Mathematics comes alive

$\frac{1}{2\pi i} \oint_C f(z)\,dz = \sum_{k=1}^{n} \operatorname{Res}_{z=a_k} f(z)$

OpenMath **ESPRIT** PROJECT 24969

Version: **Draft 0**
Date: **17 April 1999**

OpenMath LaTeX to OpenMath Converter ml2om

Richard M. Timoney

D2.1?

Abstract

The aim of this document is to explain the LaTeX commands currently understood by the ml2om converter, together with the mathematical meaning such commands are assumed to have. In rough outline, the aim is to understand all common notations used in a non-special context, without imposing arbitrary assumptions on the style of LaTeX used. By its nature, this aim can never be 100% satisfied.
A subsequent document will explain how to tailor ml2om to achieve translation of more advanced material, where symbols may have specific meaning according to the field.

Contents

1 Algebraic expressions

At the lowest level ml2om will assume that single letters like a, b, x, etc. represent unknown (numerical) quantities. The symbols $+$, $-$, $/$ and $\div$ are assumed to reprsent the usual algebraic operations of addition, subtraction and division, while juxtaposition (as in xy) is taken to mean multiplication. Also fractions are assumed to represent division, both in the LaTeX form `\frac{ac}{b+1}` and the Plain TeX form `{ac \over b + 1}`. Both of these typeset as $\frac{ac}{b+1}$ and are taken to mean the product a times c divided by the sum of b and the integer 1.

Numerical constants are recognised by an ad-hoc rule that a sequence of digits `0-9` (with no spaces) is assumed to represent an integer quantity. Other ad-hoc rules are that a sequence like `12.54` is assumed to represent a decimal number and `0x12ae` is assumed to represent a hexadecimal integer. This represents a first breach of the rules of LaTeX.

A superscript is taken to represent a power, so that x^{a+1} is taken to mean x raised to the power of the sum of a and 1. The rules about integers and floating point numbers have the drawback that `x^21.34` typesets as $x^21.34$ but is treated by ml2om as though it was $x^{21.34}$.

Actually all LaTeX commands and all letters (except the digits and some symbols which have a special significance to LaTeX such as `$%^_`) are interpreted by means of a configuration file read at run time, and their interpretations are largely a reflection of the default configuration file. Many LaTeX commands such as `\alpha` and `\beta` which typset as single 'letters' are also treated as unknown variables, but there are some exceptions.

π (i.e. the input `\pi`) is assumed to mean the mathematical constant which is approximately 3.14. The letter f is taken as the name of a function (see section 2) and the letters d and ∂ are assumed to relate to differentials if they occur within an integrand or within a `\frac{ }{ }` (see sections 6 and 7). The symbols `=<>` are treated as binary relations, as are several other symbols such as $\leq, \geq, \neq, \in, \notin, \subset$.

Grouping is done on the basis of futher ad-hoc rules, where an attempt is made to match various types of left parenthesis with some corresponding right even though there is no rule in LaTeX to enforce such matching. To compensate, missing right brackets are assumed at places where they are deemed to be overdue. The LaTeX rule that enforces matching braces `{ }` in the input is enforced at a higher precedence than the ad-hoc rule on visible brackets.

At the moment, delimeters such as $|$ and $\|$ which serve a dual purpose as left and right delimeters cause errors unless they are preceded by an explicit `\left` or `\right`.

2 Functions

There are a small number of functions which have standard command sequences in LaTeX that clearly relate to the way these functions are commonly represented. Examples are the trigonometric functions such as $\cos$, $\sin$ (input as `\cos` and `\sin`), the inverse trigonometric functions such as $\arctan$, the hyperbolic and functions such as $\sinh$ and a handful of others such as $\ln$, $\exp$, $\log$, $\gcd$,

$\det$.

These are recognised and an attempt is made to find the argument. If no argument is found, the command is taken to be a reference to the function in the abstract as opposed to an instance of the application of the function to some argument.

Subscripted variables are treated as synomyms for functions. For example an infinite sequence $x_1, x_2, \ldots$ may be viewed as a function with domain the natural numbers and value x_n at n, and this is the justification for treating y_α as a function named by the variable y applied to a value α in some domain.

As mentioned earlier, the letter f is also declared as a function, though this is partly an experimental interpretation. There is an interim way to declare that f should just be a variable called f, by changing it to `\omvar{f}` at each occurrence.

