The Geometry of $\mathbf{C}^n$ is Important for the Algebra of Elementary Functions

James H. Davenport

November 30, 2001

Abstract

On the one hand, we all "know" that $\sqrt{z^2}=z$, but on the other hand we know that this is false when $z=-1$. We all know that $\ln e^x = x$, and we all know that this is false when $x=2\pi i$. How do we imbue a computer algebra system with this sort of "knowledge"? Why is it that $\sqrt{x}\sqrt{y}=\sqrt{xy}$ is false in general ($x=y=-1$), but $\sqrt{1-z}\sqrt{1+z}=\sqrt{1-z^2}$ is true everywhere?

It is the contention of this paper that, only by considering the geometry of $\mathbf{C}$ (or $\mathbf{C}^n$ if there are n variables) induced by the various branch cuts can we hope to answer these questions even semi-algorithmically. This poses questions for geometry, and calls out for a more efficient form of cylindrical algebraic decomposition.

1 Introduction

It is possible to do algebra with the elementary functions of analysis, using only the tools of algebra, especially differential algebra, *if* one is content to regard these functions as abstract solutions of differential equations, so that

$$(\log f)' = \frac{f'}{f}; \qquad (\exp f)' = f' \exp f \tag{1}$$

[11], but of course these differential equations only define the functions "up to a constant". Furthermore, in this context, "constant" means something that diffierentiates to zero. So $\log\frac{1}{z}+\log z$ is a "constant", since $\left(\log\frac{1}{z}+\log z\right)' = \frac{\frac{-1}{z^2}}{\frac{1}{z}}+\frac{1}{z}=0$, but in fact this constant is 0 off the branch cut for log, but $2\pi i$ on the branch cut.

It is the thesis of this paper that, if instead one wishes to reason with these functions as genuine one-valued functions $\mathbf{C}\to\mathbf{C}$, or even $\mathbf{R}\to\mathbf{R}$, the geometry of $\mathbf{C}^n$ (or $\mathbf{R}^n$: n being the number of independent variables) that the branch cuts of these functions induce has to be taken into account. We claim also that this is true even for functions like

$$\sqrt{x} = \exp\left(\frac{1}{2}\log x\right), \tag{2}$$

since it is conventional to align the branch cuts so that equation (2) is true for these genuine one-valued functions, as well as being algebraically true *in abstracto*.

For the purposes of this paper, the precise definitions of log, arctan etc. as functions $\mathbf{C} \to \mathbf{C}$ (or $\mathbf{R} \to \mathbf{R}$) is that given in [5], which repeats, with more justification, the definitions of [1]. In particular, the branch cut of log (and $\sqrt{\ }$) is along the negative real axis, with the negative axis itself being continuous with the upper half-plane, so that $\lim_{\epsilon\to 0+} \log(-1+\epsilon i) = i\pi = \log(-1)$, but $\lim_{\epsilon\to 0-} \log(-1+\epsilon i) = -i\pi \neq \log(-1)$.

We will consider, as motivating examples, a variety of potential equations involving elementary functions.

$$\sqrt{z^2} \stackrel{?}{=} z \tag{3}$$

is one: it is patently false, even over $\mathbf{R}$, with $z=-1$ as a counterexample. Another pseudo-equation is

$$\log \overline{z} \stackrel{?}{=} \overline{\log z} : \tag{4}$$

this is false on the branch cut for log, and instead $\log \overline{z} = \overline{\log z} + 2\pi i$ on the cut. Similarly,

$$\log(\frac{1}{z}) \stackrel{?}{=} -\log z \tag{5}$$

is false on the branch cut: instead $\log(\frac{1}{z}) = -\log z + 2\pi i$ on the cut.

Another pseudo-equation that causes problems over the reals as well as over the complexes is

$$\arctan(x) + \arctan(y) \stackrel{?}{=} \arctan\left(\frac{x+y}{1-xy}\right). \tag{6}$$

2 How to handle multi-valued functions

[6] and [3] discuss various treatments of the inherently multi-valued functions such as $\sqrt{\ }$, log and arctan. The four approaches are given in the following sub-sections.

2.1 Signed zeros

This goes back to work of Kahan [9, 10], and distinguishes 0^+ from 0^-. This technique can resolve the issues raised by pseudo-equations (4) and (5), provided that we state that $\log(-1+0^+i) = i\pi$ but $\log(-1+0^-i) = -i\pi$. This method is no help in those situations where the pseudo-equation is invalid on a set of measure greater than 0, such as (3) and (6).

