Integration to obtain expressions valid on domains of maximum extent

D.J. Jeffrey
Department of Applied Mathematics,
The University of Western Ontario,
London, Ontario, Canada N6A 5B7
djj@uwo.ca

As published in Proceedings of ISSAC 93, M. Bronstein, Ed. ACM Press, pp 34 – 41

Abstract

In certain circumstances, the integration routines used by computer algebra systems return expressions whose domains of validity are unnecessarily restricted by the presence of discontinuities. It is argued that this is undesirable and that integration routines should meet an additional requirement: they should return expressions that are valid on domains of maximum extent. The contention is supported by general mathematical arguments, by an examination of existing practises and by a demonstration that two standard algorithms can be modified to meet the requirement.

1 Introduction

This paper identifies and discusses a general design goal for integration packages, with a view to establishing it as an implementation objective and as a research topic. Several of the difficulties pointed out here are not solved. We introduce the class of problems addressed by discussing two examples that illustrate many general points. First we consider how the following integral is evaluated by two CAS.

$$\int (\cot x + \csc x) dx \ . \qquad (1)$$

The indefinite integration routines of Mathematica and Maple evaluate this integral to the following expressions.

$$2\ln(\sin \tfrac{1}{2}x) \ , \qquad \text{Mathematica;}$$
$$\ln \sin x + \ln(\csc x - \cot x) \ , \qquad \text{Maple.}$$

Suppose, now, that a user were to attempt to evaluate the definite integral

$$\int_{\pi/3}^{3\pi/2} (\cot x + \csc x) dx = \ln 2 \qquad (2)$$

by substituting $3\pi/2$ and $\pi/3$ into one of the above indefinite integrals and then subtracting the results. The Mathematica expression would give the correct answer, while the Maple expression would give $\ln 2 + 2\pi i$, an incorrect result. The error is due to a discontinuity in the Maple expression at $x = \pi$. It is the presence of this discontinuity that we shall be discussing.

As a second example, we define a function $g(x)$ by the integral

$$g(x) = \int_0^x \frac{3}{5 - 4\cos t} dt \ ,$$

and consider the problem of plotting it using Mathematica (version 2.1). Within Mathematica, we can define the function g in two apparently equivalent ways: using numerical integration, and using symbolic integration. The numerical definition is

```
g1[x_ ] := NIntegrate[3/(5-4*Cos[t]),{t,0,x}],
```

while the symbolic one is

```
g2[x_ ] := Integrate[3/(5-4*Cos[t]),{t,0,x}].
```

Upon plotting these functions, however, we see that they are not the same, because the function `g1` is continuous everywhere, while `g2` is discontinuous at odd multiples of π. Other CAS show similar behaviour. Specifically, the following indefinite integral evaluates as shown, using the CAS indicated.

$$\int \frac{3}{5-4\cos x}dx \Rightarrow \tag{3}$$

$$\begin{aligned} &2\arctan(3\tan x/2)\ ; && \text{Maple, Mathematica} \\ &2\arctan(3\sin x/(\cos x+1))\ ; && \text{Macsyma} \\ &-\arctan(-3\sin x/(5\cos x-4))\ ; && \text{Axiom.} \end{aligned}$$

None of these expressions is continuous everywhere.

The two examples show that CAS might return different expressions for an integral that differ from each other with respect to their continuity properties, and in addition those properties might be different from the ones expected by the user. To explore this further, and to define some notation, we recall the fundamental theorem of calculus (Rudin 1976). If f is continuous on an interval $[a, b]$, the function defined by

$$g(x) = \int_a^x f(t)dt \tag{4}$$

is continuous and differentiable on (a, b), and

$$g'(x) = f(x)\ . \tag{5}$$

The interval $[a, b]$ is an essential part of this theorem, and yet indefinite integration is mostly performed by CAS without reference to any interval. Moreover, present users do not expect to have to specify an interval when posing an indefinite integration problem to a CAS. When the system returns an expression as an answer to an integration problem, it may or may not apply to the interval envisioned by the user. Thus the Maple response to the first example is correct on the interval $(0, \pi)$ and only incorrect if one accepts that the interval should be $(\pi/3, 3\pi/2)$.

