of π. Other CAS show similar behaviour. Specifically, the following indefinite integral evaluates as shown, using the CAS indicated.

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

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

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

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

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 $\ln 2 + 2\pi i \,\mathrm{sgn}(x - \pi)$.

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 $2 \,\mathrm{arctanh} \sqrt{1 + \tan x} + \ln \tan x$ 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 emwhich 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 b

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

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

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 he expres [illegible] 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 eval-
i etvi and le on

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* [illegible] *s*