysis is made of the independent variables of the one that produces the greatest decrease of the unexplained variation in the dependent or response variable; this process continues until every variable has become included in the multiple regression or until no further decrease of the unexplained variation is discernible. A complementary approach involves inclusion of every feasible variable at the beginning of an analysis and the successive elimination of that variable for which indicators of goodness of fit show a lack of significance. With the duration of computation for data sets of at least moderate size becoming almost negligibly small, the latter approach is highly practical.

criteria of goodness of fit for linear regression

Among quantities to evaluate a goodness of fit, the primary quantity is the sum of weighted squared errors, or in practice a sum of weighted squared residuals, denoted SSE here and commonly χ^2 elsewhere, that we employ to derive the normal equations. Rather than the mean weighted squared error $\dfrac{SSE}{n}$ or $\dfrac{\chi^2}{n}$, or an estimate of this quantity, a more statistically meaningful quantity, the square of the standard error of the estimate about the regression relation, is denoted σ^2 and is related to SSE through a formula

$$\sigma^2 = \frac{SSE}{n-f}$$

in which sets of data number n and the number of components in vector β for the fitted parameters is f; the number of <u>degrees of freedom</u> is thus $n - f$. This effective mean weighted squared error is also the <u>variance</u> of a fit; the magnitude of its square root σ is the <u>standard deviation</u> of a fit and represents an expected error of any measured value of dependent variable y_i, or its counterpart for a weighted fit, or a value of y predicted with the equation of best fit for any value x. We calculate this variance of a fit in terms of explicit algebraic quantities,

$$\sigma^2 = \frac{n}{(n-2)\left(\sum\limits_{i=1}^{n} w_i\right)} \sum\limits_{i=1}^{n} w_i\, (y_i - m\,x_i - b)^2$$

in which appears a factor $\dfrac{n}{\sum\limits_{i=1}^{n} w_i}$ to take into account that weights might not be <u>normalized</u> (to

sum to n). A weighted fit yields a *normalized variance* or *normalized standard deviation*.

The usage of χ^2 is this context differs from that associated with <u>Pearson's</u> <u>test statistic</u> according to which χ^2 is a sum of *ratios* of squared deviations -- i.e., squares of differences observed $-$ calculated values -- and calculated values,

$$\chi^2 = \sum\limits_{j=1}^{n} \frac{(x_{o,j} - x_{c,j})^2}{x_{c,j}}$$

over the range of the data.

For an unweighted fit, either quantity variance σ^2 or standard deviation σ is a measure of the overall success of that fit. In contrast, for a weighted fit such a variance or standard deviation is instead a measure of the care in assigning weights: if weights be correctly assigned, the weighted standard deviation or variance is exactly unity. With such a standard deviation σ of a weighted fit in a range [0.9, 1.1], the consequences are likely negligible; if one obtain from a fit of weighted data a value of σ much outside that range, one should reassess input data, especially standard deviations of each individual datum. The variance is also expressible as

$$s^2 = \frac{1}{n-1} \sum\limits_{i=1}^{n} (x_i - \overline{(x_i)})^2$$

in which $\overline{(x_j)}$ denotes the mean of those values for a particular factor or independent variable and the sum runs over all values of that particular factor; the covariance arising from two separate factors j and k is analogously expressible as

$$cov_{j,k} = \frac{1}{n-1} \sum\limits_{i=1}^{n} (x_{i,j} - \overline{(x_i)})\, (x_{i,k} - \overline{(x_k)})$$

with $j \neq k$. The standard deviation s for factor x is thus $s = \sqrt{s^2}$, the positive square root of the

variance.

As an outcome of regression analysis, we clearly require other indications, or indicators, of a quality of the fit, such as statistical significance of derived values of individual parameters β_i. For this purpose we form a *dispersion matrix* $\mathbf{v}$ as the product of σ^2 with a matrix containing as elements the coefficients of β_i, which is proportional to an inverse of the *information matrix*.

$$\mathbf{v} = \sigma^2 \, (\, \mathbf{x}^T \, \mathbf{w} \, \mathbf{x} \,)^{(-1)}$$

The variances of parameters β_j are the diagonal elements of this dispersion matrix.

For simple linear regression, estimated *variances of parameters* are, for $\beta_1 = m$ as slope,

$$\sigma_m^{\,2} = \frac{\displaystyle\sum_{i=1}^{n} w_i}{\Delta_x}$$

and for $\beta_0 = b$ as intercept on the ordinate axis,

$$\sigma_b^{\,2} = \frac{\displaystyle\sum_{i=1}^{n} w_i \, x_i^{\,2}}{\Delta_x} \; ;$$

the positive square roots of these variances are the respective *estimated standard error*s of these parameters.

