How to Search

MATHEMATICAL HANDBOOK

Version 1

Sotirios Persidis

Professor Emeritus

University of Thessaloniki

Practical problems in writing and reading mathematics with a computer

Computers and mathematics

The relationship between computers and mathematics has been always close and constantly evolving. When computers first appeared, their machine characteristics made them attractive for repetitive, precise and fast work, provided some clear rules were given to their hardware. Starting about sixty years ago, computers have been used for numerical calculations, typesetting mathematics and algebraic calculations. Today, computers do mathematics much better, but their performance, especially in relation to writing and reading mathematics, is far from perfect and needs further improvement.

The typesetting of mathematics remains a problem that has been partially and inadequately solved. The important step made with the introduction of TEX more than 40 years ago remains essentially at the level of ASCII, while today our Unicode fonts have many thousands of characters. The keyboards are not suitable for typing mathematical symbols. Most important, the software for typesetting is old in its conceptual base and does not exploit the power of personal computers. In conclusion, with a computer today we can typeset anything, but we have to put a lot of effort for something that should be done more easily. Reading mathematics by computers is also very problematic, since OCR systems cannot read mathematical documents. Furthermore, mathematical content in files cannot be searched inside mathematical formulas. This last and small problem of searching inside mathematical formulas is solved in this work.

Solutions to all these problems are continuously sought, but progress is slow. The mathematical community seems to be content with the present situation and little effort is made to improve it. Software companies are not actively interested in building better software for faster and more effective typesetting of mathematics. Hardware companies reproduce the well known keyboard with no essential improvements. In general, people are trying to do their job with whatever is available and that is all.

The changes that are needed are the following:

1) **A better keyboard has to be constructed.** We must be able to type from the keyboard about 1000 characters to cover the Latin and Greek alphabets (both lower case and upper case), all punctuation marks, all mathematical symbols, plus a few new symbols for special use (e.g. to represent the bar of a fraction). We must be able to type each character by pressing one or two keys (at the same time or sequentially), without memorizing anything. It is not difficult to construct such a keyboard with some minimal changes of the existing keyboard.

2) **The typesetting software must format mathematics automatically.** This means that a mathematical formula must take its proper form as soon as we type the symbols sequentially from the keyboard. TEX has such a mechanism, but it remains adapted to standards of some 40 years ago (fonts, notation, procedures, etc.). Today, the typesetting software that is used for books and magazines is much more powerful and able to combine many forms of information (text, drawings, pictures and mathematics in graphics, etc.), but it does not have an intrinsic mechanism to form mathematics automatically and appropriately. Furthermore, drawings, pictures and even constructions evolving in time must be an integral part of mechanism presenting mathematics. The ability that a computer gives us to sketch the evolution of phenomena in time is a new and powerful tool for transmitting knowledge.

3) **Computers must learn to read mathematics from a file or a printed page.** This will help to use all the extensive material that has been accumulated in the libraries for hundreds of years by mathematicians and other scientists. At present, the degree of character recognition for complicated printed material is not satisfactory and for mathematics is unacceptable.

The previous problems for writing mathematics are quite difficult and cannot be solved all together. The best software available today has some surprising limitations, that it should not have. For example, Adobe Acrobat is an excellent software for presenting "closed" files (i.e. files that cannot be altered). When, however, it comes to searching for "a_n", we also find "a^n" and "an", since inside the Find or Search box we cannot type exponents, indices, etc. This deficiency seems to be the easier to overcome but requires some change of mentality from the people who build software. Since most documents contain plain text, we search mostly for words. But that is not how it should be. We must be able to search for graphics of any kind and for elements that are parts of a graphic. The present work makes a step towards this direction by replacing the graphics that contain mathematics with plain text composed of letters and symbols. Simplicity of the method is another main objective.

A simple language for writing mathematics

The essential problems for typesetting or reading mathematics are the following:

1) A mathematical expression can be written in many ways, all correct and with the same meaning to a mathematician, but different in appearance. For example, we can write

$$\sqrt{a} \ \text{ or } \ a^{1/2}, \qquad ax^2 + bx \ \text{ or } \ (ax + b)x, \qquad \frac{\sin x}{\sin x + \cos x} \ \text{ or } \ \frac{\tan x}{\tan x + 1}$$

Since we are forced to choose one in each case, a subjective criterion in writing mathematics is introduced. We prefer the "simpler expression", but what is "simpler" depends on our personal preferences.

2) The characters needed to write mathematics are many more than those required to write plain text. In addition to the Latin and Greek alphabets, we need many symbols, such as $\neq, \leq, \sqrt{\ }, \nabla, \partial, \int, \infty$ and many others. Since we cannot type them from the keyboard, we are forced to introduce them with the mouse or to type whole words (as in TEX) and such procedures are very time consuming.

