

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

ARTIFICIAL INTELLIGENCE PROJECT
AI Memo 97A

MAC-M-327
October 13, 1966

SYMBOLIC INTEGRATION - II

by Joel Moses

In this memo we describe the current state of the integration program originally described in AI Memo 97 (MAC-M-310). Familiarity with Memo 97 is assumed. Some of the algorithms described in that memo have been extended. Certain new algorithms and a simple integration by parts routine have been added. The current program can integrate all the problems which were solved by SAINT and also the two problems attempted by it and not solved. Due to the addition of a decision procedure the program is capable of identifying certain integrands (such as e^x or e^x/x) as not integrable in closed form.

Methods of solution to certain types of differential equations have also been coded. These use the integration program as a sub-routine.

The integration program has been dubbed SIN (for Symbolic INtegration.) Below we present the methods of integration used by SIN.

Here $R(z)$ and $S(z)$ are rational functions (ratio of two polynomials) in z . $\text{Elem}(z)$ is an elementary function of z . (See Slagle, page 6, for definition of elementary function.)

- I. $x^c \text{Elem}(x^{k_i})$, where c, k_i are integers, and where $k = \text{g.c.d. of the } k_i \text{ divides } (c + 1)$.

Substitute $y = x^k$

$$\int x^7 / (x^4 + 1) dx \text{ becomes } 1/4 \int y / (y^3 + 1) dy$$

$$\int 1/x \frac{x^2 - 1}{x^2 + 1} dx \text{ becomes } 1/2 \int \frac{W - 1}{W + 1} dy$$

- II. (Rational) $R(x)$. Chains to a program written by Manove which integrates rational functions by a method described in Hardy.³

- III. $\text{Elem}(\text{base}_1^{a_i + b_i x}, \text{base}_2^{a_i + b_i x}, \dots)$, where base_i, a_i, b_i are constants.

a. $\text{base}_i^{a_i + b_i x}$ is converted to $\text{base}_1^{(a_i + b_i x) \log_{\text{base}_1} \text{base}_i}$

b. $y = \text{base}_1^x$ is substituted

$$\int e^x (2 + 3e^{2x}) dx \text{ becomes } \int 1 / (2 + 3y^2) dy$$

$$\int 10^k e^k dx \text{ becomes } \int y^{k \log_e 10} dy$$

- IV. (Chebyshev) $Ax^r (c_1 + c_2 x^q)^{r_2}$. A, r, r_2, q, c_1, c_2 are constants.

Convert A to A/q , let $r_1 = \frac{r+1}{q} - 1$

a. r_1 integer, $r_1 > 0$

Substitute $y = \frac{c_1}{c_2} + c_2 x^q$

b. r_2 integer, r_1 rational number with denominator d_1

Substitute $y = x^{q/d_1}$

c. r_1 integer, $r_1 < 0$, r_2 rational with denominator d_2

Substitute $y = (c_1 + c_2 x^q)^{1/d_2}$

d. $r_1 + r_2$ is an integer

Substitute $y = \left(\frac{c_1 + c_2 x^q}{x^q} \right)^{1/d_2}$

V. Elem($x, \sqrt[n_1]{\frac{ax+b}{cx+d}}, \sqrt[n_2]{\frac{ax+b}{cx+d}}, \dots$) where the n_i are rational numbers.
 a, b, c, d are constants with $ad-bc \neq 0$.