Correlations arise both between variables, here x and y, and between parameters, here $\beta_0 = b$ corresponding to β_1 in the vector of parameters and $\beta_1 = m$ corresponding to β_2 in that vector. To indicate an extent of correlation between these parameters, we define a <u>covariance</u>, which is a measure of the way in which two observable quantities vary together,

$$\sigma_{m\,b}^{\,2} = - \, \frac{\displaystyle\sum_{i=1}^{n} w_i \, x_i}{\Delta_x}$$

When the error in one quantity or parameter is positive, for the error in another quantity or parameter to be likely also positive implies a positive covariance. In terms of matrices, the variances of parameters lie along the principal diagonal of dispersion matrix $\mathbf{v}$, whereas the off-diagonal elements are covariances: explicitly $v_{i,j}$ is a measure of linear association of β_i and β_j ; if β_i and β_j be independent, $v_{i,j} = 0$, but the converse is not necessarily true; in a case of a single regressor, $v_{i,j} = \sigma_{m\,b}^{\,2}$.

Although we might compare a covariance with associated variances, we form another useful matrix $\mathbf{c}$ of which each element off the principal diagonal becomes a coefficient of correlation between parameters, indicating an extent to which two errors are correlated. For a general case of simple linear regression, there are two parameters, m and b; their *product-moment* <u>coefficient of correlation</u> conforms to this relation:

$$c_{mb} = \frac{\sigma_{mb}^2}{(\sigma_m^2 \, \sigma_b^2)^{\left(\frac{1}{2}\right)}}$$

The range of such a correlation coefficient is $[-1, 1]$; a positive value of correlation coefficient indicates that a positive error in one parameter is likely to be accompanied by a a positive value of another parameter, just as for a positive covariance. For a corresponding [matrix of correlation coefficients](#) for multiple regressors, we form its elements on dividing covariance $cov_{i,j} = v_{i,j},\ i \neq j$, by the square root of a corresponding product of variances:

$$c_{i,j} = \frac{v_{i,j}}{\sqrt{v_{i,i}\, v_{j,j}}} = c_{j,i}$$

For single or multiple linear regression, unity, according to $\dfrac{v_{i,i}}{\sqrt{v_{i,i}\, v_{i,i}}}$, appears along the principal diagonal because each coefficient is perfectly correlated with itself; for simple linear regression, the value c_{mb} appears in either off-diagonal position, because this matrix of order 2 is symmetric. We form a further matrix $\mathbf{u}$ with reciprocal square root of elements of matrix $\mathbf{v}$ along the diagonal, defined as

$$u_{i,j} = \frac{\delta_{i,j}}{\sqrt{v_{i,i}}}$$

with $\delta_{i,j} = 1$ if $i = j$, and 0 otherwise. In matrix notation,

$$\mathbf{u} = diag(\frac{1}{\sqrt{v_{i,i}}}, i = 1 .. f)$$

The correlation matrix becomes

$$\mathbf{c} = \mathbf{u\,v\,u}.$$

Element $c_{i,j}$ of that matrix that represents a coefficient of correlation between parameters i and j is sensitive to the mean of values of the independent variable: for example, for data comprising these four values $y_i = 2.3, 2.8, 7.9$ and 11 and four corresponding values $x_i = i,\ 1 \leq i \leq 4,\ c_{m,b} = -0$ .913; with the same values of y_i but with x_i shifted to $101 \leq i \leq 104$, $c_{m,n}$ becomes -0.99994, and with x_i shifted to $1001 \leq i \leq 1004$, $c_{m,b}$ becomes $-.9999994$. For this reason, this statistic seems to be a reliable indicator of goodness of fit only when the standard deviation of values of an independent variable is comparable with their mean.

Two other indicators of goodness of a linear fit are *correlation coefficient of a sample* and *F statistic*. The former indicates whether increasing values of x_i are likely to be associated with increasing or decreasing values of y_i; a preferable symbol for this correlation might be r_{xy} but, to avoid a complication of notation when this quantity appears in the right side of an equation or an assignment, we use simply r. The quantity r^2, called a *sample squared correlation* or *sample coefficient of determination*, measures the strength of association between variables and expresses the proportion of total variation in the levels of dependent variable y that is explained through a

variation of x for the particular sample on which measurements are made, and is thus an estimate of the coefficient of determination of an entire population; being such a proportion, the range of its values is [0, 1], between limits of entire lack of association between variables and a perfect association, but the latter value still implies no causal relation. To evaluate the correlation coefficient r that is $\sqrt{r^2}$, we define a further determinant, of form equivalent to that of Δ_x but with x_i replaced with y_i:

$$\Delta_y = \begin{vmatrix} \sum_{i=1}^{n} w_i & \sum_{i=1}^{n} w_i y_i \\ \sum_{i=1}^{n} w_i y_i & \sum_{i=1}^{n} w_i y_i^2 \end{vmatrix}$$

