SYMBOLIC AND NUMERIC INTEGRATION IN MAPLE

Keith Geddes and George Labahn
Department of Computer Science
University of Waterloo, Waterloo, Ontario, Canada
e–mail: {kogeddes, glabahn}@daisy.uwaterloo.ca

1 Introduction

The Maple system is designed to support both symbolic and numeric modes of mathematical computation. In this paper, we give a survey of the algorithms used in Maple to deal with the integration problem of calculus, in both symbolic and numeric contexts.

2 Indefinite Integration

Given a function $f(x)$, the indefinite integration problem asks whether there exists a function $g(x)$ such that $g'(x) = f(x)$. We assume that $f(x)$ and $g(x)$ are both represented as Maple expressions; specifically, as rational expressions in x and various mathematical functions. We will write

$$\int \mathrm{f}(x)\, dx = \mathrm{g}(x)$$

where we do not explicitly write the arbitrary constant of integration.

The algorithmic approach to the indefinite integration problem is based on the Risch integration algorithm. In this presentation, it is only possible to give a very brief outline of a part of the Risch algorithm. For a more detailed exposition of the algorithm, see [1],[4].

2.1 Liouville's Principle

Recall that a field $K(x)$ of rational functions in the variable x over a constant field K consists of all functions which can be represented as a quotient of polynomials, where the numerator and denominator are each polynomials in x with coefficients from the constant field K. For our purposes, the constant field K will be a subfield of the field of complex numbers. For example, K could be the field Q of rational numbers or, more generally, an extension field

$$K = \mathrm{Q}\,(\alpha_1, ..., \alpha_k)$$

where each α_i is an algebraic extension (i.e. the solution of a polynomial equation over a lower field in the hierarchy). Examples of constant algebraic extensions are the complex number i satisfying $i^2+1=0$, $\sqrt{2}$, or any n-th root of a constant.

To define a new class of functions, we extend the field $K(x)$ to include additional functions. The notation

$$F_n = \mathrm{K}\,(x, \theta_1, ..., \theta_n)$$

denotes the set of all rational expressions in the symbols $x, \theta_1, ..., \theta_n$ over the constant field K. Each extension θ_i will be a new function of x. The extension θ_i is called an *elementary extension* of the field $F_{i-1} = K(x, \theta_1, ..., \theta_{i-1})$ if it can be expressed in one of the following forms: (i) $\theta_i = log(u)$ for some $u \in F_{i-1}$; (ii) $\theta_i = exp(u)$ for some $u \in F_{i-1}$; or (iii) θ_i is *algebraic* over F_{i-1} (i.e. there exists a polynomial $p \in F_{i-1}[z]$ such that $p(\theta_i) = 0$).

The more restrictive class of *transcendental elementary extensions* are those elementary extensions which are not algebraic. A function field F_n is called a *field of elementary functions* if it can be expressed in the above form where each θ_i is an elementary extension, and it is called a *field of transcendental elementary functions* if each θ_i is a transcendental elementary extension.

Note that the field of elementary functions includes all of the mathematical functions commonly encountered in a beginning calculus course, including the exponential, logarithm, trigonometric, inverse trigonometric, hyperbolic, and inverse hyperbolic functions, as well as the square root, cube root, and all other algebraic functions. This relies on the fact that functions such as `sin(x)`, `arctan(x)`, `cosh(x)`, `arcsech(x)` (and the others) can be expressed in terms of exponentials and logarithms if we allow complex numbers in the constant field. The advantage of expressing all of the elementary functions in terms of exponentials, logarithms, and algebraic functions is that this imposes some structure on the possible form of the integral.

For the case of rational function integration, if $f(x) \in K(x)$ then there exists an elementary function $g(x)$ which is an indefinite integral of $f(x)$. We cannot say that $g(x)$ is necessarily a rational function. However, we can state that $g(x)$ lies in an extension field $K(x, \theta_1, ..., \theta_n)$ where each θ_i is a logarithmic extension (with possibly some algebraic number extensions of the constant field K). In other words, the only new functions required are logarithms. Surprisingly, a similar result holds for the field of elementary functions.