Let k = least common multiple of the n_i

Substitute $y = \sqrt[k]{\frac{ax+b}{cx+d}}$

$\int \cos \sqrt{x} dx$ becomes $\int 2y \cos(y) dy$

$\int x \sqrt{x+1} dx$ becomes $\int 2(y^2 - 1)y^2 dy$

VI. $R(x, \sqrt{a+bx^2})$

a. $a > 0, b > 0$ Substitute $\tan(y) = \sqrt{a/b} x$

$\int 1/(4+x^2) dx$ becomes $\int 1/2 \sec(y) dy$

b. $a > 0, b < 0$ Substitute $\sin(y) = \sqrt{a/b} x$

$\int x/(1-x^2)^{3/2} dx$ becomes $\int \sin(y) dy$

c. $a < 0, b > 0$ Substitute $\sec(y) = \sqrt{-a/b} x$

VII. $R(x, \sqrt{cx^2+bx+a})$

a. $c > 0$ Substitute $y = \sqrt{c} x + \sqrt{cx^2+bx+a}$

b. $a \geq 0$ Substitute $y = (\sqrt{cx^2+bx+a} - \sqrt{a})/x$

VIII. $R(x) F(S(x))$, where $\int R(x) dx$ is rational

a. $F = \log$

Solution is $\log(S(x)) \int R(x) dx - \int \left[\int R(x) dx \right] S'(x)/S(x) dx$

$\int x \log(x) dx$ becomes $1/2 x^2 \log(x) - \int x/2 dx$

b. $F = \arctan$

Solution is $\arctan(S(x)) \int R(x) dx - \int \left[\int R(x) dx \right] S(x)/(1+S^2(x)) dx$

c. $F = \arcsin$

Solution is $\arcsin(S(x)) \int R(x) dx - \int \left[\int R(x) dx \right] S(x)/(1-S^2(x))^{1/2} dx$

$\int x^2 \arcsin(x) dx$ becomes $1/3 x^3 \arcsin(x) - \int 1/3 x^3/(1-x^2)^{1/2} dx$

IX. $R(x) \text{Elem}(\log(a+bx))$

Substitute $y = \log(a+bx)$

$\int \log(x)/(1+\log(x))^2 dx$ becomes $\int e^y y/(1+y)^2 dy$

X. Elem($F_i(a + bx)$) where the F_i are trigonometric functions.

I. Preparatory step $y = a + bx$ yielding $1/b$ Elem($F_i(y)$)

Note: Some code has been written for the case in which the arguments of the trigonometric functions are linear but not identical.

II. $\int \sin^m(y) \cos^n(y) dy$

a. $m < n$; transform into $\int (1/2 \sin 2y) (1/2 + 1/2 \cos 2y)^{n/2} dy$

b. $m \geq n$; transform into $\int (1/2 \sin 2y) (1/2 - 1/2 \cos 2y)^{m/2} dy$

$\int \cos^2(x) dx$ becomes $\int (1/2 + 1/2 \cos 2y) dy$

III. i) Transform all trigonometric functions into sines and cosines and test to see if the result is of the form

a. $\sin^{2n+1}(y) \text{Elem}(\sin^2(y), \cos(y))$

Substitute $Z = \cos(y)$

$\int \sin^3(y) \cos(y) dy$ becomes $\int (z - z^3) dz$

b. $\cos^{2n+1}(y) \text{Elem}(\cos^2(y), \sin(y)) dy$, substitute $Z = \sin(y)$

ii) Transform all trigonometric functions into secants and tangents and determine if the resulting expression is of the form

a. $\text{Elem}(\tan(y), \sec^2(y))$, substitute $z = \tan(y)$

$\int \sec^2(y) / (1 + \sec^2(y) - 3 \tan(y)) dy$

becomes $\int 1/(z^2 - 3z + 2) dz$

b. $\tan^{2n+1}(y) \text{Elem}(\sec(y), \tan^2(y)) dy$,
substitute $z = \sec(y)$

iii) Substitute $z = \tan 1/2 y$

XI. $R(x)e^{P(x)}$ where P is a polynomial function of x . See below for a description.

XII. a. In a sum each of the summands is integrated separately.

b. If the integrand is a small, positive integer power of a sum, the integrand is expanded and the result is integrated.

XIII. (Derivative Divides) See below for description.

XIV. (Integration by Parts) See below for description.

Procedure for Integrating $R(x)e^{f(x)}$ - Algorithm X.