Standard terminology often blurs the connection between the function defined by the fundamental theorem of calculus and any apparently similar function defined by an indefinite integral. In addition, because the integral g has a derivative equal to the integrand f, the term anti-derivative of f is frequently used for g, and this leads to the common assumption that the problem of integrating f is solved by finding any function whose derivative is equal to f. The shortcomings of this assumption are the focus of this paper. To discuss them, we introduce the following definitions.

Definition. An anti-derivative of a function $f(x)$ is any function $F(x)$ that satisfies $F'(x) = f(x)$.

Comment. This definition requires further qualification on the meaning of differentiation, and is not independent of the CAS under consideration. We take the differentiation to be the differentiation used by the CAS. Fateman (1992) pointed out that an anti-derivative of $f = 0$ is $F = \arctan x + \arctan(1/x)$ in most systems. In contrast, systems differ significantly over the derivative of the signum function. `DIF( SIGN(x),x)` is evaluated as 0 by Derive, but Maple and Mathematica return a formal derivative of signum.

Definition. An indefinite integral of a function $f(x)$ on an interval $[a, b]$ is any function $G(x)$ that satisfies $G(x) = \int_a^x f(t)dt + K$, where K is constant on the interval. Since most CAS are willing to consider functions f that are integrable but not continuous on $[a, b]$, we shall suppose that the integral is interpreted as a Lebegue integral (Rudin 1976) or at least as the generalized Riemann integral defined by Botsko (1991). So f need only be Riemann integrable rather than continuous, and G is differentiable almost everywhere on (a, b) rather than everywhere.

In terms of these definitions, the central question that this paper raises is whether it is sufficient for an integration routine to return an anti-derivative of the function passed to it, or whether it should strive to return an indefinite integral, and if so, on what interval. With regard to the question of what interval, we introduce the following definitions.

Definition. Given a function $f(x)$ that is integrable on an interval $[a, b]$, and an anti-derivative F of f that has a discontinuity somewhere on $[a, b]$, we shall call the discontinuity in F spurious, because the fundamental theorem of calculus asserts that another function g exists that does not contain this discontinuity.

Definition. Given a function f that is integrable on one or more intervals of the real line, a function g will be called an integral on the domain of maximum extent if it can serve as an indefinite integral of f on all of the intervals on which f is integrable.

As an example, the expression $\arctan(\tan x)$ is an integral of $f = 1$ on the interval $(-\pi/2, \pi/2)$, but the expression x is the integral on the domain of maximum extent, namely the real line. Both x and $\arctan(\tan x)$ are antiderivatives of 1, but $\arctan(\tan x)$ contains spurious discontinuities at odd half multiples of π.

This paper contends that routines should return indefinite integrals valid on domains of maximum extent, and that it is important to develop and implement such routines. The first argument in favour of this is the expectation held by most users. A user who wishes to derive the result

$$\int_0^{4\pi} \frac{3}{5-4\cos t}dt = g(4\pi) - g(0) = 4\pi\ , \tag{6}$$

expects to be able to do so using the indefinite integral returned by the system, because it is obvious that the integral exists everywhere. Designers of CAS might reply that if the above definite integral is posed *as a definite integral*, then Maple V release 2, Macsyma and Mathematica get the correct answer. Clearly these systems are using something more than their own anti-derivatives to evaluate the definite integral, and they make it the users' responsibility to know this. Users must convert the statement of their problem into a particular form in order to get the correct answer. Surely users would prefer an indefinite integral that made this unnecessary.

