ering (to non-integral powers) and also the n-th root. These functions are built in, to a greater or lesser extent, into many computer algebra systems (not to mention other programming languages [12, 17]), and are heavily used. As abstract algebraic solutions to integrals and/or differential equations, they satisfy well-known properties [16]. However, reasoning with them as functions $\mathbb{C} \to \mathbb{C}$ is more difficult than is usually acknowledged, and all algebra systems have one, sometimes both, of the following defects:

- they make mistakes, be it the traditional schoolchild one

$$1 = \sqrt{1} = \sqrt{(-1)^2} = -1 \tag{1}$$

  or more subtle ones;

- they fail to perform obvious simplifications, leaving the user with an impossible mess when there "ought" to be a simpler answer. In fact, there are two possibilities here: maybe there is a simpler equivalent that the system has failed to find, but maybe there isn't, and the simplification that the user wants is not actually valid, or is only valid outside an exceptional set. In general, the user is informed neither what the simplification might have been nor what the exceptional set is.

Faced with these problems, the user of the algebra system is not convinced that the result is correct, or that the algebra system in use understands the functions with which it is reasoning. An ideal algebra system would never generate incorrect results, and would simplify the results as much as practicable, even though perfect simplification is impossible, and not even totally well-defined: is $1 + x + \cdots + x^{1000}$ "simpler" than $(x^{1001} - 1)/(x - 1)$?

Throughout this paper, z and its decorations indicate a complex variable, while x, y and t indicate real variables. The symbol $\Im$ denotes the imaginary part, and $\Re$ the real part, of a complex number. For the purposes of this paper, the precise definitions of the inverse elementary functions in terms of log are those of [5]: these are reproduced in Appendix A for ease of reference.

2 The Problem

The fundamental problem is that log is multi-valued: since $\exp(2\pi i) = 1$, its inverse is only valid up to adding any multiple of $2\pi i$. This ambiguity is traditionally resolved by making a *branch cut*: usually [1, p. 67] the branch cut $(-\infty, 0]$, and the rule (4.1.2) that

$$-\pi < \Im \log z \leq \pi. \tag{2}$$

This then completely specifies the behaviour of log: on the branch cut it is continuous with the positive imaginary side of the cut, i.e. counter-clockwise continuous in the sense of [14].

1. $z = \ln e^z + 2\pi i \mathcal{K}(z)$.
2. $\mathcal{K}(a \ln z) = 0 \, \forall z \in \mathbb{C}$ if and only if $-1 < a \leq 1$.
3. $\ln z_1 + \ln z_2 = \ln(z_1 z_2) + 2\pi i \mathcal{K}(\ln z_1 + \ln z_2)$.
4. $a \ln z = \ln z^a + 2\pi i \mathcal{K}(a \ln z)$.
5. $z^{ab} = (z^a)^b e^{2\pi i b \mathcal{K}(a \ln z)}$.

5′. $e^{ab} = (e^a)^b \; (-\pi < \Im a \leq \pi)$.

Table 1: Some correct identities for logarithms and powers using $\mathcal{K}$.

2.2 Uniformly valid transformations

The authors of [6] point out that most "equalities" do not hold for the complex logarithm, e.g. $\log(z^2) \neq 2 \log z$ (try $z = -1$), and its generalisation

$$\log(z_1 z_2) \neq \log(z_1) + \log(z_2). \tag{7}$$

The most fundamental of all non-equalities is $z \stackrel{?}{=} \log \exp z$, whose most obvious violation is at $z = 2\pi i$. (A similar point w

Similarly (4) can be rescued as

$$\log \overline{z} = \overline{\log z} - 2\pi i \mathcal{K}(\overline{\log z}). \tag{10}$$

Note that, as part of the algebra of $\mathcal{K}$, $\mathcal{K}(\overline{\log z}) = \mathcal{K}(-\log z) \neq \mathcal{K}(\log \frac{1}{z})$. $\mathcal{K}(z)$ depends only on the imaginary part of z. $\mathcal{K}(\log z) = 0$ for all z.

$$\mathcal{K}(-\log z) = \begin{cases} -1 & z \text{ real negative} \\ 0 & \text{elsewhere} \end{cases}. \tag{11}$$

2.3 Multi-valued "functions"

Although not formally proposed in the same way in the computational community, one possible study elementary functions (they have been used successfully for algebraic functions [10]). In order to answer this, a computational interpretation of a Riemann surface must first be given, because the standard textbook descriptions do not offer one.