In the previous memo we mentioned a decision procedure for integrands of the form $R(x)e^{f(x)}$ where R is rational and P is a (non-constant) polynomial. This is an extension of the decision procedure found in Ritt for the case $f(x) = x$. This algorithm proceeds by making guesses at the solution. This is the way the Edge heuristic would have proceeded. We have modified the heuristic slightly to arrive at this procedure.

$$\text{Let } R(x) = \frac{C_1 x^{m_1} + S_1(x)}{Q(x)}, \text{ where } S_1, Q \text{ are polynomials}$$

S_1 is a polynomial of degree $< m_1$
 C_1 is a constant

We know from Ritt that the integral (if any) will be a multiple of $e^{f(x)}$. Suppose the integral is represented by $(a_1(x) + b_1(x))e^{f(x)}$, then

$$P'(x)a_1 + a_1' + P'(x)b_1 + b_1' = R(x) = \frac{C_1 x^{m_1} + S_1}{Q}$$

Now, if we guess $a_1(x) = \frac{C_1 x^{m_1}}{P'Q}$ then

$$a_1' = \frac{m_1 C_1 x^{m_1-1} - \frac{C_1 x^{m_1} Q'}{P'Q} - \frac{C_1 x^{m_1} P''}{P'^2}}{Q}$$

$$\text{and } P'b_1 + b_1' = R(x) - P'a_1 - a_1' = \frac{S_1 - \frac{m_1 C_1 x^{m_1-1}}{P'} + \frac{C_1 x^{m_1} P''}{P'^2} + \frac{C_1 x^{m_1} Q'}{P'Q}}{Q}$$

Now, consider the numerator of $P'b_1 + b_1'$ as a polynomial $T_1(x)$ and a rational function remainder $U_1(x)$. Let the leading term of $T_1(x)$ be $c_1 x^{m_1}$. Continue the process indicated above until some T_i (T_n , say) is 0. If at that time the corresponding U_i (i.e. U_n) is also 0, then the expression $R(x)e^{f(x)}$ is integrable and the integral is the sum of the $a_i(x)$ multiplied by $e^{f(x)}$. If U_n is not 0, then the expression is not integrable.

Note that if $U_n = 0$, then $R(x) - P' \sum_{i=1}^n a_i - \sum_{i=1}^n a_i' = 0$.

Let $a = \sum_{i=1}^n a_i$, then we claim that $U_n = 0$ implies that

$$\int R(x)e^{f(x)} dx = a(x)e^{f(x)}, \text{ or}$$

$$aP'e^P + a'e^P = R \cdot P$$

$$(R - aP' - a')e^P = 0$$

but $e^P \neq 0$, hence $R - aP' - a' = 0$, but this is assured.

Examples

$$\int \frac{e^x x}{(x+1)^2} dx = \frac{1}{x+1}, \int e^{x^2} dx \text{ is not integrable}$$

$\int \frac{e^x}{x} dx$ is not integrable

Derivative Divides - Algorithm XIII

Let the integrand be $f(x)g(x)$ where $f(x)$ is each of the non-constant terms in the integrands considered in turn of the whole integrand if it is not a product of terms.

- a) If $f'(x) = \text{constant} * g(x)$, then $\int f(x)g(x) = \frac{f^2(x)}{2 * \text{constant}}$
- b) If f is a unary operator such as \sin , \cos and its argument is $k(x)$ and $k'(x) = \text{constant} * g(x)$ then substitute $y = k(x)$ and obtain $\int 1/\text{constant} f(y)dy$ by looking it up in a table. This table contains about ten entries.
- c) If f is a binary operator (i.e., expt or \log) then $k(x)$ is considered to be the non-constant argument (assuming only one argument is non-constant). Proceed as with step b above.

This routine is rather conservative in searching for a subexpression whose derivative divides the integrand. Since it is applied to each problem, our main considerations were speed and infallibility in making substitutions. Considering its limitations, it is a highly successful routine. For example, it helped solve forty-five out of the eighty-six SA.INT problems with an average time of 0.6 seconds on the 7094.