The *coefficient of linear correlation of a sample* or *distribution*, which measures correlation between variables, is then

$$r = \frac{\sum_{i=1}^{n} w_i \left(x_i - \dfrac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \right) \left(y_i - \dfrac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \right)}{\sqrt{\left(\left(\sum_{i=1}^{n} w_i x_i^2\right) - \dfrac{\left(\sum_{i=1}^{n} w_i x_i\right)^2}{\sum_{i=1}^{n} w_i} \right) \left(\left(\sum_{i=1}^{n} w_i y_i^2\right) - \dfrac{\left(\sum_{i=1}^{n} w_i y_i\right)^2}{\sum_{i=1}^{n} w_i} \right)}}$$

which simplifies to

$$r = \frac{\left(\sum_{i=1}^{n} w_i x_i y_i\right) - \dfrac{\left(\sum_{i=1}^{n} w_i x_i\right)\left(\sum_{i=1}^{n} w_i y_i\right)}{\sum_{i=1}^{n} w_i}}{\sqrt{\left(\left(\sum_{i=1}^{n} w_i x_i^2\right) - \dfrac{\left(\sum_{i=1}^{n} w_i x_i\right)^2}{\sum_{i=1}^{n} w_i} \right) \left(\left(\sum_{i=1}^{n} w_i y_i^2\right) - \dfrac{\left(\sum_{i=1}^{n} w_i y_i\right)^2}{\sum_{i=1}^{n} w_i} \right)}}$$

or succinctly,

$$r = \frac{m_d}{\sqrt{\Delta_x \Delta_y}} \; .$$

Expressed in terms of means of particular factors and neglecting here the weighting coefficients w_i, this correlation coefficient $r_{j,k}$ between two separate factors x_j and x_k, with $j \neq k$, becomes

$$r_{j,k} = \frac{\text{cov}(j,k)}{\sqrt{s_j^2\, s_k^2}} = \frac{\sum\limits_{i=1}^{n} (x_{i,j} - \overline{(x_j)})\,(x_{i,k} - \overline{(x_k)})}{\sqrt{\left(\sum\limits_{i=1}^{n} (x_{i,j} - \overline{(x_j)})^2\right)\left(\sum\limits_{i=1}^{n} (x_{i,k} - \overline{(x_k)})^2\right)}}$$

with the sums running over all values of x_j and x_k. The correlation matrix has accordingly values unity along the principal diagonal and elsewhere values $r_{j,k}$ according to that formula.

The factors within the square root in the denominators of both formulae are proportional to variances of weighted values of the independent and dependent variables, respectively; the quantity in the numerator is proportional to a corresponding covariance. Including the weights, we express this in matrix form as

$$r^2 = (\mathbf{x}^T \mathbf{w}\, \mathbf{y})^T\, ((\mathbf{x}^T \mathbf{w}\, \mathbf{x})\,(\mathbf{y}^T \mathbf{w}\, \mathbf{y}))^{(-1)}\, (\mathbf{x}^T \mathbf{w}\, \mathbf{y})$$

or alternatively for unweighted data as

$$r^2 = (\beta^T \mathbf{x}^T \mathbf{y} - n\,\overline{y}^2)\,(\mathbf{y}^T \mathbf{y} - n\,\overline{y}^2)^{(-1)}$$

in which $\overline{y}$ is the mean of y_i; $\overline{y} = \dfrac{\sum\limits_{i=1}^{n} y_i}{n}$ for unweighted data or $\overline{y} = \dfrac{\sum\limits_{i=1}^{n} w_i y_i}{\sum\limits_{i=1}^{n} w_i}$ for weighted data. An interpretation of r is the fraction of the total variance that the model explains. In some cases r^2, also known as *coefficient of determination*, might assume even a negative value, such as when an objective formula fits the data worse than a horizontal line at $\overline{y}$, which is the mean value of y. For almost all data collected in chemical or physical experiments involving large samples, $|r|$ has a value near unity, and is practically useless as a direct measure of goodness of fit of these data to a linear relation.

This sample correlation coefficient r, or its square r^2, is neither a measure of the slope of a regression line nor a measure of the appropriateness of a linear model.

For a straight line as model, r takes positive values if $m > 0$, i.e. for a positive slope, and negative values if $m < 0$, thus for a negative slope. The range of values of r is $[-1, 1]$, each extreme implying perfect correlation, whereas with $r = 0$ there is no correlation between the variables; magnitudes of r near unity hence indicate that a model might be satisfactory. As specific cases to exemplify the limiting cases of this indicator, with all points exactly on a fitted straight line, if the line have a positive slope, $r = 1$, or if the line have a negative slope, $r = -1$. In contrast, for four points at respective corners of a square and fitted to a straight line, $r = 0$. According to a method to calculate r given above from r^2, only the magnitude of r, in a range $[0, 1]$, becomes evaluated: one loses information about its sign.

In general, statistical parameters gain enhanced accuracy through use of $\overline{x}$ and $\overline{y}$, hence with

centred data, relative to directly calculated sums of values of variables, because the latter sums tend to involve only positive quantities and thus lead to comparably large magnitudes that might cause loss of precision on subtraction of large positive quantities. A disadvantage of a use of $\bar{x}$ and $\bar{y}$ is either that all data must be collected before a mean of x_i and y_i can be calculated or that a running mean must be calculated throughout a collection of data with constant upgrading. A line fitted according to a least sum of squares of errors contains a centre of gravity $(\bar{x}, \bar{y})$ of fitted data.