2.2 Genuinely exact equations

[8] claims that the most fundamental failure is $z \stackrel{?}{=} \log\exp z$. They therefore introduce the unwinding number, $\mathcal{K}$, defined by

$$\mathcal{K}(z) = \frac{z - \log\exp z}{2\pi i} = \left\lceil \frac{\Im z - \pi}{2\pi} \right\rceil \in \mathbf{Z}. \tag{7}$$

Equation (5) is then rescued as:

$$\log\frac{1}{x} = -\log x - 2\pi i \mathcal{K}(-\log(x)). \tag{8}$$

They would write the equation (6) as:

$$\arctan(x) + \arctan(y) = \arctan\left(\frac{x+y}{1-xy}\right) + \pi\mathcal{K}\left(2i(\arctan(x) + \arctan(y)\right). \tag{9}$$

2.3 Multi-valued functions

Though not explicitly advocated in the computational community, mathematical texts often urge us to treat these functions as multi-valued, defining, say, $\mathrm{Arctan}(x) = \{y \mid \tan x = y\}$ (the notational convention of using capital letters for these set-valued functions seems helpful). In this interpretation, equation (6), considered as

$$\mathrm{Arctan}(x) + \mathrm{Arctan}(y) \stackrel{?}{=} \mathrm{Arctan}\left(\frac{x+y}{1-xy}\right),$$

with the $+$ on the left-hand side representing element-by-element addition of sets, is valid.

In general, all the "algebraic" rules of simplification are valid in this context, e.g. with $\mathrm{Sqrt}(x) = \{y \mid y^2 = x\} = \{\pm\sqrt{x}\}$, it is the case that

$$\mathrm{Sqrt}(x)\,\mathrm{Sqrt}(y) = \mathrm{Sqrt}(xy), \tag{10}$$

whereas

$$\sqrt{x}\sqrt{y} \stackrel{?}{=} \sqrt{xy} \tag{11}$$

is not true: consider $x = y = -1$.

There are a few caveats that must be mentioned, though.

1. Cancellation is no longer trivial, since in principle $\mathrm{Arctan}(x) - \mathrm{Arctan}(x) = \{n\pi \mid n \in \mathbf{Z}\}$, rather than being zero.

2. Not all such multivalued functions have such simple expressions as

$$\mathrm{Arctan}(x) = \{\arctan(x) + n\pi \mid n \in \mathbf{Z}\}.$$

For example, $\text{Arcsin}(x) = \{\arcsin(x) + 2n\pi \mid n \in \mathbf{Z}\} \cup \{\pi - \arcsin(x) + 2n\pi \mid n \in \mathbf{Z}\}$. This problem combines with the previous one, so that, if $A = \arcsin(x)$,

$$\begin{aligned} A - A &= \{2n\pi \mid n \in \mathbf{Z}\} \cup \{2\arcsin(x) - \pi + 2n\pi \mid n \in \mathbf{Z}\} \\ &\quad \cup \{\pi - 2\arcsin(x) + 2n\pi \mid n \in \mathbf{Z}\} \\ &= \{2n\pi \mid n \in \mathbf{Z}\} + \{0, 2\arcsin(x) - \pi, \pi - 2\arcsin(x)\}. \end{aligned}$$

Note that this still depends on x, unlike the case of $\text{Arctan}(x) - \text{Arctan}(x)$.

In the case of non-elementary functions, as pointed out in [7] in the case of the Lambert W function, there may be no simple relationship between different branches.

3. However, some equations take on somewhat surprising forms in this context, e.g. the incorrect (23) simply becomes the correct

$$\text{Arcsin}(z) \subset \text{Arctan}\left(\frac{z}{\text{Sqrt}(1-z^2)}\right), \tag{12}$$

and if we want an equality of sets, we have

$$\text{Arcsin}(z) \cup \text{Arcsin}(-z) = \text{Arctan}\left(\frac{z}{\text{Sqrt}(1-z^2)}\right), \tag{13}$$

in which both sides take on four values in each 2π period. It is an open question to produce an alternative characterisation of just $\text{Arcsin}(z)$.