Liouville's Principle: Let F be a field of elementary functions with constant field K. For $f \in F$ suppose that the integral

$$\mathrm{g}(\,x\,) = \int \mathrm{f}(x)\, dx$$

is an elementary function. Specifically, suppose that $g \in G$ where G is an elementary extension of the field F having the same constant field K. Then there exist functions $v_0, v_1, ..., v_m \in F$ and constants $c_1, ..., c_m \in K$ such that

$$\mathrm{g}(x) = v_0 + \left(\sum_{i=1}^{m} c_i \ln(v_i)\right)$$

2.2 The Risch algorithm for a transcendental log extension

In the space available here, the general form of the Risch algorithm for the case of transcendental elementary functions will be outlined by concentrating on the case of a logarithmic extension. The case of an exponential extension is quite similar. Computation with algebraic functions involves a rather different set of mathematical concepts and will not be pursued in this presentation.

Let $f \in K(x, \theta_1, ..., \theta_n)$ and suppose that the last extension $\theta = \theta_n$ is a transcendental log extension, say $\theta = log(u)$ for $u \in F_{n-1} = K(x, \theta_1, ..., \theta_{n-1})$. We view the integrand as a rational function in the variable θ; specifically, $f(\theta) = p(\theta)/q(\theta) \in F_{n-1}(\theta)$. Without loss of generality, we may assume that $f(\theta)$ is normalized such that $p(\theta), q(\theta) \in F_{n-1}[\theta]$ satisfy $GCD(p(\theta), q(\theta)) = 1$ and that $q(\theta)$ is monic.

The steps in the Risch algorithm are as follows.

1. Apply Euclidean division to the polynomials $p(\theta), q(\theta) \in F_{n-1}[\theta]$ yielding polynomials $s(\theta), r(\theta) \in F_{n-1}[\theta]$ such that $p(\theta) = q(\theta)s(\theta) + r(\theta)$ with $r(\theta) = 0$ or $deg(r(\theta)) < deg(q(\theta))$. Then
$$\int \mathrm{f}(\theta)dx = \int \mathrm{s}(\theta)dx + \int \frac{\mathrm{r}(\theta)}{\mathrm{q}(\theta)}dx$$
The first term on the right hand side is called the integral of the *polynomial part* and the second term is called the integral of the *rational part*.

Integration of the Rational Part

2. Compute the square-free factorization of the denominator $q(\theta) \in F_{n-1}[\theta]$:
$$\mathrm{q}(\theta) = \prod_{i=1}^{k} q_i(\theta)^i$$
where each $q_i(\theta)$ $(1 <= i <= k)$ is monic and square-free.

3. Compute the partial fraction expansion
$$\frac{\mathrm{r}(\theta)}{\mathrm{q}(\theta)} = \sum_{i=1}^{k}\sum_{j=1}^{i} \frac{r_{i,j}(\theta)}{q_i(\theta)^j}$$
where each $r_{ij}(\theta) \in F_{n-1}[\theta]$ and $deg(r_{ij}(\theta)) < deg(q_i(\theta))$. Then
$$\int \frac{\mathrm{r}(\theta)}{\mathrm{q}(\theta)}dx = \sum_{i=1}^{k}\sum_{j=1}^{i} \int \frac{r_{i,j}(\theta)}{q_i(\theta)^j}dx$$

4. Apply Hermite reduction to reduce any power $j > 1$ in the denominator. For a particular term $r_{ij}(\theta)/q_i(\theta)^j$ with $j > 1$, by applying the Extended Euclidean Algorithm it is possible to compute polynomials $s(\theta), t(\theta) \in F_{n-1}[\theta]$ such that

$$\mathrm{s}(\theta)q_i(\theta) + \mathrm{t}(\theta)\frac{\partial q_i(\theta)}{\partial x} = r_{i,j}(\theta)$$