The *F statistic* is related to a F distribution, but, for practical purposes, we can view it as a further test of significance of one model for comparison with another model; if we assume only one model, such as a straight line, to be pertinent for a particular set of data, this quantity has limited use, but, in view of subsequent applications with multiple independent variables or polynomial functions, we define it here as

$$F = \frac{r^2(n-f)}{f(1-r^2)},$$

in which r^2 is again the coefficient of determination and $n-f$ is the number of degrees of freedom. The *F* distribution is formally applicable to data provided that the residuals conform to a gaussian distribution and that all sample observations are independent. Even with a straight line as a fitting model or objective function, one might wonder whether one should include a possibility of a non-zero value of constant term *b* in a fitting model, pertaining to a finite intercept on the ordinate axis for a plot of that line; if σ_b were comparable with the magnitude of *b*, fits of the same data with and without *b* as a fitted parameter would yield disparate values of *F*, the larger of which would indicate a statistically preferable model. In matrix notation a formula for the *F* statistic is

$$F = \frac{n-f}{n} \left(\frac{\mathbf{y}^T . \mathbf{w} . \mathbf{y} \ -1}{\mathbf{y}^T . \mathbf{w} . \mathbf{y} - ((\mathbf{x}^T . \mathbf{w} . \mathbf{x})^{(-1)} . \mathbf{x}^T . \mathbf{w} . \mathbf{y})^T . \mathbf{x}^T . \mathbf{w} . \mathbf{y}} \right)$$

in terms of the same matrices as defined above.

With a division of the variance between explained and unexplained components, this *F* statistic represents a ratio,

$$F = \frac{variance\ explained\ by\ regression}{variance\ unexplained}$$

In terms of a null hypothesis, according to which the observed data are characteristic of a merely random occurrence, and an alternative hypothesis, according to which the data are uncharacteristic of a merely random occurrence, under the null hypothesis one expects values of *F* near unity; typical values of *F* for fits of precise chemical or physical data with a theoretically appropriate objective function attain large magnitudes, consistent with strongly correlated independent and dependent variables: under these conditions r^2 approaches unity, hence $(1-r^2)$ approaches zero, and the expression defining *F* above clearly attains large magnitudes.

Another approach to testing the significance of a parameter obtained from replicate observations is called analysis of variance, according to which means are compared through their

corresponding sample variances; in *Maple*'s superseded package stats and its replacement package Statistics that is a collection of tools for mathematical statistics and analysis of data, there are provided a few commands related to analysis of variance, for which there are three assumptions:

- the observations are independent,

- the sample data conform to a gaussian distribution, and

- scores in separate groups have homogeneous variances.

Various other indicators of goodness of fit are available, with varied ease of use or applicability. Among these is Akaike's criterion of information [K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference*, second edition, Springer, New York, USA, 2002], corrected to take into account a number of data sets in a sample; this quantity is defined as

$$AIC = \ln(SSE) + \frac{2p}{n}$$

or in related forms, in which parameters number p and data or observations number n; by means of this criterion that is a compromise between the number of parameters and the fit of a model or objective function, one might select among various models. For a single objective formula this statistic is not highly meaningful or descriptive, but it is useful to rank competitive models, in which case a larger value of $-AIC$ is preferable; small differences of AIC between such models are not necessarily crucial. Unlike most indicators described above, this criterion is applicable to both linear and non-linear regression in various forms.

In summary, the optimal conditions for the results of a fit to a straight line, of form $y = m\,x + b$ with parameters m and b, follow. F has a large value; σ_m and σ_b are small, meaning that ratios $\left|\dfrac{\sigma_m}{m}\right|$ and $\left|\dfrac{\sigma_b}{b}\right|$ have small magnitudes, except with $m \sim 0$ or $b \sim 0$; in the latter case an alternative objective function or model is likely indicated. For a fit of unweighted data, a ratio $\left|\dfrac{\sigma}{\max(y) - \min(y)}\right|$ between standard deviation σ of the fit and the magnitude of a difference between largest and smallest values of y_i is small. Sample correlation coefficient $|r|$ is near unity whereas parameter correlation coefficient c_{mb} is near zero, signifying that variables x and y are strongly correlated whereas parameters m and b are weakly correlated. For $|c_{mb}| < 0.9$ one can generally ignore a mutual dependence of parameters, except for calculation of the effects of propagation of error from fitted parameters to predicted variables, whereas for $|c_{mb}| > 0.97$ one ought to scrutinize both data and fitting model to discover whether an alternative treatment might be preferable. For a fit of weighted data in which each weight w_i is an inverse of a properly evaluated standard deviation of corresponding y_i, a value of standard deviation σ of the fit near unity is preferable; for a weighted fit, a value of σ much greater than unity indicates a possible presence of a gross outlier or discordant values, arising from blunder, or that weights are ineptly assigned, whereas a value of σ much less than unity indicates strongly that weights are assigned too conservatively. If these desirable conditions apply to results of a particular regression, one accepts that a fitting model might be satisfactory and that parameters are well evaluated, with

satisfactory statistical significance. Even when all these criteria hold, one can not distinguish between a merely empirical correlation and a causal relation. If not all these desirable conditions be applicable, either the data or the model requires reassessment. If all error be not associated with only a dependent variable, values of parameters m and b become distorted because of a neglect of the error in the independent variable. For a comparison of multiple possible linear models or objective formulae used for fitting, a model with maximum F statistic has maximal statistical significance. An iterative regression with a re-weighting of data based on results of a preceding iteration might increase the robustness of the method but is susceptible to bias.