Example

$$\int \tan x \sec^2 x \, dx = 1/2 \tan^2 x$$

$$\int e^{\sin x} \cos x \, dx = e^{\sin x}$$

$$\int x^n \, dx = 1/n+1 x^{n+1}$$

$$\int \cos(ax + b) \, dx = 1/a \sin(ax + b)$$

Integration by Parts - Method XIV

At the present time the Edge heuristic discussed in our previous memo has not been written. We have, however, written a simple integration by parts routine similar to the one found in SAINT. The organization that this routine requires is closer to the organization found in SAINT than any other part of the program.

Consider any partition of the integrand into a product of non-constant factors g and h , where $H = \int h dx$ can be found without calling the integration by parts routine. (This is a modification of the SAINT method which required that the integral be found with an allotment of no resources.) Now consider $\int g h dx$. We require that this integral be found with the number of calls to the integration by parts routine fewer than a constant called Maxparts. This constant is calculated to be twice the maximum absolute value of an exponent of a top level term of the integrand. Thus, $x \cos(x)$ may use only two calls to integration by parts and $x^2 \cos(x)$ may use four.

If both integrals can be obtained, then the solution is

$$\int g h dx = g H - \int g' H dx.$$

There are some obvious difficulties with the approach taken above. First, there is no attempt to devise a reasonable partition of the integrand. Second, there is no routine which determines whether progress has been made in obtaining a solution. Both of these problems must be solved in a successful implementation of the Edge heuristic.

Note also that the value of Maxparts can be set only at the top level call to the integration by parts routine. Otherwise, we would wind up in an infinite loop. This requires the program to determine whether it is at the top level call to this routine. Although it is debatable whether we really require such a top-level determination in this routine, we predict that complex programs will increasingly make use of information regarding the history of the program. For a further discussion of this point see Teitelman.

Example

$$\int x \cos(x) dx = x \sin(x) - \int \sin(x) dx$$

Differential Equations

Routines for solving the following classes of differential equations have been written. The integration program is called as a subroutine in each case.

a) Separable

$$A(x)B(y)dx + C(x)D(y)dy = 0$$

$$\int A/C dx + \int D/B dy = 0$$

$$x(y^2 - 1)dx - y(x^2 - 1)dy = 0$$

$$\text{becomes } \int \frac{x}{x^2 - 1} dx - \int y/(y^2 - 1)dy = C_0$$

b) Linear

$$y' + P(x)y + Q(x) = 0$$

$$\text{becomes } ye^{\int P dx} + \int Qe^{\int P dx} dx = C_0$$

$$y' + y + x = 0 \text{ becomes } ye^x + \int xe^x dx = C_0$$

c) Bernoulli

$$y' + P(x)y + Q(x)y^n = 0$$

Substitute $v = y^{1-n}$ yielding the linear equation (case b)
 $v' + (1-n)P(x)v + (1-n)Q(x) = 0$

d) Exact

$$P(x,y)dx + Q(x,y)dy = 0$$

$$\text{where } \frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x} \quad (\text{The matching program SCHATCHEN is}$$

used to determine the equivalence.)

$$\text{becomes } \int P dx + \int (Q - \frac{\partial}{\partial y} [\int P dx]) dy = C_0$$

$$2xydx + (x^2 + \cos(y))dy = 0$$

$$\text{becomes } x^2y + \int \cos(y)dy = C_0$$

e) Homogeneous

$$P(x,y)dx + Q(x,y)dy = 0$$

where P and Q are homogeneous functions of the same degree, n, say. Substitute $y = vx$ and factor out x^n from the equation. Substitution yields variables separable (case a).

$$(2ye^{y/x} - x)dy + (2x + y)dx = 0$$

$$\text{becomes } (2v^2e^v - 1)x dv + 2(v^2e^v + 1)dx = 0$$

f) Linear coefficients

$$y' + F\left(\frac{ax + by + c}{a'x + b'y + c'}\right) = 0, \text{ where } a, b, c, a', b', c' \text{ are constants} \\ \text{and } a'b - ab' \neq 0$$

$$\text{substitute } x^* = x - \frac{b'c - bc'}{a'b - ab'}, y^* = y - \frac{ac' - a'c}{a'b - ab'}$$

and obtain a homogeneous equation (case e).