Supporters of separate definite-integration routines might argue that there is no point in correcting the indefinite integrator because a good definite integrator must exist separately anyway. The system must check not only for cases such as (6), for which a finite solution exists, but also for integrals across singularities, where the integral may or may not be finite. Thus Mathematica returns -2 for the integral

$$\int_{-1}^{1} \frac{dx}{x^2} ,$$

and this error can only be corrected inside a definite integration routine. This indeed shows the need for a definite-integration routine, but it in no way lessens the desirability of correcting the indefinite integrator, because a continuous indefinite integral will reduce the load on the definite integration code.

As a second argument, let us recall that indefinite integration is used for more things than definite integration. The solution of differential equations usually requires repeated integration. For an example, the reader is referred to Marion & Thornton (1988, section 7.53) where the equations of planetary motion are incorrectly integrated because the authors used expressions derived from those in (3).

Another argument can be based on the emerging practises of CAS. Several systems have recently modified their indefinite integration routines to replace discontinuous expressions with continuous ones. Derive has consistently adopted the policies advocated here for several years, and both Maple and Axiom have implemented Rioboo's algorithm (Rioboo 1991). Release 1 of Maple V, for example, evaluated the integral below to a discontinuous expression. In release 2, however, the result is a continuous expression, showing that systems have already changed some routines in order to give preference to a continuous integral over a discontinuous one.

$$\begin{aligned}\int \frac{x^4-3x^2+6}{x^6-5x^4+5x^2+4}dx \\ \Rightarrow (i/2)\ln(x^3+ix^2-3x-2i) \\ -(i/2)\ln(x^3-ix^2-3x+2i) \ , \quad \text{release 1}\end{aligned}$$

$$\begin{aligned}\Rightarrow \arctan(-3x^3/2+x^5/2+x/2)+\arctan x \\ +\arctan(x^3) \ , \qquad \text{release 2}\end{aligned}$$

The first expression is discontinuous at $x=\sqrt{2}$, and hence is an indefinite integral only on the intervals $(-\infty,\sqrt{2})$ and $(\sqrt{2},\infty)$. In contrast, the second expression is an indefinite integral on the domain of maximum extent. Some comments on the Rioboo algorithm will be made later.

There is only one example, that I am aware of, in which a system has preferred a discontinuous anti-derivative to a continuous one. At some point, Reduce changed to an expression equivalent to (3) from one that was continuous, although the developers have commented (private communication) that this change did not imply a conscious policy.

2 The rectification of anti-derivatives

If, for a particular integrand, an integration method gives an anti-derivative containing spurious discontinuities, there are two strategies that might be used to obtain an integral on the domain of maximum extent. The first way is to change the integration method to one that gives the desired integral, and the other is to accept the anti-derivative and look for a way of removing the discontinuity. The latter approach regards any expression returned by an integration procedure as provisional until it has been checked somehow for discontinuities. We can call this approach the rectification of anti-derivatives; it has been used in several contexts. The algorithm of Rioboo (1991) is such a rectifying algorithm. Here we give another one.

The problem can be illustrated by Maple's answer to the first example:

$$\int \cot x + \csc x dx = \ln\sin x + \ln(\csc x - \cot x) \ .$$

The discontinuity at π is the result of two effects. First, the formula

$$\int [f'(x)/f(x)]dx = \ln f(x)$$

contains a discontinuity, in that the imaginary part of the logarithm is discontinuous whenever f passes through 0. Usually this is of no consequence, because it is masked by the singularity in the real part of the logarithm. In the case above, however, the singularities in the real parts of the two logarithms cancel, but the discontinuities in the imaginary parts do not. In the neighbourhood of $x = \pi$, the anti-derivative behaves like

$$\ln \sin x + \ln(\csc x - \cot x) \sim \ln(\pi - x) + \ln \frac{2}{\pi - x} \ .$$

To leading order, the right-hand side is asymptotically equal to $\ln 2 + 2\pi i \operatorname{sgn}(x-\pi)$.

In general, given two functions f_1 and f_2 that behave near a point x_s asymptotically according to

$$f_1 \sim (x - x_s)^n \quad \text{and} \quad f_2 \sim (x - x_s)^{-n} \ ,$$

then

$$\frac{f_1'}{f_1} + \frac{f_2'}{f_2}$$

will be integrable, but the expression

$$\ln f_1 + \ln f_2$$

will be discontinuous.

