– 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 not informed 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. F on the branch cut: instead $\log \overline{z} = \overline{\log z} + 2\pi i$ on the cut. Similarly,

$$\log\left(\frac{1}{z}\right) \neq -\log z \tag{5}$$

on the branch cut: instead $\log(1/z) = -\log z$

This definition has several attractive features: $\mathcal{K}(z)$ is integer-valued, and familiar in the sense that "everyone knows" that the multivalued logarithm can be written as the principal branch "plus $2\pi ik$ for some integer k"; it is single-valued; and it can be computed by a formula not involving logarithms. It does have a numerical difficulty, namely that you must decide if the imaginary part is an odd integer multiple of π or not, and this can be hard (or impossible in some exact arithmetic contexts), but the difficulty is inherent in the problem and cannot be repaired e.g. by putting the branch cuts elsewhere.
Some correct identities for elementary functions using $\mathcal{K}$ are given in Table 1.

1. $\quad = \log e^z + 2\pi i\mathcal{K}(\ )$.
2. $\mathcal{K}(a \log\ ) = 0\, \forall\ \ \in \mathbf{C}$ if and only if $-1 < a \leq 1$.
3. $\log\ _1 + \log\ _2 = \log(\ _1\ _2) + 2\pi i\mathcal{K}(\log\ _1 + \log\ _2)$.
4. $a \log\ \ = \log\ ^a + 2\pi i\mathcal{K}(a \log\ )$.
5. $\ ^{ab} = (\ ^a)^b e^{2\pi i b\mathcal{K}(a \log z)}$.

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

(7) can then be rescued as

$$\log(z_1 z_2) = \log z_1 + \log z_2 - 2\pi i\mathcal{K}(\log z_1 + \log z_2). \qquad (\ )$$

Similarly (4) can be rescued as

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

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

– Although not formally proposed in the same way in the computational community, one possible solution, often found in texts in complex analysis, is to accept the multi-valued nature of these functions (we adopt the common convention of using capital letters, e.g. Ln, to denote the multi-valued function), defining, for example

$$\operatorname{Arcsin} z = \{y \mid \sin y = z\}.$$

This leads to $\sqrt{z^2} = \{\pm z\}$, which has the advantage that it is valid throughout $\mathbf{C}$. Equation 7 is then rewritten as

$$\operatorname{Ln}(z_1 z_2) = \operatorname{Ln} z_1 + \operatorname{Ln} z_2\,, \qquad (11)$$

where addition is addition of sets ($A + B = \{a + b : a \in A, b \in B\}$) and equality is set equality[4].

[4] "The equation merely states that the sum of one of the (infinitely many) logarithms of $\ _1$ and one of the (infinitely many) logarithms of $\ _2$ can be found among the

However, it seems to lead in practice to very large and confusing formulae. More fundamentally, this approach does not say what will happen when the multi-valued functions are replaced by the single-valued ones of numerical programming languages.

A further problem that has not been stressed in the past is that this approach suffers from the same aliasing problem that naïve interval arithmetic does [6]. For example,

$$\mathrm{Ln}(z^2) = \mathrm{Ln}\, z + \mathrm{Ln}\, z \neq 2\,\mathrm{Ln}\, z\ ,$$

since $2\,\mathrm{Ln}(z) = \{2\log(z) + 4k\pi i : k \in \mathbf{Z}\}$, but $\mathrm{Ln}(z) + \mathrm{Ln}(z) = \{2\log(z) + 2k\pi i : k \in \mathbf{Z}\}$: indeed if $z = -1$, $\log(z^2) \notin 2\,\mathrm{Ln}(z)$. Hence this method is unduly pessimistic: it may fail to prove some identities that are true.

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 [4], showing that two alternative definitions of arccos are in fact equal:

(infinitely many) logarithms of ${}_1\ {}_2$, and conversely every logarithm of ${}_1\ {}_2$ can be represented as a sum of this kind (with a suitable choice of [elements of] Ln ${}_1$ and Ln ${}_2$)." [3, pp. 259–260] (our notation).

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

Theorem 1.

$$\frac{2}{i}\log\left(\sqrt{\frac{1+z}{2}}+i\sqrt{\frac{1-z}{2}}\right) = -i\log\left(z+i\sqrt{1-z^2}\right). \tag{12}$$

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.

$$\sqrt{z_1 z_2} = \sqrt{z_1}\sqrt{z_2}(-1)^{\mathcal{K}(\log z_1+\log z_2)}. \tag{13}$$

Proof.