In terms of $s(\theta), t(\theta)$, the Hermite reduction formula takes the following form.

$$\int \frac{r_{i,j}(\theta)}{q_i(\theta)^j}dx = -\frac{\mathrm{t}(\theta)}{(j-1)q_i(\theta)^{(j-1)}} + \int \frac{\mathrm{s}(\theta) + \frac{1}{j-1}\frac{\partial t(\theta)}{\partial x}}{q_i(\theta)^{(j-1)}}dx$$

By repeated application of this reduction process, the integral of the rational part takes the following form:

$$\int \frac{\mathrm{r}(\theta)}{\mathrm{q}(\theta)}dx = \frac{\mathrm{c}(\theta)}{\mathrm{d}(\theta)} + \int \frac{\mathrm{a}(\theta)}{\mathrm{b}(\theta)}dx$$

where $a(\theta), b(\theta), c(\theta), d(\theta) \in F_{n-1}[\theta]$, $deg(a(\theta)) < deg(b(\theta))$, and $b(\theta)$ is monic and square-free.

5. *Rothstein/Trager method.* For the integration of the rational part, it remains to specify how to compute the integral of a proper rational function $a(\theta)/b(\theta)$. By a method attributed independently to Rothstein [9] and Trager [12], we compute the following *polynomial resultant* $R(z) \in F_{n-1}[z]$:

$$\mathrm{R}(z) = res_\theta(\mathrm{a}(\theta) - z\frac{\partial \mathrm{b}(\theta)}{\partial x}, \mathrm{b}(\theta))$$

Then the integral of $a(\theta)/b(\theta)$ is an elementary function if and only if all the roots of the polynomial $R(z)$ are constants (i.e. free of the integration variable x). If all roots of $R(z)$ are constants, let c_i $(1 <= i <= m)$ be the distinct roots of $R(z)$ and define $v_i(\theta)$ $(1 <= i <= m)$ by

$$v_i(\theta) = \mathrm{GCD}(\mathrm{a}(\theta) - c_i\frac{\partial \mathrm{b}(\theta)}{\partial x}, \mathrm{b}(\theta))$$

Then

$$\int \frac{\mathrm{a}(\theta)}{\mathrm{b}(\theta)}dx = \sum_{i=1}^{m} c_i \ln(v_i(\theta))$$

Example 2.1

If we apply the Risch algorithm to the integral

$$\int \frac{1}{x\ln(x)}dx$$

the resultant computation for this problem is

```
> a := 1:  b := x*theta:
> diff_b := diff(x*ln(x), x):
> diff_b := subs(ln(x) = theta, diff_b);
```

$$diff\_b := \theta + 1$$

```
> R := resultant(a - z*diff_b, b, theta);
```

$$R := x(-1+z)$$

The single root of $R(z)$ is $z = 1$ and the integral is elementary. Specifically, $c_1 = 1$, $v_1 = \theta$ and the integral is $c_1\, ln(v_1) = ln(\theta)$. In other words,

$$\int \frac{1}{x\ln(x)} dx = \ln(\ln(x))$$

Integration of the Polynomial Part

6. The problem is to compute (if it exists as an elementary function) the integral $q(\theta)$ defined by

$$\int \mathrm{p}(\theta) dx = \mathrm{q}(\theta)$$

where the given integrand is $p(\theta) \in F_{n-1}[\theta]$ and $\theta = ln(u)$ for some $u \in F_{n-1}$. A careful argument based on Liouville's Principle determines that if

$$\mathrm{p}(\theta) = p_l \theta^l + p_{l-1}\theta^{(l-1)} + \ldots + p_0$$

then $q(\theta)$, if it exists as an elementary function, must be of the form

$$\mathrm{q}(\theta) = q_{l+1}\theta^{(l+1)} + q_l\theta^l + \ldots + q_1\theta + q_0$$