To remove this behaviour, we note that although the transformation

$$\ln f_1(x) + \ln f_2(x) \Rightarrow \ln[f_1(x) f_2(x)]$$

is usually valid only if f_1 and f_2 are real and at least one of them is positive, for anti-derivatives it is always permissible. This is because

$$(\ln f_1 + \ln f_2)' = f_1'/f_1 + f_2'/f_2 = (\ln[f_1 f_2])' \ ,$$

and hence both expressions are anti-derivatives of the same function. We are therefore free to choose either form to express an anti-derivative, or integral. We can see that the collected form $\ln(f_1 f_2)$ is always preferable by the following argument. If x_s is a singular point of the expression $\ln(f_1 f_2)$, then it will also be a singular point of $\ln f_1 + \ln f_2$, but the converse is not true. There can be singular points of $\ln f_1 + \ln f_2$ that are not singular points of $\ln(f_1 f_2)$. The proof is an obvious generalization of the example above, and will not be written out.

Applied to the example above, the transformation gives

$$\ln \sin x + \ln(\csc x - \cot x) \Rightarrow \ln(1 - \cos x) \ ,$$

which is equivalent to the expression returned by Mathematica, and equal to the one returned by Axiom.

A more general transformation is needed to handle some cases, such as Axiom's evaluation of the following integral.

$$\int \left(\frac{\pi - 2x}{x(\pi - x)} - \csc x \right) dx \Rightarrow$$
$$\tfrac{1}{2} \ln(\cos x + 1) - \tfrac{1}{2} \ln(\cos x - 1) + \ln(x^2 - \pi x) \ ,$$

which has a spurious discontinuity at 0. The transformation needed is for

$$\alpha \ln f_1 + \beta \ln f_2 \ ,$$

where α and β are rational coefficients. The obvious transformation to $\ln(f_1^\alpha f_2^\beta)$ is unsatisfactory because fractional powers will again introduce discontinuities. Instead we find integers m,n,p,q such that $\alpha = mp/n$ and $\beta = mq/n$ and p and q are mutually prime. Then the transformation is

$$\alpha \ln f_1 + \beta \ln f_2 \Rightarrow \frac{m}{n} \ln(f_1^p f_2^q) \ .$$

Applied to the above, it gives

$$\tfrac{1}{2} \ln(\cos x + 1) - \tfrac{1}{2} \ln(\cos x - 1) + \ln(x^2 - \pi x)$$
$$\Rightarrow \tfrac{1}{2} \ln \left[\frac{\cos x + 1}{\cos x - 1} (x^2 - \pi x)^2 \right] \ .$$

Once an integral has been reduced to a single logarithmic term, a 'tidy-up' transformation $\ln K f^\gamma \Rightarrow \gamma \ln f$, for K,γ constants, can be used, justified as above by differentiating both forms. This turns the last result into

$$\tfrac{1}{2} \ln \left[\frac{\cos x + 1}{\cos x - 1} (x^2 - \pi x)^2 \right] \Rightarrow \ln[(x^2 - \pi x) \cot \tfrac{1}{2} x] \ .$$

For the two examples above, Mathematica obtains integrals on domains of maximum extent without using this approach, and it might seem that Mathematica's methods are better than these. However, Mathematica fails to obtain a continuous integral for

$$\frac{\sec^2 x}{1 + \tan x + \sqrt{1 + \tan x}} \ .$$

It gives $2 \operatorname{arctanh} \sqrt{1 + \tan x} + \ln \tan x$ which has spurious discontinuities at integer multiples of π, whereas $2 \ln(1 + \sqrt{1 + \tan x})$ does not.

3 Discontinuity from substitution

The technique of integration by substitution is a standard topic in calculus textbooks and is one that is used by some integration routines, in particular, by those in Derive. An aspect of the technique that is rarely emphasised is the fact that spurious discontinuities can be introduced if a substitution is singular in the domain of the integrand.