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

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 schoolchild just squares both sides, but in fact that loses information, and the identity requires proof. To show this, consider the apparently similar "result"[6]:

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

If we take $z = i/2$, the left-hand side becomes $\sqrt{-3i/2}\,\sqrt{-i/2}$: the inputs to the square roots[7] have $\arg = -\pi/2$, so the square roots themselves have $\arg = -\pi/4$, and the product has $\arg = -\pi/2$, and therefore is $-i\sqrt{3}/2$. The right-hand side is $\sqrt{-3/4} = i\sqrt{\ }$

by the previous lemma. Also $2\log a = \log(a^2)$ if $\mathcal{K}(2\log a) = 0$, so we need only show this last stipulation, i.e. that

$$-\frac{\pi}{2} < \arg\left(\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}}\right) \leq \frac{\pi}{2}.$$

This is trivially true at $z = 0$. If it is false at any point, say z

4.3 arcsin and arctan

The aim of this section is to prove the correct expression for arcsin in terms of arctan. We note that we need to add unwinding number terms to deal with the two cuts $\Re z < -1$, $\Im z = 0$ and $\Re z > 1$, $\Im z = 0$.

Theorem 2.

$$\arcsin z = \arctan \frac{z}{\sqrt{1-z^2}} + \pi\mathcal{K}(-\log(1+z)) - \pi\mathcal{K}(-\log(1-z)). \tag{14}$$

We start from equations (1) and (21). Then

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

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

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

For $|z| < 1$, the real part of the input to the logarithm is positive and hence has argument in $(-\pi/2, \pi/2)$; therefore $\mathcal{K} = 0$. 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

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

Repeating the procedure with

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

shows that

5 The Unwinding Number: Insertion

We have seen that the systematic insertion of unwinding numbers while applying many "simplification" rules is necessary for mathematical correctness.

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

$$\log\left(\frac{z_1}{z_2}\right) = \log z_1 - \log z_2 -$$

- An unwinding number may divide the complex plane into two regions, one where it is non-zero and one where it is zero. A typical case of this is given in section 4.2. Here the proof methodology consists in examining the critical case, i.e. when the input to $\mathcal{K}$ has imaginary part $\pm\pi$, and examining when the functions contained in the input to $\mathcal{K}$ themselves have discontinuities.
- An unwinding number may correspond to the usual $+n\pi$: $n \in \mathbf{Z}$ of many trigonometric identities: examples of this are given in appendix B.

7 Conclusion

Unwinding number insertion permits the manipulation of logarithms, square roots etc., as well as the cancellation of functions and their inverses, while retaining mathematical correctness. This can be done completely algorithmically, and we claim this is one way, the only way we have seen, of guaranteeing mathematical correctness while "simplifying".

Unwinding number removal, where it is possible, then simplifies these results to the expected form. This is not a process that can currently be done algorithmically, but it is much better suited to current artificial intelligence techniques than the general problems of complex analysis.

When the unwinding numbers cannot be eliminated, they can often be converted into a case analysis that, while not ideal, is at least comprehensible while being mathematically correct.

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 Mathematical Functions with Formulas, Graphs, and Mathematical Tables. US Government Printing Office, 1964. 10th Printing December 1972.
2. Bradford,R.J., Algebraic Simplification of Multiple-Valued Functions. Proc. DISCO '92 (Springer Lecture Notes in Computer Science 721, ed. J.P. Fitch), Springer, 1993, pp. 13–21.
3. Carathéodory,C., *Theo y of functions of a complex va iable* (trans. F. Steinhardt), 2nd. ed., Chelsea Publ., New York, 1958.
4. Corless,R.M., Davenport,J.H., Jeffrey,D.J. & Watt,S.M., "According to Abramowitz and Stegun". To appear in SIGSAM Bulletin. OpenMath Project (Esprit 24969) deliverable 1.4.6.
5. Corless,R.M. & Jeffrey,D.J., The Unwinding Number. SIGSAM Bulletin **0** (1996) 2, pp. 28–35.
6. Davenport,J.H. & Fischer,H.-C., Manipulation of Expressions. *Imp oving Floating-Point P og amming* (ed. P.J.L. Wallis), Wiley, 1990, pp. 149–167.