Here are some examples

Input	**Typesets as**	**Interpretation by ml2om**
`\cos x + 1`	$\cos x + 1$	Sum of $\cos$ applied to a variable named x and 1
`\cos { x{y+z}}`	$\cos xy + z$	$\cos$ applied to a product of a variable named x times the sum of variabls y and z
`\tan ( x + y)`	$\tan(x + y)$	Tan of the sum of x and y
`f x^2`	fx^2	Apply function named by variable f to result of taking variable x to the power 2
`{\ln x}^2`	$\ln x^2$	Apply function $\ln$ (natural log) to a variable named x and raise the result to the power 2
`\sinh x^2 y`	$\sinh x^2 y$	Multiply $\sinh$ applied to square of x by y
`\cos^2 y`	$\cos^2 y$	Apply power 2 to result of applying the function $\cos$ to a variable named y

3 Relations

Binary relations are recognised as such, but only one relation is allowed in a chain at present. So $x^2 + 4x \leq 5y^3 + z$ will be digested in a reasonably senseible way (apply the relation $\leq$ to the two expressions $x^2 + 4x$ and $5y^3 + z$). However $0 \leq x < 1$ is not currently recognised and neither is $x \in S \subset T$.

Relations in lower limts for integrals and summations are interpreted separately.

4 Set theory

Set-theoretic operations of $\cup$ and $\cap$ are recognised as infix operators that act on the two adjacent symbols. Also element relations such as $\in$ and $\ni$ are recognised.

The input `S \cap T \cup U \owns x` typesets as the rather confusing $S \cap T \cup U \ni x$ and is interpreted by ml2om as meaning that a variable called x is an element of the set $(S \cap T) \cup U$. Thus $S \cap T$ is the set intersection operation applied to variables named S and T and $S \cap T \cup U$ is interpreted by left association as the application of the set union operation to that intersection and a variable names U.

The more clear input `S \cap (T \cup U)` typesets as $S \cap (T \cup U)$ and is interpreted correctly: the set intersection operation applied to the variable S and the result of the set union operation applied to variables T and U.

'Big' versions of unions and intersections such as $\bigcup_{i=1}^{n} S_i$ or $\bigcap_{j=1}^{m} T_j$ are interpreted with the same limitations as for summations or products. (See section 5.)

5 Sigma summations and Pi products

$\sum_{i=1}^{n} i^2+1 = s$ (input as `\sum_{i=1}^n i^2 + 1 = s` or `\sum^{n}_{i=1} i^2 + 1 = s`) is dealt with as follows. The fact that the summation variable is i is detetected as the leftmost part of the realtion $i = 1$ in the subscript (a rather ad-hoc rule). The trailing expression is then treated as a lambda expression with variable i, so that a mapping $i \mapsto i^2 + 1$ is the result. According to the current experimental CDs, this function is then used to generate a list, the list of values of that function on the integer interval from the lower limit 1 to the upper limit n. The operation $+$ is then applied to the entries of that list. Finally the relation $=$ is applied to the result and a variable named s.

Shorthand notations $\sum_{1}^{n} i^2 + 1$ for the same summation generate partially correct output with a missing lambda expression replaced just by the formula $i^2 + 1$ together with an XML comment.

The same technique is applied to $\prod_{j=1}^{k} i^3 + i$ and to $\bigcup_{k=2}^{5} S_k$, except that the operation of addition is replaced by times or set union.

6 Calculus — integrals

Definite integrals are assumed to have the forms $\int_{x=a}^{b} f(x)\,dx$ or $\int_{t=a}^{b} f(t)$ or $\int_{a}^{b} f(u)\,du$ with the variable of integration preferably indicated with a lower limit $x = a$ or alternatively by a trailing dx. x can be relaced by any variable letter and the integral ends with a differential (eg dx) or a relation or anything that is not a valid part of an expression. The lower and upper limits a and b can be expressions.

Indefinite integrals are assumed to have a trailing dx (or dt or x replaced by any variable).

If the integrand as considered by the rules has no dx and there is no $x =$ in the lower limit of the integral, then the resulting may well be incomplete. That is the integrand may be a simple expression where it should be a lambda expression.