Derive uses substitution to evaluate the following integral. Given the integrand

$$f(x) = \begin{cases} \dfrac{\exp(1/x)}{x^2(1 + \exp(1/x))^2} & x \neq 0, \\ 0 & x = 0, \end{cases}$$

which satisfies the conditions of the fundamental theorem, we make the substitution $s = 1/x$ to get

$$\int \frac{e^{1/x}}{(1+e^{1/x})^2} \frac{dx}{x^2} = \int \frac{e^s\, ds}{(1+e^s)^2} \ .$$

Integrating the last expression and substituting for s in the usual way, we obtain

$$\frac{1}{1+e^{1/x}} \ ,$$

which contains a spurious discontinuity at $x = 0$. In order to develop an algorithm that will remove this discontinuity, we must decide first where it comes from. The method of integration will influence, to some extent, how we assign the cause. Since our focus here is the method of substitution, we turn to the following theorem (Jeffrey & Rich 1993)

Theorem. Given a function f that is continuous on an interval $[a, b]$ and a function ϕ that is differentiable and monotonic on the interval $[\phi^{-1}(a), \phi^{-1}(b)]$, the function

$$g(x) = \int_{\phi^{-1}(a)}^{\phi^{-1}(x)} f(\phi(t))\phi'(t)\ dt = \int_a^x f(s)\ ds$$

is continuous for $x \in [a, b]$.

The relevance of this theorem comes about as follows. We suppose that we wish to obtain an integral of $f(x)$ that is valid on the domain of maximum extent. We further suppose that our integration system can already return an indefinite integral on a domain of maximum extent for the function $f(\phi(t))\phi'(t)$. The theorem states that the second integral might be discontinuous at points where ϕ is singular. For our example, it follows that the point $x = 0$ might be (and is) a point of discontinuity. We could at this point jump immediately to introducing a rectifying transformation, but we can place it on a more formal footing as follows.

Suppose f is integrable on an interval $[a, c]$, and ϕ is differentiable and monotonically increasing at all points in $[a, c]$ except the isolated point b. On such an interval, the function we wish to find is g, defined by

$$g(x) = \int_a^x f(s)\ ds \ ,$$

and it will be continuous by the fundamental theorem; the function we actually find, however, is $\hat{g}$.

$$\hat{g}(x) = \int f(\phi(x))\phi'(x)\ dx \ .$$

So long as $x < b$, the function g can be expressed, by the above theorem, in terms of $\hat{g}$.

$$g(x) = \hat{g}(x) - \hat{g}(a) \ .$$

For $x > b$, the connection between g and $\hat{g}$ is obtained as follows.

$$\begin{aligned} g(x) &= \int_a^c f(s)\ ds - \int_x^c f(s)\ ds \\ &= \int_a^c f(s)\ ds - \int_{\phi^{-1}(x)}^{\phi^{-1}(c)} f(\phi(t))\phi'(t)\ dt \\ &= g(c) - \hat{g}(c) + \hat{g}(x) \ . \end{aligned}$$

To eliminate $g(c)$ from this equation, we calculate

$$\begin{aligned} &\lim_{x\to b-} \hat{g}(x) - \lim_{x\to b+} \hat{g}(x) \\ &= \lim_{x\to b-} (g(x) + \hat{g}(a)) - \lim_{x\to b+} (g(x) - g(c) + \hat{g}(c)) \\ &= g(c) - \hat{g}(c) + \hat{g}(a) \ , \end{aligned}$$

where the limits have been evaluated using the continuity of g at b. We can combine the expressions for g in the intervals $[a, b]$ and $[b, c]$ into a single equation using the Heaviside, or step, function H.

$$g(x) = \hat{g}(x) - \hat{g}(a) + H(x-b)\left[\lim_{x\to b-} \hat{g}(x) - \lim_{x\to b+} \hat{g}(x)\right] \ .$$

This gives a continuous expression for the desired function g in terms of the computable function $\hat{g}$.