9. IEEE Standard 754 for Binary Floating-Point Arithmetic. IEEE Inc., 1985.
10. 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, 1987, pp. 165–211.
11. 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 1999.
12. Steele,G.L.,Jr., Common LISP: The Language, 2nd. edition. Digital Press, 1990.

A Definition of the Elementary Inverse Functions

These definitions are taken from [4]. 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 [4].

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

$$\arccos(z) = \frac{\pi}{2} - \arcsin(z) = \frac{2}{i}\log\left(\sqrt{\frac{1+z}{2}} + i\sqrt{\frac{1-z}{2}}\right). \tag{20}$$

$$\arctan(z) = \frac{1}{2i}\left(\log(1+iz) - \log(1-iz)\right). \tag{21}$$

$$\operatorname{arccot} z = \frac{1}{2i}\log\left(\frac{z+i}{z-i}\right) = \arctan\left(\frac{1}{z}\right). \tag{22}$$

$$\operatorname{arcsec}(z) = \arccos(1/z) = -i\log(1/z + i\sqrt{1-1/z^2}), \tag{23}$$

with $\operatorname{arcsec}(0) = {}^{\pi}$

B Formulae for inverse functions

These formulae are taken from [11]. They make use of the secondary function csgn, which we define below in terms of $\mathcal{K}$ and was first defined by Dr. D. E. G. Hare as the piecewise function on the right hand side[8]:

$$\operatorname{csgn}(z) = (-1)^{\mathcal{K}(2\log(z))} = \begin{cases} +1 & \Re(z) > 0 \text{ or } \Re(z) = 0;\ \Im(z) \geq 0 \\ -1 & \Re(z) < 0 \text{ or } \Re(z) = 0;\ \Im(z) < 0 \end{cases}.$$

$$\arcsin(\sin(z)) = \begin{cases} z - 2\pi\mathcal{K}(zi) & \operatorname{csgn}(\cos z) = 1 \\ \pi - z - 2\pi\mathcal{K}(i(\pi - z)) & \operatorname{csgn}(\cos z) = -1 \end{cases}. \tag{31}$$

$$\arccos(\cos z) = \begin{cases} z - 2\pi\mathcal{K}(zi) & \operatorname{csgn}(\sin z) = 1 \\ -z - 2\pi\mathcal{K}(-zi) & \operatorname{csgn}(\sin z) = -1 \end{cases}. \tag{32}$$

$$\arctan(\tan z) = z + \pi\left(\mathcal{K}(-zi - \log\cos z) - \mathcal{K}(zi - \log\cos z)\right) \tag{33}$$

provided $z \neq \frac{\pi}{2} + n\pi$: $n \in \mathbf{Z}$.

$$\operatorname{arcsinh}(\sinh(z)) = \begin{cases} z - 2\pi i\mathcal{K}(z) & \operatorname{csgn}(\cosh z) = 1 \\ i\pi - z - 2\pi i\mathcal{K}(i\pi - z) & \operatorname{csgn}(\cosh z) = -1 \end{cases}. \tag{34}$$

$$\operatorname{arccosh}(\cosh z) = \begin{cases} z - 2\pi\mathcal{K}(z) & \operatorname{csgn}(\sinh z)\cos(n\pi) = 1 \\ -z - 2\pi i\mathcal{K}(-z) & \operatorname{csgn}(\sinh z)\cos(n\pi) = -1 \end{cases} \tag{35}$$

where $n = \mathcal{K}\left(\log(\cosh(z) - 1) + \log(\cosh(z) + 1)\right)$.

$$\operatorname{arctanh}(\tanh z) = z + i\pi\left(\mathcal{K}(z - \log\cosh z) - \mathcal{K}(z - \log\cosh z)\right) \tag{36}$$

provided $z \neq \frac{\pi}{2}i + in\pi$: $n \in \mathbf{Z}$.

[8] This function simplifies $\sqrt{\ \ ^2}$ to csgn( ). Dr. J. Carette observed that if we put $\omega = \exp(2\pi i/n)$, then the function defined by $\omega^{\mathcal{K}(n\log z)}$ and sometimes abbreviated by $C_n(\ )$, that generalizes csgn, is useful in simplifying $(\ \ ^n)^{1/n}$ (private communication).