THE CLOSED-FORM INTEGRATION OF ARBITRARY FUNCTIONS

A.D. Fitt, Faculty of Mathematics, University of Southampton

G.T.Q. Hoare, Dr. Challoner's Grammar School, Amersham, Bucks

1 Introduction

Consider the common experience of a student who progresses far enough in school mathematics to begin to study the calculus. After motivation of the topic of 'rates of change', simple differentiation is learned. More advanced functions are then considered, and eventually the student meets the product, quotient and chain rules. The result: with enough algebraic accuracy and persistence the student can determine derived functions for virtually any sufficiently well-behaved combination of standard functions. A natural continuation of calculus is the determination of areas under curves and volumes of revolution, which entails discussion of the fundamental theorem of calculus. This asserts that the 'area function' $g(x)$ defined by

$$g(x) = \int_a^x f(t)dt$$

has derivative $f(x)$. This leads to the problem of finding 'antiderivatives' - given a function $f(x)$, can we determine another function $g(x)$ such that $g'(x) = f(x)$? Again, some standard examples are usually considered first, before integration by substitution, parts and other methods are examined. Indeed, the beginnings of an algorithmic approach are considered based essentially on Hermite's result that the integral of a rational function of one real variable is elementary (see below for definition) since it is a linear combination of logarithms, inverse tangents and rational functions.

Consider now the student's plight ! Does he or she not have a right to expect that, since (i) the link between differentiation and integration has been emphasized throughout and (ii) he or she can calculate 'any' derivative, it should be possible to perform any integration? Evidence that this is not the case normally involves such functions as e^{-x^2}, $\sin x/x$, or $\sqrt{\sin x}$ but the message to the student, once understood, is the same; although the problem of differentiation is formally solved, finding an antiderivative for a given function $f(x)$ at best involves locating the correct technique in a somewhat arbitrary manner, and at worst complete failure. The fact that similar integrands (consider, for example $(1+x)^{-2}$, $(1+x)^{-1}$, $x(1+x^2)^{-1}$, $(1+x^2)^{-2}$ and $(1-x^2)^{-1}$) give wildly different answers serves only to add to the confusion.

Bearing this in mind, is there any way of tackling the integration problem, other than by using a combination of random techniques, experience and luck ? The purpose of this note is to point out that not only are there better ways of proceeding, (some of which are of great practical use and are employed in the new generation of symbolic algebra computation packages) but also that in some senses, the integration problem may be regarded as completely solved. It is surprising how little known this result seems to be. Even more surprising is the fact that the theory was initially established by Liouville (who was probably in turn motivated by some conjectures of Abel) more than 150 years ago, and, for the most part, does not require a great degree of mathematical sophistication. It should be stressed that in all of the discussion below, we are concerned only with the *formal* problem of finding antiderivatives, rather than foundational notions such as the Lebesgue integral or measure theory.

2 Some Definitions

We assume throughout that we are interested in the purest form of the integration problem: given $f(x)$ can we determine an antiderivative $g(x)$ such that $g'(x) = f(x)$? As usual, we do not expect $g(x)$ to be determined uniquely, as an arbitrary constant may always be added. This additive constant is irrelevant for the purposes of the discussion below, however, and is taken to be zero throughout.

Of course, we have to decide what we regard as 'allowable candidates' for a solution. In this article, we shall restrict our attention to **elementary** functions. By elementary functions, we mean those built up from rational functions of x by successively exponentiating, taking logarithms, and performing algebraic operations (that is, solving polynomial equations whose coefficients are previously defined functions). We also choose to use not only real, but complex numbers throughout, since the set of elementary functions then includes sines, cosines etc. as well as their inverses, by using, for example, Euler's formula $\cos x = (e^{ix} + e^{-ix})/2$. For our purposes, this definition of the elementary functions will prove sufficient, and the reader is encouraged to equate 'elementary functions' with 'functions that may be built up with a scientific calculator using a finite number of operations'. It is worth making the point however that for total rigour a little more care is required in defining elementary functions, since technical complications may occur when dealing with multivalued functions. Full details are given in Ritt (1948).