Applying differentiation, the relationship between the given integrand $p(\theta)$ and the unknown result $q(\theta)$ yields the following system of equations to solve.

$$0 = \frac{\partial q_{l+1}}{\partial x}$$

$$p_l = (l+1) q_{l+1} \frac{\partial \theta}{\partial x} + \frac{\partial q_l}{\partial x}$$

$$p_{l-1} = l q_l \frac{\partial \theta}{\partial x} + \frac{\partial q_{l-1}}{\partial x}$$

$$\ldots$$

$$p_1 = 2 q_2 \frac{\partial \theta}{\partial x} + \frac{\partial q_1}{\partial x}$$

$$p_0 = q_1 \frac{\partial \theta}{\partial x} + \frac{\partial q_0}{\partial x}$$

These equations can be solved systematically in the order stated, by recursively applying the integration algorithm in the field F_{n-1} which is free of the extension θ. Note that the given coefficients $p_i, 0 <= i <= l$ all lie in the field F_{n-1} and it is important when solving these equations that the solutions $q_i, 1 <= i <= l+1$ must all lie in the field F_{n-1} (possibly with some extensions to the constant field). However, the final coefficient q_0 is allowed to contain new log extensions.

The integration of the "polynomial part" involves recursive application of the Risch algorithm. If any of the sub-integrals which arise is not elementary then the original integral is not elementary. Furthermore, if any of the sub-integrals is elementary but contains a new log extension (other than the particular extension θ) then the original integral is not elementary.

2.3 Extensions of the Risch algorithm

When the Risch algorithm determines that a particular integral "cannot be expressed", this conclusion is relative to a particular class of "allowable functions". The Risch algorithm is usually presented in the context of the class of *elementary functions* described earlier. Ideally, one would like to broaden the class of functions to include the various Special Functions such as the error function, gamma function, exponential integrals, and many others. In fact, the theoretical framework of the Risch algorithm has been extended to only a few non-elementary extensions (see [2], [10].

A computer algebra system such as Maple has knowledge about various Special Functions and it expresses integrals in terms of these functions. However, failure to express a particular indefinite integral in closed form does not imply that a decision procedure has determined that there is no closed-form expression for the integral. For the case of definite integrals, various techniques are applied to attempt to express the result as a symbolic formula (see [6] and the references therein for details).

3 Definite Integration: Hybrid Symbolic-Numeric Techniques

We turn now to a different topic: *numerical integration*. For many definite integration problems, a computer algebra system may be unable to express the result in symbolic closed form. Even if there is a closed-form formula, perhaps all that is actually required is a numerical value for the definite integral (to some precision).

Our purpose here is to examine the potential advantages to be gained by exploiting a hybrid *symbolic-numeric* approach to the problem of obtaining a numerical value for a definite integral. For some additional details on this topic see [5].

Traditional numerical integration codes typically require the interval of integration to be finite and the integrand to be finite at every point of the interval, although some numerical routines have been written to handle specific classes of singularities. The Maple computer algebra system is a natural environment in which to treat singularities. By using generalized series expansions, variable transformations, and other techniques, it is possible to automatically handle the numerical integration problem in the presence of various types of singularities. Moreover, integrals can be computed to arbitrarily high precision.

Example 3.1

Maple provides a facility to "evaluate in floating-point mode" those definite integrals which are left unevaluated by the symbolic techniques of the `int` command, or which are deliberately left unevaluated by using the inert `Int` command.

```
> f := sin(x)*ln(x)*exp(-x^3):
> int(f, x=0..infinity);
```

$$\int_0^\infty \sin(x)\ln(x)\mathrm{e}^{(-x^3)}\,dx$$

```
> evalf(");
```

$$-.1957885158$$

```
> evalf("", 25);
```

$$-.1957885158487997538390572$$

The computation of the above numerical results employs several of the techniques discussed in the following sections.