 Although much discussion above applies specifically to simple linear regression, involving only a single independent variable, the beauty of a formulation in terms of matrices enables a direct extension to multiple linear regression.

propagation of error

 To calculate the error propagated to a dependent w variable from uncorrelated independent variables x, y, z with known standard deviations or errors $\sigma_x, \sigma_y, \sigma_z$ and the corresponding variances of variables of both types as the squares of those standard errors, we apply the following formulae, in which a, b, c denotes precisely known constant parameters.

- proportionality   -- $w = c\,x$ $\sigma_w = c\,\sigma_x$

- addition or subtraction   -- $w = a\,x + b\,y - c\,z$ $\sigma_w = \sqrt{a^2\,\sigma_x^2 + b^2\,\sigma_y^2 + c^2\,\sigma_z^2}$

- multiplication or division -- $w = \dfrac{x\,y}{z}$ $\dfrac{\sigma_w}{w} = \sqrt{\left(\dfrac{\sigma_x}{x}\right)^2 + \left(\dfrac{\sigma_y}{y}\right)^2 + \left(\dfrac{\sigma_z}{z}\right)^2}$

- exponential   -- $w = a\,x^b$ $\dfrac{\sigma_w}{w} = \dfrac{b\,\sigma_x}{x}$

- exponential   -- $w = a^{(b\,x)}$ $\dfrac{\sigma_w}{w} = b\,\ln(a)\,\sigma_x$

- logarithmic   -- $w = a\,\ln(b\,x)$ $\sigma_w = \dfrac{a\,\sigma_x}{x}$

These formulae are usable in combinations as appropriate.

non-linear regression

 In contrast with linear regression, parameters in non-linear regression might appear in various products and to various powers, such that a derivative of a residual with respect to some particular parameter depends on that or other parameter. Non-linear regression is then fraught with all complications associated with any non-linear problem: both the uniqueness of the solution that must be a concern and the non-linear propagation of error might cause difficulty with any solution. A standard strategy for a solution is to make linear the non-linear equations of the conditions in the vicinity of a solution, and then to solve them iteratively to improve the values of the fitted parameters for an improved fit, until the iterations terminate on achievement of a criterion that defines an acceptable accuracy.   For a regression calculation under these conditions, some initial values of parameters must be provided, even if zero for instance, which might be default initial

values; even if a fit converge, there is no assurance that a global minimum of a sum *SSE* or χ^2 of squared residuals is obtained: a derived solution depends in general on the initial values of parameters that a user provides. Such a fitting procedure is analogous to seeking the deepest valley on a mountainous landscape in multiple dimensions: depending on a particular starting point, the result of a fit might attain not the deepest valley but a subsidiary valley or even just a pass or col between two valleys. The theoretical disadvantages of non-linear regression are that one can not state rigorous expressions for estimates according to a criterion of a least sum of squares of residuals, corresponding to the normal equations for linear regression, and that, in general, estimators as parameters lack exact distributional properties. The former factor leads to iterative methods that might succeed satisfactorily when a functional form for fitting is sensitive to data and when initial estimates of parameters are properly selected; the derived values of estimates of uncertainties of parameters might be based on a linear approximation in the region of a minimum of χ^2. The latter disadvantage implies that indicators of goodness of fit might be unreliable. Despite these disadvantages, a necessity exists to fit the data with parameters of specified uncertainties, even if these uncertainties fail to correspond exactly to the standard errors that would prevail in a linear regression; an imperfect knowledge of these uncertainties is superior and preferable to no knowledge, but one must take care not to attach a full quantitative significance to the particular values of the parameters so derived. For these reasons considerable care in application of non-linear regression is recommended, as is testing of results with disparate initial estimates of parameters.

If a well defined theoretical functional relation be unavailable for a particular data set, care must be taken in the selection of a model or objective formula; for instance, curves of $y = b \sqrt{x}$, $y = a\,(1 - \mathrm{e}^{(b\,(x+c))})$ and $y = a\left(1 - \dfrac{1}{x+b}\right)$ for appropriate values of parameters a, b, c have similar shapes within a limited domain of x, or perhaps merely a polynomial of sufficient terms serves the purpose of finding a model to which data conform. An arbitrary choice of one such formula might produce a subsequent deficient interpretation of the results.

An algorithm associated with <u>Newton</u> and <u>Gauss</u> embodies a method to minimize a sum of squared residuals χ^2 or $\chi^T \chi$, in which $\chi = y - \eta(\beta_i, x_i)$ is typically a vector of residuals; if a response involve multiple dependent variables, a minimum of a determinant from the matrix is applicable instead. According to this approach, one expands an expectation function $\eta(\beta_i, x_i)$ in a multivariate Taylor series, as discussed in section 5.401, and retains terms to only first order. According to a geometric interpretation, beginning at some point on a response surface $y(\beta_i, x_i)$ defined with initial estimates of parameters β_i, this procedure uses first derivatives, which are the coefficients of linear terms in that expansion and which thus become approximate first derivatives of an objective function, to generate improved values of the parameters; these derivatives imply slopes of a surface in various directions, and become elements of a gradient matrix. These values can become further improved in an iterative manner until convergence to a minimum is attained, if initial estimates be not too remote from correct values. According to a particular criterion for convergence, this process is most rapid, or takes fewest iterations, when the direction from the initial point on the surface or hypersurface of χ^2 is that of a steepest descent, but finding that

direction on a hypersurface is difficult.