Applying this theory to our example, we see that an indefinite integral valid on the real line, the domain of maximum extent, is

$$\int \frac{e^{1/x}}{(1+e^{1/x})^2} \frac{dx}{x^2} = \frac{1}{1+e^{1/x}} + H(x) \ .$$

There remains a removable singularity at 0 in this expression for the integral, so the user would have to use a limit to evaluate a definite integral for which 0 was an endpoint; CAS definite-integation routines would mostly do this automatically. Other ways of removing the discontinuity are possible: Derive adds $\frac{1}{2}\operatorname{sgn} x$ instead of $H(x)$.

The example used in this section is tackled by many systems using Risch integration instead of substitution. The algorithm given here could not be used by such systems, and the development of procedures for obtaining continuous expressions from that method of integrating is an important topic. In this connection, the 'cause' of the discontinuity must be decided, since clearly we can no longer blame the substitution $\phi = 1/x$.

A limitation of the present method is the fact that the singularities of ϕ must be ascertainable by the system. Although this will be possible for most functions arising in practise, it is in general an outstanding problem in its own right.

4 Complex-valued integrands

We now consider some problems associated with the integration of a complex-valued function $f(x)$ of a real variable x. The emphasis here will be less on establishing algorithms, and more on discussing the problems that exist and ways of tackling them. There are two reasons for considering integration problems of this type. First, because of the way in which many algorithms work, an integration problem posed by a user entirely in real terms might be evaluated symbolically by converting the integral into one taking complex values. Second, a contour integration in the complex plane is typically converted into a complex integral over a real variable by describing the contour of integration parametrically. In this latter case, many textbooks of complex analysis give the impression that all such integrals can be reduced to real integrals. Thus they write that the contour integral

$$\int_C f(z)dz$$

can be evaluated by describing the contour parametrically as $z = \phi(s)$ and then separating the integrand into real and imaginary parts $f(z) = f(\phi(s)) = u(s) + iv(s)$. In practice, however, this last step is purely formal. In many cases the decomposition into $u + iv$ is impractical and, even if it were successful, would only lead to real integrals too difficult for the CAS to evaluate. For example, the integral

$$\int [1 + (1 + i)x]^{1/2}\, dx = \frac{2}{3(1+i)}[1 + (1 + i)x]^{3/2}$$

is simple as a complex function, but the separation into Cartesian form makes it too difficult for Maple or Derive (although not Axiom):

$$\begin{aligned} &[1 + (1 + i)x]^{1/2} = \\ &\quad \frac{1}{\sqrt{2}}\sqrt{\sqrt{2x^2+2x+1}+x+1} \\ &\quad +\frac{i \operatorname{sgn} x}{\sqrt{2}}\sqrt{\sqrt{2x^2+2x+1}-x-1}\ . \end{aligned}$$

The main aim of this section is to discuss some unsatisfactory aspects of the formula

$$\int \frac{f'(s)}{f(s)} ds = \log f(s)\ , \qquad (7)$$

and to discuss ways of improving it. To focus the discussion, we consider a specific example. Different CAS give different expressions for the following integral, two such expressions being

$$\int \frac{a\, dx}{a + \exp(ix)} \quad \Rightarrow \quad i \ln(ae^{-ix} + 1)\ , \qquad (8)$$

$$\Rightarrow \quad i \ln(a + e^{ix}) + x\ . \qquad (9)$$

This is just the contour integral of $1/z$ around the path $z = 1 + a\exp(-ix)$, which is a circle of radius a with centre at $z = 1$. If x increases by an amount greater than 2π, and $a > 1$, the circle encloses the origin, but if $0 < a < 1$ it does not. Using residues, we see that

$$\int_0^{2\pi} \frac{dx}{a + \exp(ix)} = \begin{cases} 0 & \text{for } 0 < a < 1, \\ 2\pi\ , & a > 1. \end{cases}$$

If we wish to evaluate the same integral using the antiderivatives above, we see that we must use (8) for the case $0 < a < 1$ and (9) for $a > 1$.