3.1 The case of an analytic integrand on a finite interval

Along with the techniques for symbolic analysis discussed here, we need an efficient routine for numerical integration in the simple case where the interval of integration is finite and the integrand has no singularities on the interval. Most traditional numerical integration methods could serve for this purpose. However, since the aim of the symbolic analysis phase is to present the numerical integration phase with an analytic integrand on a finite interval, we have found that the preferred default method is Clenshaw-Curtis quadrature [8], [3]. This method, named `ccquad`, is based on computing the Chebyshev series expansion of the integrand on the specified interval and then integrating term-by-term. Clenshaw-Curtis quadrature is at its best when the integrand is analytic in a sufficiently large region of the complex plane surrounding the interval of integration, in which case the Chebyshev series expansion converges quickly.

In Maple, `ccquad` is programmed to detect when convergence is too slow (i.e. there is a nearby singularity, perhaps on the real line, perhaps somewhere else in the complex plane) and to return control to the main routine when this occurs. The singularity-handling techniques are then invoked. If attempts to remove the singularity are unsuccessful, the code invokes another numerical integration routine named `quadexp`, an adaptive double-exponential quadrature method [11]. The strength of this method is that it is less sensitive than `ccquad` to the presence of singularities near to the interval of integration.

3.2 Generalized series expansions

The main tool for analyzing integrand singularities is a facility for computing generalized series expansions of expressions, called hierarchical series [7]. The standard Maple syntax for generating series expansions is

```
series(expr, var = value, ord);
```

although some of the most general series expansions which are needed are not returned by the `series` command in current versions of Maple. (There are special series expansion routines associated with the numerical integration routines.)

A generalized series expansion may not be a pure power series in the variable x but it may involve terms in $ln(x)$, $exp(-1/x)$, $exp(-1/x^2)$, etc. These series expansions will be understood to be one-sided expansions for $x > 0$. As long as the non-polynomial functions introduced into the series expansion can be bounded by powers of x, the mathematical validity of the series can be verified.

3.3 Transforming to an analytic integrand

If an integrand is found to be non-analytic at a point of the interval of integration, the first technique tried is to look for a transformation which will remove the singularity. The types of transformations which are attempted include: (i) subtracting off the singularity, and (ii) an algebraic transformation of variables. In each case, the fundamental tool for symbolic analysis is the generalized series expansion discussed in the preceding section.

Let f be the integrand, x be the variable of integration, $x = 0$ be the singular point, and s be the generalized series expansion of f about $x = 0$.

For case (i), the method used is to test each term in the expansion s to determine which terms are regular (i.e. have a Taylor series expansion at $x = 0$). If the number of non-regular terms is less than half the number of terms in s, then make the conjecture that the expression $f - q$ might be analytic at $x = 0$, where q denotes the sum of the non-regular terms. Test this conjecture, and if it is true then the numerical integration of $f - q$ can proceed normally. In some cases, the non-regular part q will be integrable by the symbolic integrator; otherwise, it will be passed on to the general techniques discussed in subsequent sections.

Example 3.2

```
> h := ln(1 - cos(2*x)):
```

The generalized series expansion of h at $x = 0$ takes the following form.

```
> series(h, x=0, 8);
```

$$\left(\ln(2) - 2\ln\left(\frac{1}{x}\right)\right) - \frac{1}{3}x^2 - \frac{1}{90}x^4 - \frac{2}{2835}x^6 + \mathrm{O}(x^8)$$