According to an alternative procedure due to Levenberg and Marquardt, a compromise between a direction of a Newton-Gauss increment and a direction of steepest descent tends to avoid problems due to a nearly singular matrix of derivatives caused by collinearity of its columns. For both procedures, the derivatives of an expectation function $\eta(\beta_i, x_i)$, or residual $y - \eta(\beta_i, x_i)$, with respect to parameters are required, but, for an objective function of a continuous algebraic form in a context of symbolic computation, this condition is generally trivial. Convergence is expedited with an hessian matrix -- second derivatives of an expectation function with respect to the parameters -- particularly when residuals are relatively large. Such an hessian matrix might not be positive definite, particularly when a starting point is not near a solution, or this hessian matrix might be nearly singular. In both procedures *nonlinfit* and *mnlfit* to follow, such an hessian is approximated rather than calculated directly. Although, with linear regression, the standard errors of parameters are well defined through their direct relation to the information matrix, for non-linear regression, only estimates of standard errors of the parameters are practicable; in some cases these values might be unreliable.

linear programming

Linear programming is an optimization of a type in which a model or objective function is a linear function or formula and the constraints are linear equations or inequalities. An optimization differs from a regression in accommodating in its procedure constraints in the fits. In a design of experiments, we seek to optimize the conditions to yield the best or most meaningful outcome, for instance the conditions to investigate the kinetics of a chemical reaction to obtain the most significant values of rate coefficients; one might, for instance, seek to identify the likely dominant source of error in an experiment, and then to minimize that source by experimental design. Such optimization involves either a maximum or minimum of some function, generally subject to certain constraints: if that function be linear in its parameters, this problem is susceptible to attack with linear programming *in standard form*. A slack variable is a non-negative variable added to, or subtracted from, a linear inequality to attain a linear equation. A standard method of solving a problem in linear programming involves the use of a simplex; this term has multiple definitions even in a mathematical context. In geometry, a simplex is the most elementary geometric figure of a given dimension -- a point in zero dimension, a line in one dimension, a triangle -- not necessarily equilateral -- in two dimensions, a tetrahedron in three dimensions, and so forth for hyperspace; a face of a simplex is a simplex of lower dimension. Such a geometric simplex can serve as a basis of design of a simplex lattice in planning an experiment. A simplex can also imply an abstract topological space. A simplex search is a method to maximize, or to minimize, a function of several variables that proceeds through a choice of a direction of descent, or ascent, with an ordered sorting of vertices of an admissible polyhedral set; an idea of a simplex method is to proceed from one feasible solution, which represents an extreme point or vertex of a simplex, of a constrained set of a problem in standard form to another solution in such a way that the value of an objective function continually decreases until a minimum is attained, or increases toward a maximum. A feasible solution constitutes the variables in a set that satisfies all specified constraints. A minimum of a function for parameters in one set implies a maximum in another set, according to the duality theory of linear programming. A simplex method of sequential optimization implies an approach to solve equations, in an over-determined system and involving

non-linear parameters, and is an alternative to non-linear regression, as discussed below. Each such application of this name *simplex* implies a geometric relation or an interpretation of the properties of a geometric figure. For problems in linear programming or optimization involving numerous variables and conditions, matrix or array forms -- hence linear algebra -- provide an elegant and efficient means to specify the conditions and to test the results.

   Further information, general and specific,on linear programming and a simplex algorithm is available in books such as by

   V. Chvatal, *Linear Programming* (Freeman, New York, USA, 1983), or by

   R. J. Vanderbei, *Linear Programming: Foundations and Extensions* (Kluwer, Boston, USA, 1996), or by

   S. Venit and W. Bishop, *Elementary Linear Algebra* (Prindle, Weber and Schmidt, Boston, USA, 1982), especially chapter 9.

Another book by

   R. Fletcher*, Practical methods of Optimization* (Wiley, Chichester, UK, second edition, 1987)

contains helpful information related to section groups 8.3 - 8.6. To solve problems in linear programming one can alternatively apply methods based on an ellipsoid or an interior point, also discussed in these books.

optimization

   An optimization implies a determination of an optimal value, typically minimal or maximal, of a model or objective function subject to constraints. A problem in an optimization has a general form,

$$\min(f(x)) \text{ or } \max(f(x)) \text{ for which } x \text{ is in } \mathbf{R}^n$$

$$\text{subject to } c_i(x) \leq 0 \text{ with } i \text{ in } \mathbf{Z} \text{ or } c_i(x) = 0 \text{ with } i \text{ in } \mathbf{Z}.$$