The essence of the Riemann surface approach is geometric. We associate paths with each quantity that we compute with. Consider the pair of functions $z = e^w$ and $w = \ln z$ as an example. The approaches above have concentrated on the values taken by w, or in other words on the behaviour of $\ln z$ in the w-plane; in the Riemann-surface treatment, attention is shifted to the z-plane, by considering the mapping $z = e^w$. Starting from a path, say a straight vertical segment, joining two points w_1 and $w_2 = w_1 + 2\pi i$, we consider the path it maps to in the z plane. Let $z_1 = e^{w_1}$ and $z_2 = e^{w_2} = e^{w_1+2\pi i}$ be the endpoints of the path; usually these are considered to be the same point, but in the Riemann approach, we continue to distinguish between them. In order to create a distinction, we label the two z-points with the property that makes them different, namely the value of the imaginary part of w. In [18, 7], a 3-dimensional surface is constructed by plotting $\Im w$ against complex z. This can be shown to be isomorphic to any faithful representation of the Riemann surface for $\ln z$.

Therefore, in the Riemann approach, the complex number z now becomes a number and an index: z_{w_1} and z_{w_2}. Using the index, we can decide which value to assign $\ln z$. Thus $w_1 = \ln(z_{w_1})$ and $w_2 = \ln(z_{w_2})$. This is effectively how students in a first course on complex numbers compute the n values of $z^{1/n}$. They are taught to start with the equivalence $z = re^{i\theta} = re^{i\theta+2\pi ik}$ and then they are mysteriously ordered to apply the rule $(e^{i\phi})^{1/n} = e^{i\phi/n}$ to the second form of z rather than the first. Thus they are replacing z with z_k, an equivalent point on the kth Riemann sheet, and then computing $(z_k)^{1/n}$.

Taking the point of view of a computer algebra system, we see that a complex number z does not reveal its full significance until we know what Riemann sheet it is on. Thus, to compute logs correctly in the Riemann approach, a system would have to create a new data structure, so that the correct sheet of the Riemann surface could be recorded along with the value of z. This would have to be equivalent to a path from some reference point. This does not complete the solution to the problem, however, because the Riemann surface depends upon the function being considered: the Riemann sheet for log being different from that for $\arcsin z$, for instance.

Further problems arise when we consider combining functions. Thus, when we write $\ln(u + v)$, are u and v on the same sheet? More importantly, what sheet is the sum $u + v$ on? Products are easier: we can write $u = re^{i\theta}$ and record its sheet as a multiple of 2π in θ, and likewise for v; then the index of the sheet of the product uv is just the sum of the indexes of the sheets of u and v. But sums are awkward. Moreover, if we write the expression $(z - 1)^{1/2} + \ln z$, the Riemann surface for the combined function is different from either of the component Riemann surfaces. How do we label z?

We have to distinguish between a conceptual scheme and a computational scheme. Computer systems are about computation. Often computation assists in conception, but computers must be able to compute. Riemann surfaces are a beautiful conceptual scheme, but at the moment they are not computational schemes.

3 The rôle of the Unwinding Number

We claim that the unwinding number provides a convenient formalism for reasoning about these problems. Inserting the unwinding number systematically allows one to make "simplifying" transformations that *are* mathematically valid. The unwinding number can be evaluated at any point, either symbolically or via guaranteed arithmetic: since we know it is an integer, in practice little accuracy is necessary. Conversely, removing unwinding numbers lets us genuinely "simplify" a result. We describe insertion and removal as separate steps, but in practice every unwinding number, once inserted by a "simplification" rule, should be eliminated as soon as possible. We have thus defined a concrete goal for mathematically valid simplification.[5]

The following section gives examples of reasoning with unwinding numbers. Having motivated the use of unwinding numbers, the subsequent sections deal with their insertion (to preserve correctness) and their elimination (to simplify results).