It is also important to realize that the point at issue here does not concern 'whether an answer exists', but if an answer exists *in a certain form*. An analogy may be drawn with the solution of polynomial equations: given say a quadratic equation we could either determine the roots by using the standard formula, or employ a numerical method such as Newton-Raphson. In the spirit of the present discussion we are interested only in the former 'method of solution', that is, *is there a formula*? Of course for the general polynomial equation it was proved by Galois that a general formula exists only when the polynomial has degree less than or equal to 4, but what can be said about the integration problem? Some feel for the sort of results that may be established may be gained from the result that

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

for $n \neq -1$. What can be said in the special case $n = -1$? It is possible that the answer could still be a rational function? That this cannot be the case may be easily shown by contradiction, for if it were true that

$$\int \frac{1}{x} dx = \frac{P(x)}{Q(x)} \tag{1}$$

where P and Q are coprime (all common factors cancelled) polynomials, then by differentiation

$$\frac{1}{x} = \frac{QP' - PQ'}{Q^2}.$$

This gives

$$Q^2 = x(QP' - PQ')$$

and thus Q has a zero at $x = 0$. Assuming that $Q = x^n R(x)$ where $R(0) \neq 0$ and dividing by x^n we have

$$x^n R^2 = xRP' - nPR - xPR',$$

but this implies that the term PR is zero at $x = 0$, contradicting the assumption that neither P nor R has a zero at $x = 0$. Thus the integral of $1/x$ cannot have the form given by (1).

3 The Theorem of Liouville

Liouville's most important theorem on the problem of integration concerns the form that a primitive must take, if it is to be elementary. Davenport et al. (1988) state the theorem as follows:

Theorem (Liouville, 1833) Let f be a function from some function field K. If f has an elementary integral over K, it has an integral of the form

$$\int f = v_0 + \sum_{i=1}^{n} c_i \log v_i,$$

where v_0 belongs to K, the v_i belong to $\hat{K}$, an extension of K by a finite number of constants algebraic over K, and the c_i belong to $\hat{K}$ and are constant.

There may be terms in this statement with which some readers are unfamiliar, but the technical details of the theorem are not required for what follows. The theorem essentially states that if f has an elementary integral, then (by differentiation) it must be of the form

$$f = v_0' + \sum_{i=1}^{n} \frac{c_i v_i'}{v_i}.$$

For the context we have in mind here, the function field K is simply that of the elementary functions.

This theorem is obviously one of great power, but the proof (see, for example Rosenlicht (1972)), which is based on induction, is quite short and requires only a minimal amount of specialist knowledge. We content ourselves here by observing that if, for example, f is algebraic, (any function constructed in a finite number of steps from the operations of addition, subtraction, multiplication and division, the extraction of integral roots, and from the inverses of any functions already constructed) then $\int f$ cannot include exponentials since, roughly speaking, exponentials survive differentiation. The same applies to any logarithmic term involved in $\int f$ unless it enters the expression in a linear fashion. Also the v_i may be shown to be algebraic since logarithms of elementary functions, for example, also 'partially' survive.

4 Functions without Elementary Primitives

In what follows, we exploit Liouville's theorem to obtain a stronger version of the general result that is applicable to some special cases. Specifically, it is asserted that:-

Theorem ('Rational Liouville theorem') Let f and g be algebraic functions of x and assume that g is not a constant. Then if the integral of fe^g is elementary, it is given by

$$\int fe^g dx = Re^g$$

where R is rational in f, g and x.

For the sake of completeness, we give a proof of this result adapted from Ritt (1948) p. 47. This proof is not difficult, but a knowledge of elementary partial differentiation is necessary and the reader may wish to omit it at the first reading. The key section of the proof makes use of the result that in an *identity* (as opposed to an equality) between two functions, the independent variable may be replaced by any other variable (including one that contains a parameter) and the expression may be differentiated and integrated. As an example of this, consider the well-known trigonometrical identity

$$\sin 2\theta = 2\sin\theta\cos\theta.$$

Suppose that we replace θ by $\mu\theta$ and differentiate the identity with respect to μ. This yields

$$2\theta\cos 2\mu\theta = 2\theta\cos^2\mu\theta - 2\theta\sin^2\mu\theta.$$

Now setting $\mu = 1$ yields the other 'double angle' formula

$$\cos 2\theta = \cos^2\theta - \sin^2\theta.$$

Many readers will no doubt have encountered manipulations similar to this which commonly occur in the calculation of the envelopes of families of curves. The only other result that we ask the reader to assume is that the exponential of a non-constant algebraic function cannot itself be an algebraic function, and is henceforth referred to as 'transcendental'.