3) For mathematics presented in pdf files, we can type directly into the Find or Search box **only the characters that appear on the keyboard and only in one line**. We cannot type directly Greek letters and many mathematical symbols ($\neq$, $\leq, \infty, \int, \sqrt{\ }$, etc.) and we cannot type subscripts or superscripts. Consequently, we need a method that will enable us to **write all symbols and all mathematical expressions in one line**.

The first step towards that goal should be **the definition of a new "alphabet"** that will have many additional characters to those available in ASCII, in order to write mathematical expressions. Then we need some **rules for writing all mathematical expressions in a meaningful way and in one line**. At the same time we have to **keep the whole system as simple as possible**.

To address these problems we have made the following choices:

1) The "basic characters" of the new alphabet are the characters that appear on the keyboard only. These include the numbers 0 to 9, the lower case letters a to z, the upper case letters A to Z, and the various punctuation marks that can be typed by pressing a single key or Shift plus a key.

2) Greek letters can easily be typed in Windows, if we use an international keyboard that allows to switch between Latin and Greek. If we use Greek letters seldom, the Character Map offers another method. Finally, a third way to introduce Greek letters into the Find or Search box is by Copy and Paste.

3) Other characters or symbols that are needed are represented as follows:

  $\neq$ by \=, $\leq$ by \<, $\geq$ by \>, ∞ by \o, $\pm$ by \+, ∂ by \d, ∇ by \D, $\in$ by \e,
  $\simeq$ by \~, $\mp$ by \-

4) The operations of addition, subtraction, multiplication and division are denoted explicitly by +, -, *, /. Exponentiation and indexing are denoted by ^ and _.

5) Some usual operations in mathematics are represented as follows:

  the root $\sqrt[n]{a^m}$ by a^(m/n)

  the scalar product of two vectors **A·B** by A*B

  the vector product of two vectors **A×B** by A\xB

  the complex conjugate of A by #(A)

  the limit $\lim_{x \to a} f(x)$ by lim[x,a]{f(x)}

  the derivative of a function $f(x)$ $\frac{d}{dx} f(x)$ by (d/dx){f(x)} or d(f(x))/dx or f'(x)

  the derivatives of higher order by (d/dx)^2{f(x)}, (d/dx)^3{f(x)}, etc.

  the partial derivatives $\frac{\partial \Phi}{\partial x}$, $\frac{\partial^2 \Phi}{\partial x \partial y}$, etc. by (\d/\dx){Φ}, (\d/\dx){(\d/\dy){Φ}}

  integration $\int f(x)dx$ by \I[x]{f(x)} and $\int_a^b f(x)dx$ by \I[x,a,b]{f(x)}

   any integral $\int f(x,y,...)dx\,dy...$ by \I{f(x,y,...)*dx*dy...}

   a line integral along a curve C $\int_C \Phi ds$ by \I_C{Φ*ds}

   a surface integral over a surface S $\int_S \Phi dS$ by \I_S{Φ*dS}

   a volume integral over a volume V $\int_V \Phi dV$ by \I_V{Φ*dV}

  expansion of $f(x)$ into a sum or series by \S{f(x)}

   a sum for some values of x $\sum_x a(x)$ by \S[x]{a(x)}

   a sum $\sum_{k=1}^{n} a(k)$ by \S[k,1,n]{a(k)}

  expansion of $f(x)$ into a product by \P{f(x)}

   a product $\prod_{k=1}^{n} a(k)$ by \P[k,1,n]{a(k)}

  expansion of $f(x)$ into a Fourier series by \F{f(x)}

  the Fourier transform (of any kind) of $f(t)$ by \F{f(t)}

  the Laplace transform of $f(t)$ by \L{f(t)}

6) Some letters and combinations of letters represent special quantities, functions, operations, etc. as follows:

e, π, i for the base of natural logarithms, the usual pi, and the imaginary unit

i, j, k for the unit vectors along the axes of a Cartesian coordinate system (in practice there is no danger of confusing the two i's)

sinx, cosx, tanx, cotx for the respective trigonometric functions of x

sin^(-1)x, cos^(-1), tan^(-1)x, cot^(-1)x for the inverse trigonometric functions of x

sinhx, coshx, tanhx, cothx for the respective hyperbolic functions of x

sinh^(-1)x, cosh^(-1)x, tanh^(-1)x, coth^(-1)x for the inverse hyperbolic functions of x

erf(x), erfc(x), Γ(x), ζ(x), sgn(x) for the error function, the complimentary error function, the gamma function, the Riemann zeta function and the sign function

mat(a,b;c,d) and det(a,b;c,d) for the matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and the determinant $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ and similarly for higher orders

When we write mathematics using this notation we accept the following rules:

1) We do not type spaces anywhere, unless they have a special meaning.