The former line here implies that one seeks to find a minimum or maximum value of an objective formula or function, with its variables numbering n, hence associated with a space of variables having n dimensions. The latter line implies that this minimum or maximum must be satisfied subject to constraints having the forms of inequalities or equalities. Procedures *wmlinfit* for linear regression and *mnlfit* for non-linear regression, to be presented in succeeding sections for those purposes, enable an optimization with no constraints. In sections 5.403 and 5.404 we present methods of constrained optimization involving linear or non-linear constraints and Lagrange multipliers, but these methods are applicable to perfect data and hence yield no estimate of uncertainties. Linear programming allows an evaluation of linear parameters with linear constraints. *Maple* includes a package for constrained optimization that extends a scope of problem that might be solved to include quadratic parameters with linear constraints and a general case of non-linear parameters and non-linear constraints; a worksheet contains examples of use of procedures provided for these purposes, but commands therein provide no direct measure of goodness of fit of individual parameters. The approaches to optimization thus include unconstrained linear and non-linear regression, in which a typical criterion is a least sum of squared residuals. An optimization involving constraints based on unrigorous inequalities involves linear and non-linear programming. *Maple* contains a package for optimization,

```
> ?Optimization
```

including interactive approaches, but the procedures therein provide also no explicit measure of the goodness of fit or the uncertainties of the parameters.

As an alternative to methods of linear and non-linear regression explained in section groups 8.3 and 8.4, a simplex method is applicable to an estimation of non-linear parameters. A simplex is here a geometrical object with vertices and faces numbering one greater than the number of parameters to be evaluated, hence a triangle for a case of two parameters or a tetrahedron for a case of three parameters. Inclusion of weights into a formula or function χ^2 poses no problem; a derivation of the estimated standard errors is practicable through a matrix formulation [G. R. Phillips and E. M. Eyring, Analytical Chemistry, **60**, 738--741 and 2656, 1988]. A simplex procedure for non-linear optimization requires no derivatives; with a careful control of the contraction of a simplex toward convergence and re-expansion to double the value of χ^2 at convergence, reliable estimates of the parameters might be obtained.

Maple 10 and beyond contains a package Statistics that is a collection of functions and interactive tools for mathematical statistics and data analysis; this package supports diverse common statistical tasks such as quantitative and graphical data analysis, symbolic manipulation of random variables, simulation and curve fitting, but procedures provided for the fitting of data yield no estimates of goodness of fitted parameters. Much functionality in this package is accessible through context menus.

In the preceding chapters we treat mostly continuous variables and functions, which we can generally differentiate and integrate one way or another with normally no concern for an uncertain value apart from numerical rounding when working with numbers of type float; we typically diminish such inaccuracy to a negligible extent by setting appropriately a number of decimal digits to be carried through a calculation. Mathematics possesses other aspects according to which one treats both discrete variables and continuous distributions in which a primary interest lies in the nature of the uncertainty of a value. In this chapter we investigate how an uncertainty can arise and how to cope with such a condition to obtain a result that we might, according to an appropriate criterion, consider optimal. We consider these statistical aspects beginning with probability, discovering how combination and permutation of objects produce varied prospective outcomes of an event. After discussion of the nature of a statistical distribution, which implies a set of possible values of a random variable or points in a sample space considered in terms of their theoretical or observed frequency, we examine how to evaluate significantly a few parameters that can represent compactly numerous experimental measurements, with models in which parameters appear either linearly or non-linearly. In a design of experiments to yield cogent information in general, or optimal values of parameters in particular, one might encounter issues of efficiency or of economy of the conduct of those experiments, which might imply some compromise between competing objectives; as a prelude to statistical analysis of experiment results, to plan the experimental conditions such as the concentrations of reactants, temperature et cetera is advisable on the grounds of efficiency, involving optimization and linear or non-linear programming so as ultimately to produce values of the parameters with maximal significance. Other methods to treat data susceptible to uncertainty to derive the value of an optimum descriptor or parameter might mimic natural processes, such as *neural networks* and *simulated annealing*. Analysis through a selection of principal components and through partial least squares employs predictor variables in

linear combination, rather than the original variables. All these methods have important applications in the conduct and analysis of chemical experiments.

reference *The Advanced Theory of Statistics*, M. G. Kendall and A. Stuart, Griffin, London UK, 1973

\>

summary of chapter 8

   A quantity specified without an estimate of its reliability, or its uncertainty, is worthless.

   A principal objective of this chapter is to facilitate a solution of problems involving reproducing numerical data through models and their parameters, so that each evaluated numerical quantity possesses an associated indicator of its quality. A basis of the presented methods is generally statistical in nature, involving applications of probability that we introduce in section 8.101, with its relation to combination and permutation. Unlike purely mathematical methods, for which a number might be meaningful in isolation, in a statistical context, a number -- whether for a quantity or for its uncertainty -- is never meaningful in isolation; in considering such numbers one must invariably bear in mind that other measurements yield other values of data, from which accordingly result other values of parameters and their estimates. To make such estimates, one has invariably to work, implicitly or explicitly, within some framework of a distribution of errors; although, particularly in chemical and physical sciences involving numerous molecular entities within a particular sample, one tends to assume a gaussian distribution and correspondingly a criterion of least squares of residuals -- weighted if practicable, each such assumption should be considered and tested for each method and sequence of measurements.