Here are some examples

Input	**Typesets as**	**Interpretation by ml2om**
`\int_0^1 x\, dx`	$\int_0^1 x\, dx$	Integral from 0 to 1 of the lambda expression $x \mapsto x$.
`\int_0^1 x dx`	$\int_0^1 x dx$	Same output.
`\int_0^1 {x\, dx}`	$\int_0^1 {x\, dx}$	Fails on `\,`.
`\int^1_{y=0} x\, dx`	$\int^1_{y=0} x\, dx$	Integral from 0 to 1 of the lambda expression $y \mapsto x$.
`\int \sin x \, dx`	$\int \sin x \, dx$	Indefinite integral of the lambda expression $x \mapsto \sin x$.
`\int \sin x`	$\int \sin x$	Indefinite integral of the expression $\sin x$ (function $\sin$ applied to variable named x) together with an XML comment that there may be an error in the integrand.
`\int \sin`	$\int \sin$	Indefinite integral of the expression $\sin$ (function $\sin$) together with an XML comment that there may be an error in the integrand. In this case the comment is not actually needed.

There is automatic recovery from use of the letter d as a variable outside of an integral, with the result that d gets treated as a variable unless it is in a place where it is 'expected' to have significance with regard to differentiation.

7 Calculus — derivatives

Support for derivatives is in a preliminary state so far.

Here are some examples

Input	**Typesets as**	**Interpretation by ml2om**
`\frac{dy}{dx}`	$\frac{dy}{dx}$	Derivative of a map (lambda expression) $x \mapsto y$, with derivative evaluated at x.
`\frac{d}{dx} x^2 + 2x - 1`	$\frac{d}{dx}x^2 + 2x - 1$	The sum of the evaluation at x of the derivative of a map (lambda expression) $x \mapsto x^2$ and the expression $2x - 1$.
`\frac{d (x^2 + 2x - 1)}{dx} z`	$\frac{d(x^2+2x-1)}{dx}z$	Product of a derivative and variable called z. The derivative is the evaluation at x of the derivative the (lambda expression) $x \mapsto x^2 + 2x - 1$.
`f'(x)`	$f'(x)$	Apply diff from calc CD to variable named f (here we are using the rule that f stands for a function by default) and then apply result to variable named x.
`g'(x)`	$g'(x)$	Error on $'$ as g is not recognised as a function name.
`f''(x)`	$f''(x)$	Apply partialdiff from calc CD to the list of two integers $(1, 1)$ and the variable named f (denotes the second partial with respect to argument 1 of f) and then apply result to a variable named x.

8 Matrices

Matrices are recognised when input using the LaTeX mechanism `\begin{array}`, the AMSLaTeX version `\begin{matrix}` or the PlainTeX mechansim `\matrix{`. The variants such as `\begin{pmatrix` `\begin{vmatrix}` and `\vmatrix{` are also recognised. In all cases, the output is filled out to a rectangular array with empty entries replaced by the integer 0.

Here are some examples

Input	Typesets as	Interpretation by ml2om

`\begin{array}{lll}`
`a & b\\`
`c & e & y`
`\end{array}`
	$\begin{array}{lll} a & b \\ c & e & y \end{array}$	Matrix with two rows, and three columns. First row has entries a, b, 0 and second row is c, e and y.

`\begin{matrix}`
`a & b\\`
`c & e \\`
`y`
`\end{matrix}`
	$\begin{matrix} a & b \\ c & e \\ y \end{matrix}$
	(with package `amsmath`)
		Matrix with three rows, and two columns. First row has entries a, b, second row c, e and third row y, 0.

`\matrix{`
`a & b\cr`
`c &  \cr`
`y & e`
`}`
	$\begin{matrix} a & b \\ c & \\ y & e \end{matrix}$	Matrix with three rows, and two columns. First row has entries a, b, second row c, 0 and third row y, e.

9 Vectors

Vectors are taken by ml2om to be comma-separated lists enclosed between left and right parentheses. It is intended to introduce a mechanism capable of recognising notations for intervals such as $(1, 2)$ or $[0, \infty)$ but there is no such mecahnism in place.

Here are some examples

Input	Typesets as	Interpretation by ml2om
`(1, 2, 3)`	$(1, 2, 3)$	List of 3 integers 1, 2, and 3. The list is attributed with the fact that it had left and right round parentheses by use of the `presentation` CD.
`[x +y, z + w, a + b, c)`	$[x + y, z + w, a + b, c)$	List of 4 items. The first 3 items are the sum of two named variables and the last item is just the variable c. Again attributes are generated to indicate the left [and right).