Expression (8) is an application of (7), but we see that it would give a wrong answer for $a > 1$. There are two general approaches one might take to correcting this error. The first way is to reinterpret (7) by saying that the logarithm appearing there is a multivalued logarithm, not the principal-branch logarithm. This is only a formal solution of the problem, however, and from the point of view of evaluation (especially numerical evaluation) it merely postpones having to face the problem. To evaluate a definite integral, one must know on which branch of the logarithm each endpoint lies. The escape to multivalued logarithms is often seen in textbooks of complex analysis, where it is feasible because definite integrals are never actually evaluated; for a CAS, however, it is not a possibility, at least not if the system hopes to transcend the purely formal success of a textbook. The alternative is to fix the logarithm as the principal branch and replace (7) with an expression valid on a wider domain. The logarithm will be discontinuous whenever $f(s)$, regarded as a contour, crosses the negative real axis, at which time $\Im(f) = 0$. We shall put aside the difficulties of solving this equation in practice. To ascertain whether the contour has crossed the negative real axis, we can use a limiting procedure. Let $\{\alpha_n, n = 1, 2, ...\}$ be the roots of $\Im(f) = 0$. We evaluate

$$K_n = \lim_{s\to\alpha_n-} \ln f(s) - \lim_{s\to\alpha_n+} \ln f(s)\ .$$

If this is non-zero, then there was a crossing. Using arguments similar to those in the last section, we can write a corrected integral in the form

$$\int \frac{f'(s)}{f(s)} ds = \log f(s) + \sum_n K_n H(s - \alpha_n)\ . \qquad (10)$$

This last equation is a purely formal replacement for (7) until the computability of the quantities K_n and α_n is known. Also the expressions produced by this last equation are unlikely to be pleasing, unless the summation can be simplified to something attractive. For example, let us apply this to the integral just given. The solution of

$$\Im(ae^{-ix} + 1) = 0$$

for real a is $x = n\pi$. Hence

$$\int \frac{a\,dx}{a+\exp(ix)} = i\ln(ae^{-ix}+1) + \sum_n K_n H(x-n\pi)\ .$$

If $|a|<1$, then $K_n = 0$ for all n, as several systems can obtain correctly. If $a>1$, then

$$K_n = \begin{cases} 2\pi\ , & n \text{ odd,} \\ 0\ , & n \text{ even.} \end{cases}$$

If $a<-1$, then

$$K_n = \begin{cases} -2\pi\ , & n \text{ even,} \\ 0\ , & n \text{ odd.} \end{cases}$$

Thus, using also the result that

$$\sum_{n=1}^{\infty} H(x-np-q) = \left\lfloor \frac{x-q}{p} \right\rfloor\ ,$$

where $\lfloor\ \rfloor$ is the floor function, we obtain

$$\int \frac{a\,dx}{a+\exp(ix)} =$$

$$\begin{array}{rl} i\ln(ae^{-ix}+1)\ , & |a|<1\ , \\ i\ln(ae^{-ix}+1) + 2\pi\lfloor (x-\pi)/2\pi \rfloor\ , & a>1\ . \\ i\ln(ae^{-ix}+1) - 2\pi\lfloor x/2\pi \rfloor\ , & a<-1\ . \end{array}$$

Further, since for $a>1$,

$$i\ln(ae^{-ix}+1) + 2\pi\lfloor (x-\pi)/2\pi \rfloor = x + i\ln(a+e^{ix})\ ,$$

with a similar identity when $a<-1$, we obtain

$$\int \frac{a\,dx}{a+\exp(ix)} = \begin{cases} i\ln(ae^{-ix}+1)\ , & |a|<1, \\ i\ln(a+e^{ix}) + x\ , & a>1, \\ i\ln(-a-e^{ix}) + x\ , & a<-1. \end{cases}$$

The conclusion to be drawn from this exercise is that (10) does indeed improve on (7) as an integration formula, but the implementation of (10) would entail significant computation; and the range of problems for which it could be completed successfully is an open question.