4 Examples of Unwinding Numbers

This section gives certain examples of the use of unwinding numbers. We should emphasise our view that an ideal computer algebra system should do this manipulation for the user: certainly inserting the unwinding numbers where necessary, and preferably also removing/simplifying them where it can.

4.1 Forms of arccos

The following example is taken from [5], showing that two alternative definitions of arccos are in fact equal:

Theorem 1

$$\frac{2}{i}\ln\left(\sqrt{\frac{1+z}{2}}+i\sqrt{\frac{1-z}{2}}\right) = -i\ln\left(z+i\sqrt{1-z^2}\right). \tag{13}$$

First we prove the correct (and therefore containing unwinding numbers) version of $\sqrt{z_1 z_2}\stackrel{?}{=}\sqrt{z_1}\sqrt{z_2}$.

Lemma 1

$$\sqrt{z_1 z_2} = \sqrt{z_1}\sqrt{z_2}(-1)^{\mathcal{K}(\ln(z_1)+\ln(z_2))}. \tag{14}$$

[5]Just to remove the terms with unwinding numbers, as is done implicitly in some software systems, could be called "over-simplification."

Proof.

$$
\begin{aligned}
\sqrt{z_1 z_2} &= \exp\left(\frac{1}{2}\left(\ln(z_1 z_2)\right)\right) \\
&= \exp\left(\frac{1}{2}\left(\ln(z_1) + \ln(z_2) - 2\pi i \mathcal{K}(\ln(z_1) + \ln(z_2))\right)\right) \\
&= \sqrt{z_1}\sqrt{z_2}\exp\left(-\pi i \mathcal{K}(\ln(z_1) + \ln(z_2))\right) \\
&= \sqrt{z_1}\sqrt{z_2}(-1)^{\mathcal{K}(\ln(z_1)+\ln(z_2))}
\end{aligned}
$$

Lemma 2 *Whatever the value of z,*

$$\sqrt{1-z}\sqrt{1+z} = \sqrt{1-z^2}.$$

This is a classic example of a result that is “obvious” — the naïf just squares both sides, but in fact that loses information, and the identity requires proof. To show this, consider the apparently similar “result”:

$$\sqrt{z-1}\sqrt{1+z} \stackrel{?}{=} \sqrt{z^2-1}.$$

If we take $z = -2$, the left-hand side p x ffff R p x ff ff p x ff R x ff f ff

This is trivially true at $z = 0$. If it is false at any point, say z_0, then a path from z_0 to 0 must pass through a z where $\left|\arg\left(\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}}\right)\right| = \frac{\pi}{2}$, i.e. $\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}} = it$ for $t \in \mathbb{R}$. Squaring because, first, arg is continuous for $|z| \leq \pi/2$, and indeed for $|z| < \pi$, and, second, that the inputs to arg are themselves discontinuous only on $z > 1$ and $z < -1$, and on these half-lines, the arguments in question are 0 and $\pi/2$, which are acceptable. Coming back to the continuity along the path, we find that by squaring both sides, $z + i\sqrt{}$

$$
\begin{aligned}
&= \ln\left([1+i\frac{z}{\sqrt{1-z^2}}]/[1-i\frac{z}{\sqrt{1-z^2}}]\right) \\
&\quad +2\pi i\mathcal{K}\left(\ln(1+i\frac{z}{\sqrt{1-z^2}})-\ln(1-i\frac{z}{\sqrt{1-z^2}})\right) \\
&= \ln[iz+\sqrt{1-z^2}]^2 \\
&\quad +2\pi i\mathcal{K}(\ln(1+i\frac{z}{\sqrt{1-z^2}})-\ln(1-i\frac{z}{\sqrt{1-z^2}})) \\
&= 2i\arcsin(z) \\
&\quad -2\pi i\mathcal{K}\left(2\ln(iz+\sqrt{1-z^2})\right) \\
&\quad +2\pi i\mathcal{K}\left(\ln(1+i\frac{z}{\sqrt{1-z^2}})-\ln(1-i\frac{z}{\sqrt{1-z^2}})\right)
\end{aligned}
$$