   For data involving a single variable in section 8.201, only two parameters -- mean and variance -- provide a summary of those data according to a distribution of specified kind. For data involving single or multiple independent variables that appear in a linear or pseudo-linear form in an objective formula in section 8.307, or involving single or multiple independent variables that appear in an arbitrary algebraic formula or function -- even defined in an extensive external procedure, in section 8.410, statistically based methods yield useful estimates of numerous parameters: these parameters are over-determined, because data outnumber parameters, but they can optimally reproduce the fitted data according to the errors in a well defined distribution. A formula or procedure that serves as a model or objective function in these cases might have either a theoretical basis or a purely empirical or intuitive origin. An analysis of data is a process during which we learn what effects the data and the degree of complexity of a usable approximating model support. A theory of information provides applicable approaches because an aim of this analysis is to separate information from accompanying disturbance called noise: information involves a structure of relations, estimates of parameters of an objective function or model, and components of variance, whereas noise or disturbance pertains to residuals or variation left unexplained. In seeking a model or objective function that describes data satisfactorily, we try optimally to separate noise, which constitutes non-information or entropy, from structural information, or negentropy, so as to minimize a loss of information from data through their reduction to parameters according to indicators of goodness of fit.

   We discuss in section 8.501 further optimization involving evaluation of quantities to yield extreme properties -- a maximum or minimum -- of an objective function that is linear in form. Because procedures *wmlinfit* for linear regression, present in section 8.306, and *mnlfit* for

non-linear regression, present in section 8.410, provide indicators of goodness of fit of parameters and model, we recommend that these procedures be a primary method of attack whenever applicable to problems arising from modeling of chemical data. A capability of procedure *nonlinfit*, in section 8.402, to fit simultaneously parameters to two objective formulae makes this procedure applicable also to optimization in favourable circumstances; the methods of linear or non-linear programming present in section groups 8.5 or 8.6, respectively, might otherwise serve to achieve optimization in chemical problems. Although methods of linear and non-linear regression, and their implementation in procedures *wmlinfit* and *mnlfit*, constitute powerful approaches to optimization in a sense of fitting and representing data, there is no provision for constraints, such as that a particular parameter have a positive sign or a value within a specified range; if error propagated into parameters from imperfect data yield such undesired or physically unrealistic results, either improved data or selection and improved treatment of weights might relieve these conditions. In contrast, the methods of linear and non-linear programming include provision for constraints of various kinds; one must be aware that, if one makes recourse to these approaches to compensate for imperfect data, the values of retained parameters might exhibit a bias and are unlikely to be as significant as one might hope. Despite the power of all present algorithms and their implementations, data of an exceptional nature might require other approaches and their respective algorithms; an implementation involving symbolic computation, with graphical display, remains however a formidable approach to an analysis and a solution of chemical problems involving fitting of numerical data of any kind.

In an application of statistical tests as a basis of a decision of significance, or hypothesis testing , one must be aware of the following aspects.

- For a particular level of confidence, most statistical tests provide only clues whether one or more elements or samples differs from others.

- Statistical tests incorporate mathematical models against which reality might be compared, for instance Student's *t* test and analysis of variance for univariate data, or linear or non-linear regression for bivariate or multivariate data.

- As functional relations, appropriate mathematical models must become specifically incorporated into a test, or data must be transformed to become testable according to standard procedures.

- A decision based on purely statistical evidence is inferior to one supported by chemical or physical experience; a statistical correlation implies no causal relation.

- A comparison of samples or models is appropriate only if no disparity is evident or plausible.

- Incomplete models and uncertain parameters derived from statistical tests provide no basis for a significant decision.

Analogously to other branches of applied statistics such as psychometrics, biometry and econometrics, chemometrics is concerned with an application of mathematical and statistical methods to aid an acquisition and interpretation of chemical data. Aspects of chemometrics include pattern recognition in chemistry, a development of clustering methods in clinical analytical chemistry, a use of simplex optimization to improve instrumental performance and a

development of rapid filters to process spectra; linear or non-linear regression is conducive to at least some attack on pattern recognition, and filters of a type to which allusion is just made are susceptible to methods of regression, to spline fitting as discussed in section 2.411 et cetera. Most applications associated explicitly with chemometrics are found in analytical chemistry, and relate to a multivariate nature of chemical data, measurable experimental error and signal processing; many methods and procedures discussed in this present chapter have thus direct application for these purposes. Like Moliere's character M. Jourdain who discovered that "I have been speaking prose all my life, and didn't even know it", a chemist who, after assimilation of this chapter, consults a book on chemometrics can discover that he or she has already encountered here many pertinent concepts and methods, which play a major role in a contemporary practice of analytical chemistry. A design of experiments for a particular application in analytical chemistry must involve consideration of a statistical significance of the eventual data and their interpretation, and mathematical tools for this purpose can usefully include both symbolic and numeric computation.

>