$$y' - \frac{4x - y + 7}{2x + y - 1} = 0 \text{ becomes } (2x^* + y^*)dy^* - (4x^* - y^*)dx^* = 0$$

g) Linear differential equations with constant coefficients

L. Ernst has recently written a program which solves certain cases of inhomogeneous differential equations with constant coefficients. The method used requires Laplace transforms which are obtained by a program written by C. Engleman². The latter program uses the above-mentioned program by Manove for integrating rational functions. We hope to use these programs as subroutines.

Extensions to SCHATCHEN

Two new modes have been added to the matching program SCHATCHEN. These have been found useful in certain situations.

Consider the expression $A(x)B(y)dx + C(x)D(y)dy$ found in the section on differential equations when discussing separable equations (case a). What is desired here is that both A and B be a product of terms involving x and y respectively. It was not previously possible to describe to SCHATCHEN a simple matching rule which would yield a subset of the terms in a product (or, for that matter, in a sum). The indicators TIMES** and PLUS** are used for this purpose. For instance, the expression above can be matched using the following pattern:

```
(PLUS (TIMES (TIMES** A FREEY) (TIMES** B FREEEX)DX)  
(TIMES (TIMES** C FREEY) (TIMES** D FREEEX)DY))
```

where FREEEX and FREEY are functions which test expressions for independence of x and y respectively.

REPLACE

A complementary function to SCRATCHEN has been written which performs the function of a right-hand-side of a Convert rule, that is, it generates an expression by replacing in a form the values obtained from a dictionary supplied by SCRATCHEN. REPLACE is a function of two arguments Dict and Expr. Dict is a list of pairs; Expr is an expression to be transformed. If we let Dict be $((u_1, v_1) (u_2, v_2) \dots (u_n, v_n))$, then the most straightforward use of REPLACE is to substitute v_i for each occurrence of u_i in Expr. REPLACE will also attempt to simplify the results obtained during the substitution as the new expression is being formed. It avoids certain redundant simplifications by simplifying only at the level at which it is working and not at lower levels. (These are assumed simplified through the recursion process.) REPLACE also uses two special indicators suggestively called Eval* and Quote*. REPLACE(DICT(EVAL* A)) is the simplified result of (EVAL(REPLACE DICT A)ALIST). REPLACE(DICT(QUOTE* A)) is simply A. By using these indicators we can get some of the dynamic capabilities of CONVERT which are lacking in most other string transformation languages.

Example

```
(REPLACE
(QUOTE ((A. - 1) (N,1)))
(QUOTE (PLUS X (TIMES A (EXPT XN)))))) = 0
(REPLACE
(QUOTE ((A,1)(B,1)(X,Y)))
(QUOTE (TIMES X (EVAL* (FACT (PLUS A B)))))) = (TIMES 2 Y),
      where FACT is the factorial function.
```

- 1) Singh, J. L., A Heuristic Program That Solves Symbolic Integration Problems in Riemann Calculus, Ph.D. Thesis, M.I.T., 1967.
- 2) Eshleman, C., Maslove, H., Blom, S., Rational Functions in MANIAC. Proc. of ISEE Working Conference on Symbol Manipulation Languages, Pisa, Italy, September, 1966.
- 3) Hardy, G. H., The Integration of Functions of a Single Variable, 2nd Ed., Cambridge U. Press, Cambridge, England, 1916.
- 4) Knott, J. P., Integration in Finite Terms, Liouville's Theory of Elementary Methods, Columbia Univ. Press, New York, 1948.
- 5) Teitleman, W., Pilot - A Step Toward Man-Computer Symbiosis Ph.D. Thesis, M.I.T., September, 1966.