Proof of Rational Liouville theorem Assume that f and g satisfy the conditions of the theorem, and that the integral of fe^g is denoted by u. By Liouville's theorem, if u is elementary then it must have the form

$$u = v_0 + \sum_{i=1}^{n} c_i \log v_i.$$

Moreover, denoting $e^{g(x)}$ by θ, this may be written

$$u = v_0(\theta, x) + \sum_{i=1}^{n} c_i \log v_i(\theta, x). \tag{2}$$

with each of the v_i rational in θ, x, f and g. The equation (2) is an identity in the variable x, but now we regard θ and x as independent variables. Then differentiating with respect to x and using the chain rule we have

$$\frac{du}{dx} = f\theta = \theta g'(x)\frac{\partial v_0}{\partial \theta} + \frac{\partial v_0}{\partial x} + \sum_{i=1}^{n} \frac{c_i}{v_i}(\theta g'(x)\frac{\partial v_i}{\partial \theta} + \frac{\partial v_i}{\partial x}). \tag{3}$$

Our goal is now to derive a partial differential equation for $u(\theta, x)$, and the fact that θ is transcendental guarantees that the identity (3) is an identity in both x and θ, and it is therefore permissible to replace θ with $\mu\theta$ whilst leaving x unchanged, even though θ is itself a function of x. (See further remarks below.) An integration with respect to x now gives

$$\mu u = v_0(\mu\theta, x) + \sum_{i=1}^{n} c_i \log v_i(\mu\theta, x) + C(\mu) \tag{4}$$

where $C(\mu)$ is an arbitrary 'constant'. From (2) we see that

$$u(\mu\theta, x) = v_0(\mu\theta, x) + \sum_{i=1}^{n} c_i \log v_i(\mu\theta, x).$$

and so, comparing this with (4) we find that

$$\mu u(\theta, x) = u(\mu\theta, x) + C(\mu).$$

Finally, differentiation with respect to μ gives, on setting $\mu = 1$,

$$\theta \frac{\partial u(\theta, x)}{\partial \theta} = u(\theta, x) + D$$

where $D = C'(1)$ is a constant. This may be solved in standard fashion to yield

$$u = -D + A(x)\theta$$

and, setting $\theta = \theta_0$, we find that $A(x)$ is determined by $u(\theta_0, x) = A(x)\theta_0 + D$ Thus

$$u = \theta \left(\frac{u(\theta_0, x) - D}{\theta_0} \right) + D$$

and is, indeed, of the form stated.

The key step in the proof concerns the operation of replacing θ by $\mu\theta$ whilst leaving x unchanged. To understand why θ must be transcendental to allow this, consider first the case where $f = g = x$. Now (2) becomes $u = x\theta - \theta$ whilst (3) becomes $f\theta = \theta(x-1) + \theta$. Clearly if θ is now replaced by $\mu\theta$ the identity still holds. Consider, however, what could happen if θ was *not* transcendental. Suppose, for example, we tried to repeat the same argument for the integral of $f\theta$ taking $f = x$ and $\theta = x^2$. (2) becomes $u = x^4/4$, and the equivalent of (3) is now $f\theta = \theta + x(2x)$. A falsehood now results if the variable θ is replaced by $\mu\theta$ whilst x is left unchanged; essentially this has come about because θ and x may be expressed in terms of each other in a non trivial way.

The significance of the rational Liouville theorem is that it may be used to show very quickly why functions such as e^{-x^2} and e^x/x do not possess elementary primitives. For example, if the integral of the function e^{-x^2} was elementary, then by the theorem the primitive would have to take the form

$$\int e^{-x^2} dx = R(x)e^{-x^2}. \tag{5}$$

Suppose that we now set $R(x) = P(x)/Q(x)$ where both P and Q are polynomials (without loss of generality we may assume that $Q \neq 0$ and P and Q have been reduced to their lowest form so that any common factors have been cancelled). Then if (5) is to be true, differentiation shows that

$$e^{-x^2} = \frac{QP' - PQ'}{Q^2} e^{-x^2} - 2x\frac{P}{Q}e^{-x^2}$$

and thus

$$Q(Q - P' + 2xP) = -PQ'. \tag{6}$$

If P and Q possess zeroes, then they cannot be shared (we have assumed that all common factors have been cancelled). Further, if Q possesses a (possibly complex) zero of order n at the point a, so that $Q(x) = (x-a)^n h(x)$ for some polynomial $h(x)$ with $h(a) \neq 0$, then Q' possesses a zero of order $n-1$. This would mean however that the left hand side of (6) has a zero of order at least n, whilst the right hand side has a zero of order $n-1$, a contradiction. Q is therefore a polynomial possessing no zeroes and in consequence must be a constant, Q_0 say. Thus (6) yields

$$P' - 2xP = Q_0. \tag{7}$$

The final step in the argument to prove that the integral of e^{-x^2} is not elementary consists of observing that (7) cannot be satisfied by any polynomial P, for if the highest power in P is $m \geq 0$ say then the highest power in P' will be $m-1$, whilst the highest power in $2xP$ is $m+1$. Evidently the difference between two such polynomials can never be a constant.