2) A single letter or a group of letters (without spaces) represents a single variable.

3) The various operations are executed in the following order: **indexing, function, exponentiation, multiplication and division, addition and subtraction**.

4) We use parentheses only when they are necessary to show and/or impose a different order of executing the operations.

5) We do not use brackets or braces of any kind (except in assumed notation).

6) We use a dot (.) in decimal numbers, not a comma.

7) We use a colon (:) to express "for" or "if".

8) We use a comma (,) to indicate separation.

9) We use a semicolon (;) to indicate a change of row, e.g. (m;n) for $\begin{pmatrix} m \\ n \end{pmatrix}$

10) We avoid starting an exponent with a negative sign, except for an exponential function $e^{-f(x)}$ which is written as e^(-f(x)).

11) We use three dots (...) to indicate many terms or details that are not needed.

Searching in mathematics

As mentioned before, a mathematical document is prepared using some kind of computer software (e.g. Word, TEX, or a more sophisticated modern software for composition and page layout). To be easily readable by man, the final form (e.g. in pdf) has mathematical expressions that appear as either plain text or a graphic. Plain text can be easily read and searched by a computer, but mathematics in graphics cannot be searched. Searching even plain text presents problems. If, e.g., we want to search for e^x, the searching mechanism in Adobe Reader will give us also e_x and ex.

To be more precise we will use the terms (a) *simple mathematical expression* for **any sequence of characters written according to the previous rules in one line** that gives a meaningful mathematical expression, and (b) *full mathematical expression* for any other mathematical expression. For example

x-a,   a*x^2+b*x+c,   e^(-z^2),   (x_n+a)^(1/2),   (sinx^2+cos(a*x))/(a*x+b)

are simple mathematical expressions representing the full mathematical expressions

$$x - a, \quad ax^2 + bx + c, \quad e^{-z^2}, \quad \sqrt{x_n + a}, \quad \frac{\sin^2 x + \cos ax}{ax + b}$$

the first two of which are written in plain text, while the other three are graphics.

We come now to the central argument that justifies the introduction of the previous structure: **Using Adobe Reader we can easily search for any simple mathematical expression or for a part of it. If we replace the full mathematical expressions by simple mathematical expressions, then we can search for anything. Hence we need a copy of the Mathematical Handbook in pdf format with simple mathematical expressions only.**

An additional remark can make the preparation of this copy of the Mathematical Handbook substantially easier. When we search for something (in text or in mathematics), we search for something simple and meaningful that comes to our mind. We do not search for something trivial (e.g., for a single and common letter) or a complicated phrase or expression, since in most cases we do not know this phrase or expression in advance. Hence, **it is enough to replace each full mathematical expression by its important parts that can be the targets of a reasonable search.**

This objective has been completed for the Mathematical Handbook and the result is the file SearchMH.pdf. The user has now **the first searchable Mathematical Handbook**. He can search through it using the Find or Search box of the Adobe Reader. This is an easy searching procedure with free, easily available and reliable software. We hope the reader will appreciate it, use it and see in it the great possibilities it opens.

Examples of full and simple mathematical expressions

Some examples of full mathematical expressions are given below as well as the corresponding simple mathematical expressions. We have to remember that when we search we do not know the full mathematical expression. Hence we will enter a simple mathematical expression or a part of it into the Find or Search box and this should lead us to the full mathematical expression. In simple mathematical expressions the order of operations according to the rules is important.

Full mathematical expressions	Simple mathematical expressions
$a + b\sin^2 x_n$	a+b*sinx_n^2 (observe the order of operations)
$\dfrac{a+b}{cx^2 + d}$	(a+b)/(c*x^2+d)
$\displaystyle\int \dfrac{dx}{\tan ax}$	\I[x]{1/tan(a*x)}
$\displaystyle\int_0^1 \dfrac{dx}{\sqrt{\lvert \ln x \rvert}}$	\I[x,0,1]{1/|lnx|^(1/2)}
$f(x) = \begin{cases} -1 & -\pi < x < 0 \\ 1 & 0 < x < \pi \end{cases}$	f(x)=-1:-π<x<0;1:0<x<π
$\sqrt{\dfrac{2}{\pi}}\,\dfrac{\sin a\omega}{\omega}$	(2/π)^(1/2)*sin(a*ω)/ω
$\exp\left[-\dfrac{(x-\mu)^2}{2\sigma^2}\right]$	e^(-(x-μ)^2/(2*σ^2))
$\dbinom{n}{x} p^x q^{n-x}$	(n;x)*p^x*q^(n-x)
$f(z) = \dfrac{1}{\sqrt{2\pi}}\, e^{-z^2/2}$	f(z)=(1/(2*π)^(1/2))*e^(-z^2/2)
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$	mat(a_(11),a_(12);a_(21),a_(22))
$\displaystyle\int_0^1 \dfrac{x^n}{1+x}\,dx = (-1)^n\left[\ln 2 + \sum_{k=1}^n \dfrac{(-1)^k}{k}\right]$	\I[x,0,1]{x^n/(1+x)}= (-1)^n*(ln2+\S[k,1,n]{(-1)^k/k})