The tendency for $\mathcal{K}$ factors to proliferate is clear. To simplify we proceed as follows. Consider first the term

$$\mathcal{K}(2\ln(iz+\sqrt{1-z^2})) \ .$$

For $|z|<1$, the real part of the input to the logarithm is positive and hence $\mathcal{K}=0$. For $|z|>1$, we solve for the critical case in which the input to $\mathcal{K}$ is $-i\pi$ and find only $z=r\exp(i\pi)$, with $r>1$. Therefore

$$\mathcal{K}(2\ln(iz+\sqrt{1-z^2})) = \mathcal{K}(-\ln(1+z)) \ .$$

Repeating the procedure with

$$\mathcal{K}(\ln(1+iz/\sqrt{1-z^2})-\ln(1-iz/\sqrt{1-z^2}))$$

shows that $\mathcal{K}\neq 0$ only for $z>1$. Therefore

$$\mathcal{K}(\ln(1+iz/\sqrt{1-z^2})-\ln(1-iz/\sqrt{1-z^2})) = \mathcal{K}(-\ln(1-z))$$

and so finally we get

$$\arctan\frac{z}{\sqrt{1-z^2}} = \arcsin(z) - \pi\mathcal{K}(-\ln(1+z)) + \pi\mathcal{K}(-\ln(1-z)) \ , \tag{16}$$

and this cannot be simplified further as an unconditional formula. Using equation (11), we can write

$$-\pi\mathcal{K}(-\ln(1+z))+\pi\mathcal{K}(-\ln(1-z)) = \begin{cases} +\pi & \text{if } z<-1 \\ -\pi & \text{if } z>1 \\ 0 & \text{elsewhere on } \mathbb{C}. \end{cases}$$

5 Unwinding Number: Insertion Rules

Unwinding numbers are normally inserted by use of equation ( ) and its converse:

$$\log\left(\frac{z_1}{z_2}\right) = \log(z_1)-\log(z_2)-2\pi\mathcal{K}\left(\log(z_1)-\log(z_2)\right). \tag{17}$$

Equation (10) may also be used, as may its close relative (also a special case of (17))

$$\log\left(\frac{1}{z}\right) =$$

7 More Automation

Here we look at a potential approach to mechanizing the reasoning required. We reconsider theorem 1: in particular we look at the branch cuts of the left-hand and right-hand sides, in the complex plane, which we will regard as the (x, y) plane.

LHS The question is the branch cut of $\ln\left(\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}}\right)$. Hence we have to solve

$$\Im\left(\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}}\right) = \pi.$$

In terms of (x, y)-coordinates, and cancelling the $\sqrt{2}$ in the denominators for simplicity[8] this means that

$$\sqrt{1+x+iy} + i\sqrt{1-x-iy} = \hat{x} + 0i : \qquad \hat{x} \in (-\infty, 0).$$

Isolating the first square root[9],

$$\begin{aligned} 1+x+iy &= \left(\hat{x} - i\sqrt{1-x-iy}\right)^2 \\ &= \hat{x}^2 - 2i\hat{x}\sqrt{1-x-iy} - (1-x-iy). \end{aligned}$$

Hence

$$(2-\hat{x}^2)^2 = 4 - 4\hat{x}^2 + \hat{x}^4 = -4\hat{x}^2(1-x-iy).$$

Equating imaginary parts, $0 = 4\hat{x}^2 y$, so $y = 0$. Equating real parts, $\frac{4+\hat{x}^4}{4\hat{x}^2} = x$, so $x \in [\sqrt{2}, \infty)$.

Hence the branch cut for the logarithm term is $(\sqrt{2} \le x < \infty, y = 0)$. We also need to add

Table 2: (20) at $z = 1 + i$

$\sqrt{\frac{1+z}{2}}$	$\sqrt{\frac{1-z}{2}}$	$\sqrt{\frac{1-z^2}{2}}$	value
+	+	+	0
−	+	+	$0.73 - 1.13i$
+	−	+	$0.73 - 1.13i$
−	−	+	0
+	+	−	$-0.73 + 1.13i$
−	+	−	0
+	−	−	0
−	−	−	$-0.73 + 1.13i$

Equating imaginary parts, $-2\hat{x}y = 0$, so $y = 0$. Using this, the real parts simplify to

$$-2x\hat{x} + \hat{x}^2 = -1,$$

i.e. $x = \frac{1+\hat{x}^2}{2\hat{x}}$ or $x \in (-\infty, -1]$.