Now that the basic method has been established, other results are easy to obtain. For example, suppose that the integral of e^x/x was elementary. Then, proceeding along similar lines to the above argument it would be true that

$$\int \frac{e^x}{x} dx = \frac{P(x)}{Q(x)} e^x.$$

Differentiation and rearrangement gives

$$Q(Q - xP' - Px) = -xPQ'$$

and again it is fruitful to consider the zeroes (if any) of $Q(x)$. First suppose that 0 is not a zero of Q. Then by the same argument used in the previous case we must have $Q = Q_0$ = constant and we are left to determine a polynomial P satisfying

$$xP' + xP = Q_0.$$

Again, by considering the term containing the highest power in x it may quickly be established that this is impossible. The only remaining hope is therefore that 0 is the only zero of $Q(x)$, in which case it must be that $Q(x) = Kx^n$ for some constant K and some $n \geq 1$. Thence

$$Kx^n(Kx^n - xP' - Px) = -xPnKx^{n-1}$$

or

$$(Kx^n - xP' - Px) = -Pn,$$

an immediate contradiction since the left hand side is zero at $x = 0$ but the right hand side is not. The result that e^x/x has no elementary primitive may also be used to obtain other conclusions. For example, a substitution $x = \log u$ shows that the integral of $1/\log u$ cannot be elementary. More generally, a substitution $x = k \log u$ shows that

$$\int \frac{u^{k-1}}{\log u} du$$

is not elementary for any non-zero k. In the exceptional case where $k = 0$, there is of course an elementary primitive, namely $\log \log u$. Further, exploiting the fact that we are working over the field of complex numbers, it is a simple matter to show by an examination of the function e^{ix}/x that the integral of $\sin x/x$ is not elementary. (For details, see Mead (1961)).

As well as providing simple proofs that certain well-known functions have no elementary primitives, arguments similar to those used above may be employed to investigate more complicated integrals. As an example of this, we determine under what (if any) circumstances we may find an elementary expression for

$$I = \int \frac{e^{-x^2}(\alpha x^3 + \beta x^2 + \gamma x + \delta)dx}{(x+1)^2}$$

where α, β, γ and δ are constants (not all zero). Arguing in the normal way that the integral, if elementary, must be of the form $P(x)e^{-x^2}/Q(x)$ where P and Q are polynomials, we find by differentiation that, if this is so, then

$$Q[Q(\alpha x^3 + \beta x^2 + \gamma x + \delta) + 2x(x+1)^2 P - (x+1)^2 P'] = -(x+1)^2 PQ'. \tag{8}$$

Assuming as usual that P and Q have no common factors, consideration of the degrees of the left and right hand sides of the equation at the zeroes of Q shows that either (i) $Q = Q_0$, a constant, or (ii) $Q = A(x+1)^n$ for some $n > 0$. We consider case (i) first: if it is true that

$$Q_0(\alpha x^3 + \beta x^2 + \gamma x + \delta) + 2x(x+1)^2 P - (x+1)^2 P' = 0,$$

then evidently unless P has degree zero so that $P = P_0$ there will be uncancelled terms of degree 4 or greater. Setting $P = P_0$ and collecting terms gives

$$x^3[\alpha Q_0 + 2P_0] + x^2[\beta Q_0 + 4P_0] + x[\gamma Q_0 + 2P_0] + [\delta Q_0] = 0.$$

If $Q_0 = 0$ then $P_0 = 0$ and the problem is the trivial one, so it must be that $\delta = 0$, in which case $\alpha = \gamma$ and $\beta = 2\gamma$ and the integral is given by

$$I = \frac{P_0}{Q_0} e^{-x^2} = \frac{-\gamma}{2} e^{-x^2}.$$

Case (ii) provides some less obvious results, for after setting $Q = A(x+1)^n$ in (8) and cancelling common factors, we find that

$$A(x+1)^{n-1}(\alpha x^3 + \beta x^2 + \gamma x + \delta) + 2x(x+1)P - (x+1)P' + Pn = 0.$$

Now n is an integer greater than zero, but if $n \geq 2$ then setting $x = -1$ in the above expression implies that P has a zero at $x = -1$, an impossibility since P and Q share no factors. Thus $n = 1$ and P is linear, say of the form $P_0 + P_1 x$. Finally, equating coefficients of x gives the equations

$$\begin{aligned} A\alpha + 2P_1 &= 0 \\ A\beta + 2P_0 + 2P_1 &= 0 \\ A\gamma + 2P_0 &= 0 \\ A\delta - P_1 + P_0 &= 0 \end{aligned}$$

which may easily be solved to yield further conditions under which I is elementary, namely that $\beta = \alpha + \gamma$ and $2\delta = \gamma - \alpha$. If these pertain, then the integral is given by

$$I = \frac{-\alpha x - \gamma}{2(x+1)} e^{-x^2},$$

so completing the task of specifying precisely the conditions under which I is elementary.