5 Other sources of discontinuity

In the previous sections, the discontinuities considered arose because of the way in which the integral was evaluated, and the behaviour of the integrand itself did not contribute to the difficulties. There are integrals, however, in which the behaviour of the integrand must be understood in order to obtain a correct indefinite integral. Consider the integral

$$\int_{-3/2}^{-1/2} \sqrt{x^{2/3}+x^{4/3}}\,dx \approx 0.0712 - 0.1261i\ .$$

The numerical evaluation used principal branch definitions for all quantities. In this example, it is now the *integrand* that crosses the branch cut along the negative real axis. It does so at $x=-1$. The integral is still well-defined, and the result of numerical evaluation is shown. An anti-derivative for this integral is known. It is

$$\frac{1}{5}\,\mathrm{sgn}^{5/3}(x)(1+x^{2/3})(3x^{2/3}-2)\sqrt{(1+x^{2/3})\,\mathrm{sgn}^{2/3}(x)}\ .$$

This expression is discontinuous at $x=-1$ because of the discontinuity in the integrand. As with the previous cases, once the location of a discontinuity is established, there is no difficulty in removing it. The point of the example, however, is that the location of the discontinuity requires a more elaborate analysis of the behaviour of the integrand in the complex plane than any current system could manage automatically.

6 Concluding remarks

Fateman (1992) has remarked that the Risch algorithm is commonly misunderstood. He distinguishes between anti-derivatives and indefinite integrals in a way similar to the way it was done here, and points out that the Risch algorithm only guarantees to return an anti-derivative of a function, not an indefinite integral. In particular there is no reason to suppose that the anti-derivative of 0 will be a constant; it might be a function such as $\arctan x + \arctan(1/x)$ which is piecewise constant only. This paper has offered a number of arguments which complement Fateman's remarks, in the hope of establishing that the goal of integration routines should be to return indefinite integrals. Most of the time they already do, of course, which is why the issue has never been considered in detail. Exceptions, such as the examples presented here, have been treated off-handedly.

A second purpose of this paper has been to establish that the goal of returning indefinite integrals can be embraced by CAS without degrading their functionality; rather they improve it. Two common sources of discontinuous anti-derivatives have been examined, namely the handling of logarithms of real arguments and the method of substitution, and the modest improvements that correct their results in many practically occurring cases have been presented. Several other sources of discontinuous behaviour have been pointed out without a well-defined correction being given. It is to be hoped that such areas will be the subjects of further research. There will be times when the application of the improved algorithms to particular cases will involve substantial computation, but that is no reason to postpone implementing them: a wrong answer is still wrong no matter how quickly and efficiently it is obtained. A final comment concerns the interaction between the standard textbooks and CAS. The fact that equation (10) is never seen in a book on complex analysis does not mean that it does not have to be taken into account. These books are written assuming a high level of abstraction and generality, and this is a luxury the CAS cannot share, if they hope to return correct and explicit results to their users.

7 References

Botsko, M.W. 1991 A fundamental theorem of calculus that applies to all Riemann integrable functions, *Mathematics Magazine*, **64**, 347–348.

Fateman, R.J. 1992 *Risch algorithm.* Posted on newsnet sci.math.symbolic 26 Sept 1992.

Geddes, K.O., Czapor, S.R. and Labahn, G. 1992 *Algorithms for computer algebra.* Kluwer Academic.

Jeffrey, D.J. and Rich, A.D. 1993 The evaluation of trigonometric integrals avoiding spurious discontinuities. *Trans. Math. Software*, to appear.

Marion, J.B. and Thornton, S.T. 1988 *Classical Dynamics, 3rd Ed.* Harcourt, Brace, Jovanovich.

Rioboo, R. 1991 *Quelques aspects du calcul exact avec les nombres reels.* These de doctorat, universite Paris 6.

Rudin, W. 1976 *Principles of mathematical analysis.* McGraw-Hill.