Examples of possible searches

Certainly, we can search for any word or scientific term. For mathematical expressions some searches that can be tried are the following:

A search for	finds	A search for	finds
s-a	$s - a$	\D\xA	$\nabla \times A$
x^(2*n)-y^(2*n)	$x^{2n} - y^{2n}$	x^2/a^2+y^2/b^2	$\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2}$
(x+y)^n	$(x + y)^n$	(d/dx){sin^(-1)x}	the derivative of $\sin^{-1} x$
tanθ	$\tan \theta$	y" or (d/dx)^2{y}	$\dfrac{d^2 y}{dx^2}$
π/n	$\dfrac{\pi}{n}$ (regular polygons)	\P{coshx}	Expansion of $\cosh x$ into a product
(a*x+b)^(1/2)	$\sqrt{ax + b}$	ρ*cosφ	$\rho \cos \varphi$ (polar coord.)
(a*x+b)/(c*x+d)	$\dfrac{ax + b}{cx + d}$	\S[k,1,n]{sin(k*x)}	$\displaystyle\sum_{k=1}^{n} \sin kx$
R	$R, \; ax^2 + bx + c, \; \sqrt{\dfrac{2}{\pi x}}$	\F{t/(t^2+a^2)}	Fourier transform of $\dfrac{t}{t^2 + a^2}$
y_k	y_k (formulas with y_k)	B_	Bernoulli numbers
\I[x,0,π/2]{x/tanx}	$\displaystyle\int_{0}^{\pi/2} \dfrac{x \, dx}{\tan x}$	\I[x]{(a*x+b)^(1/2)/x^2}	$\displaystyle\int \dfrac{\sqrt{ax + b}}{x^2} dx$
tan(a*x)	$\tan ax$	Γ	gamma function
\I[x,0,1]	integrals from 0 to 1	\L{lnt}	Laplace transform of $\ln t$
\S{cothx}	expansions of $\coth x$	(n;x)	combinations $\dbinom{n}{x}$
J_n(x)	expressions with $J_n(x)$	T_n(x)	expressions with $T_n(x)$
\S{J_n(x)}	expansions of $J_n(x)$	erf	error function
\S{sin(n*x)}	expansions of $\sin nx$	\S{(1+x)^(3/2)}	expansion of $(1 + x)^{3/2}$
\I[x]{K/sinx}	integral of $K/\sin x$ ($K = \sqrt{1 - k^2 \sin^2 x}$)	tanh^(-1)((x^2-1)^(1/2)/x)	$\tanh^{-1} \dfrac{\sqrt{x^2 - 1}}{x}$

Remaining problems

Advanced and intelligent searching is a very complex task and the software available today is not adequate. In simple word processors (e.g. Word) we can easily write plain text with subscripts and superscripts, e.g. a_0x^2. However, we cannot enter the same information in the Find or Search box. This is also true for pdf files. Hence, the first step for improvement should be the upgrading of the searching abilities of the software.

The problem becomes more difficult if we want to search inside graphics. In mathematical books or files we have three main types of graphics: mathematics given in graphics (e.g. eps files), figures (usually vector graphics) and pictures (usually bitmaps). In this Mathematical Handbook we have established a method to search inside mathematics and other graphics after replacing them by plain text using the language introduced. Although the step is significant, we are far from solving the whole problem.

The method presented here for searching inside mathematics should be regarded as a first proposal. The whole procedure is simple and works. Experience from its use and innovation will certainly improve the situation. However, the main obstacles remain: we need a better keyboard (for typing all the mathematical symbols easily) and better typesetting software (for formatting mathematics automatically). The computer industry has solved much more difficult problems. Thus, the solution of these problems is possible. We need only determination, some capital and the correct basic ideas for building a solution.

As a final remark let us compare briefly the language introduced here with TEX. It is clear that TEX is a much richer language designed to cover symbolic representation of all cases in mathematics. At the same time, because of this, TEX is very complex and even people who know TEX well find some difficulty to read all ASCII text produced using TEX. At the other hand, the language introduced here has a limited objective: to cover the Mathematical Handbook. Using the language in searching will help us to see how to expand it to cover more areas. We prefer something simple and easily understandable by everyone that covers most mathematics we use in practice.