5 Extensions to Liouville's Theorem and Implications for Computer Algebra

Having seen how Liouville's theorem allows many of the more familiar results concerning closed-form integration to be proved, the question naturally arises as to whether the theorem may be extended. In particular, it would be attractive to augment the set of 'elementary' functions by adding some of the more useful special functions such as the error function, Bessel functions and so on. In general, it turns out that this is not possible (the details are somewhat involved, but this area is one of constant evolution and the interested reader is referred to the interesting discussion in Davenport et al. (1988)) and we have to be content with the class of elementary functions defined above. It is interesting to observe in passing however that another consequence of the theorems given above is that 'special' functions such as $\mathrm{erf}(x)$ defined by

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

really are 'new' functions - however hard we try it is not possible to express them as elementary functions.

Another obvious possible generalization of Liouville's theorem concerns functions that are not themselves elementary, but satisfy equations composed only of elementary functions. For example, the function $w(x)$ defined by the equation

$$e^w - w = x$$

belongs to this class of what are usually termed 'implicitly elementary' functions. In this case, substantial generalizations are possible, though the mathematics involved becomes significantly more complicated. The interested reader is referred, for example, to Risch (1976), who proves a 53 year old conjecture of Ritt to the effect that if the integral of an elementary function is implicitly elementary, then it is elementary. Other generalizations of this sort are also possible.

Finally, it is interesting to conjecture why, after a gap of 130 years following Liouville's work during which, apart from the work of Ritt and Ostrowski (1946), hardly any literature appeared concerning the problem, (Hardy (1905) commented that '[Liouville's memoirs] seem to have fallen into an oblivion which they certainly do not deserve') there was a sudden re-awakening of interest in the 1960's and 1970's. It is no coincidence that during the last 20 years the exponential rise in computing power has made it possible to construct increasingly accomplished symbolic manipulation (computer algebra) packages that can exploit the classical theorems to perform indefinite integration. An *algorithmic* approach to the problem of determining whether an elementary function has an elementary integral was discussed by Risch (1969). In principle, this provides a *complete* solution ot the problem. However, the sophistication of the algorithm, combined with the fact that many of the integrals that computer algebra packages are commonly required to perform are of a fairly simple nature, render the full implementation of the Risch algorithm somewhat unwieldy and in practice the algorithm is usually 'chopped' and combined with some sort of pattern search. The results of such packages are impressive. To give a random example, the computer algebra package MAPLE took less than 1 second CPU time to determine that

$$\int \sqrt{\tan x} dx = \frac{1}{\sqrt{2}} \left(\arctan\left(\frac{\sqrt{2\tan x}}{1 - \tan x} \right) - \log\left(\frac{1 + \tan x + \sqrt{2\tan x}}{\sqrt{1+\tan^2 x}} \right) \right).$$

Acknowledgement

The authors are grateful to Nick Lord and James Vickers for their helpful comments.

References

1. Ritt, J.F. (1948) Integration in Finite Terms, Columbia University Press, New York.

2. Davenport, J.H., Siret, Y. & Tournier, E. (1988) Computer Algebra, Academic Press

3. Liouville, J. (1833) Sur la determination des integrales dont la valeur est algebraique, J. Ecole Polytech. **14** 124-193

4. Rosenlicht, M (1972) Integration in Finite Terms, Amer. Math. Month. **79** 963-972

5. Mead, D.G. (1961) Integration, Amer. Math. Month. **68** 152-156

6. Risch, R.H. (1976) Implicitly Elementary Integrals, Proc. Amer. Math. Soc. **57** 1-7

7. Ostrowski, A. (1946) Sur la integrabilite elementaire de quelques classes d'expressions, Comm. Math. Helv. **18** 283-308

8. Hardy, G.H. (1905) The integration of functions of a single variable, Cambridge University Press

9. Risch, R.H. (1969) The Problem of Integration in Finite Terms, Trans. Amer. Math. Soc. **139** 167-189