We need also need to consider the branch cuts of the square root, viz. $1 - (x + iy)^2 = \hat{x} + 0i$ with $\hat{x} \in (-\infty, 0)$. The imaginary part of this is $2xy = 0$, so either $x = 0$ or $y = 0$. Substituting this in gives two critical lines: $(y = 0, x^2 = 1 - \hat{x})$ and $(x = 0, y^2 = \hat{x} - 1)$. given the range of $\hat{x}$, the latter has no solution, while the latter is $(y = 0, |x| > 1)$.

Lemma 3 *The difference between the left and right hand sides is locally constant, i.e. its derivative is zero.*

The derivative of the difference is

$$2 \quad \frac{1}{\sqrt{2+2}}$$

and denominator by -1. U

More generally, we have reduced the analytic difficulties of simplifying these functions to more algebraic ones, in areas where we hope that artificial intelligence and theorem proving stand a better chance of contributing to the problem.

References

[1] Abramowitz,M. & Stegun,I., *Handbook of Mathemat*

[14] Kahan,W., Branch Cuts for Complex Elementary Functions. *The State of Art in Numerical Analysis* (ed. A. Iserles & M.J.D. Powell), Clarendon Press, Oxford, 1 87, pp. 165–211.

[15] Litt,G., Unwinding numbers for the Logarithmic, Inverse Trigonometric and Inverse Hyperbolic Functions. M.Sc. project, Department of Applied Mathematics, University of Western Ontario, December 1 .

[16] Risch,R.H., Algebraic Properties of the Elementary Functions of Analysis. *Amer. J. Math.* **101** (1 7) pp. 743–75 .

[17] Steele,G.L.,Jr., *Common LISP: The Language*, 2nd. edition. Digital Press, 1 0.

[18] Trott,M., Visualization of Riemann Surfaces of Algebraic functions. *Mathematica in Education and Research* **6** (1 7), no. 4, pp. 15–36,

A Definition of the Elementary Inverse Functions

These definitions are taken from [5]. They agree with [1, ninth printing], but are more precise on the branch cuts, and agree with Maple with the exception of arccot, for the reasons explained in [5].

$$\arcsin z = -i \ln\left(\sqrt{1-z^2} + iz\right). \tag{21}$$

$$\arccos(z) = \frac{\pi}{2} - \arcsin(z) = \frac{2}{i} \ln\left(\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}}\right). \tag{22}$$

$$\arctan(z) = \frac{1}{2i}\left(\ln(1+iz) - \ln(1-iz)\right). \tag{23}$$

$$\operatorname{arccot} z = \frac{1}{2i} \ln\left(\frac{z+i}{z-i}\right) = \arctan\left(\frac{1}{z}\right). \tag{24}$$

$$\operatorname{arcsec}(z) = \arccos(1/z) = -i\ln(1/z + i\sqrt{1-1/z^2}), \tag{25}$$

with $\operatorname{arcsec}(0) = \frac{\pi}{2}$.

$$\operatorname{arccsc}(z) = \arcsin(1/z) = -i\ln(i/z + \sqrt{1-1/z^2}), \tag{26}$$

with $\operatorname{arccsc}(0) = 0$.

$$\operatorname{arcsinh}(z) = \ln\left(z + \sqrt{1+z^2}\right). \tag{27}$$

$$\operatorname{arccosh}(z) = 2\ln\left(\sqrt{\frac{z+1}{2}} + \sqrt{\frac{z-1}{2}}\right). \tag{28}$$

$$\operatorname{arctanh}(z) = \frac{1}{2}\left(\ln(1+z) - \ln(1-z)\right). \qquad (2\ \ )$$

$$\operatorname{arccoth}(z) = \frac{1}{2}\left(\ln(-1-z) - \ln(1-z)\right). \qquad (30)$$

$$\operatorname{arcsech}(z) = 2\ln\left(\sqrt{\frac{z+1}{2z}} + \sqrt{\frac{1-z}{2z}}\right). \qquad (31)$$

$$\operatorname{arccsch}(z) = \ln\left(\frac{1}{z} + \sqrt{1 + \left(\frac{1}{z}\right)^2}\right), \qquad (32)$$