The non-regular part is $q = -2\,ln(1/x)$. The new expression $h - q$ is analytic on the interval [0, 1] and therefore it can be integrated easily by `ccquad`.

```
> r1 := evalf(Int(h + 2*ln(1/x), x=0..1));
```

$$r1 := .5797067686$$

(computing to 10 digits of accuracy). Integrating q can be done symbolically.

```
> r2 := int(-2*ln(1/x), x=0..1);
```

$$r2 := -2$$

Finally, summing the two values yields the desired result.

```
> Int(h, x=0..1)  =  r1 + r2;
```

$$\int_0^1 \ln(1 - \cos(2x))dx = -1.420293231$$

Case (ii) is the case of an algebraic transformation of variables. This method comes into effect whenever there are fractional powers of x appearing in the series expansion s, whether or not there are other non-regular functions appearing in the expansion. The idea is to transform away algebraic singularities, and if other singularities remain they will be handled on a subsequent pass. The specific method used is to compute the least common multiple n of all denominators of the fractional powers, and then to apply the change of variables $t = x^{1/n}$.

3.4 Direct integration of a generalized series

Suppose that we have an integrand with a singularity at the left end-point $x = 0$. Suppose further that we have not succeeded in finding a transformation which removes the singularity. Then the method used is to treat the generalized series expansion as an approximation of the integrand near the singularity, and to directly integrate this generalized series over its interval of accuracy. The remainder of the interval can then be handled by the numerical integration method.

The success of this technique relies on the power of Maple's symbolic integrator to handle many of the singular functions which may arise. For series involving $ln(x)$, the indefinite integral of $ln(x)\, x^k$ is an elementary function. For series involving $exp(-1/x)$, the indefinite integral of $exp(-1/x)\, x^k$ can be expressed in terms of the exponential integral $Ei(t)$. Similarly, series involving the singular function $exp(-1/x^2)$ may lead to both the exponential integral and the error function $erf(t)$.

References

[1] M. Bronstein, Integration of elementary functions. J. Symbolic Comp, 9 (2), 1990, pp. 117-173.

[2] G.W. Cherry. Algorithms for Integrating Elementary Functions in Terms of Logarithmic Integrals and Error Functions. Ph.D Thesis, Univ. of Delaware, August 1983.

[3] K.O. Geddes. Remark on Algorithm 424: Clenshaw-Curtis quadrature. ACM Trans. Math. Software, 5 (2), 1979, p. 240.

[4] K.O. Geddes, S.R. Czapor, and G. Labahn. Algorithms for Computer Algebra. Kluwer Academic Publishers, Boston, 1992.

[5] K.O. Geddes and G.J. Fee. Hybrid symbolic-numeric integration in Maple. Proceedings of ISSAC '92, ACM Press, New York, 1992, pp. 36-41.

[6] K.O. Geddes, M.L. Glasser, R.A. Moore, and T.C. Scott. Evaluation of classes of definite integrals involving elementary functions via differentiation of special functions. Applicable Algebra in Engineering, Communication and Computing, 1 (2), November 1990, pp. 149-165.

[7] K.O. Geddes and G.H. Gonnet. A new algorithm for computing symbolic limits using hierarchical series. Symbolic and Algebraic Computation, P. Gianni (ed.), Lecture Notes in Computer Science, No. 358, Springer-Verlag, Berlin, 1989, pp. 490-495.

[8] W.M. Gentleman. Algorithm 424: Clenshaw-Curtis quadrature. Comm. ACM, 15 (5), 1972, pp. 353-355.

[9] M. Rothstein. Aspects of Symbolic Integration and Simplification of Exponential and Primitive Functions. Ph.D. Thesis, Univ. of Wisconsin, Madison, 1976.

[10] M. Singer, B. Saunders, and B.F. Caviness. An extension of Liouville's theorem on integration in finite terms. SIAM J. Computing, pp. 966-990, 1985.

[11] Hidetosi Takahasi and Masatake Mori. Double exponential formulas for numerical integration. Publ. RIMS, Kyoto Univ., 9, 1974, pp. 721-741.

[12] B. Trager. Integration of Algebraic Functions. Ph.D. Thesis, Dept. of EECS, M.I.T., 1984