4. The equation

$$\text{Log}(z^2) = 2\,\text{Log}(z) \tag{14}$$

is not valid if we interpret $2\,\text{Log}(z)$ as $\{2y \mid \exp(y) = z\}$, since this has an indeterminacy of $4\pi i$, and the left-hand side has an indeterminacy of $2\pi i$. Instead we need to interpret $2\,\text{Log}(z)$ as $\text{Log}(z) + \text{Log}(z)$, and under this interpretation, equation (14) is true, as a specialisation of the correct equation

$$\text{Log}(z_1 z_2) = \text{Log}(z_1) + \text{Log}(z_2). \tag{15}$$

5. Pseudo-equations in which ambiguity disappears, such as $\sqrt{z^2} \stackrel{?}{=} z$, cannot be encoded as $\text{Sqrt}(z^2) = z$, since this is trying to equate a set and a number, but rather have to be encoded as $z \in \text{Sqrt}(z^2)$.

2.4 Riemann surfaces

It is often said that one should "consider the Riemann surface". Unfortunately, in the case of equation (6), $\arctan x$ is defined on one surface, $\arctan y$ on a second, and $\arctan\left(\frac{x+y}{1-xy}\right)$ on a third. It is not clear how "considering the Riemann surface" solves practical questions such as this identity.

Table 1: Decomposition of complex plane for $\sqrt{z^2}$

Region	$\Re(z)$	$\Im(z)$	$\sqrt{z^2}$
R_1	<0	—	$-z$
R_2	$=0$	<0	$-z$
R_3	$=0$	$=0$	$z=-z$
R_4	$=0$	>0	z
R_5	>0	—	z

3 $\sqrt{z^2}$

It follows from equation (10) that $(\mathrm{Sqrt}(z))^2 = \mathrm{Sqrt}(z^2)$. Since $z = (\sqrt{z})^2$ (not all equations are false!), this means that $z \in \mathrm{Sqrt}(z^2)$, i.e. $z = \pm\sqrt{z^2}$. The question then resolves to "plus or minus", or, in line with the thesis of this paper, "where on **C** is it plus, and where is it minus"?

$\sqrt{z}$ has its branch cut on the negative real axis, more precisely on $(-\infty, 0]$. Therefore the branch cut relevant to $\sqrt{z^2}$ is when $z^2 \in (-\infty, 0]$, i.e. $z \in [0, i\infty)$ or $z \in (-i\infty, 0]$. Note that, as we will see later, it is not correct to write $z \in [0, i\infty) \cup (-i\infty, 0] = (-i\infty, i\infty)$. Therefore the complex plane is divided into five regions, as shown in table 1. This analysis of cases can be reduced[1] (though as yet there is no known algorithm for doing this) to $\sqrt{z^2} = \mathrm{csgn}(z)z$, which can also be expressed as

$$\sqrt{z^2} = (-1)^{\mathcal{K}(2\log z)} z. \tag{16}$$

4 Equations (4) and (5)

These are very similar, and we will only deal with equation (5) in detail. The branch cut for $\log z$ is along the negative real axis $(\infty, 0)$, which is transformed into itself by $z \mapsto \frac{1}{z}$, and so is also the branch cut for $\log \frac{1}{z}$. This does not disconnect the complex plane, so there are only two cases: on the cut and off it. We know that $\log \frac{1}{z} \in -\mathrm{Log}(z)$, so the only corrections required to equation (5) are by adding multiples of $2\pi i$. Off the branch cut, an evaluation at $z = 1$ shows that no correction is necessary, whereas an evaluation at -1 shows that, on the branch cut, a correction of $2\pi i$ is necessary. If we wish, we can write this as

$$\log \frac{1}{z} = -\log z - 2\pi i \mathcal{K}(-\log z). \tag{17}$$

Similarly, equation (4) is rescued as

$$\log \overline{z} = \overline{\log z} - 2\pi i \mathcal{K}(\overline{\log z}). \tag{18}$$

[1] The csgn function was first defined in Maple. There is some uncertainty about csgn(0): is it 0 or 1, but for the reasons given in [3], we choose $\mathrm{csgn}(0) = 1$.

5 Additivity of arctan

In this section, we investigate the pseudo-equation (6), and attempt an algorithmic correction to it.

If we differentiate both sides of this equation with respect to x, we get

$$\frac{1}{1+x^2} \stackrel{?}{=} \frac{1}{1+x^2},$$

which is true, and if we differentiate both sides with respect to y, we get

$$\frac{1}{1+y^2} \stackrel{?}{=} \frac{1}{1+y^2},$$

which is also true.

Hence the difference between the two sides is a local constant, in the sense that its derivatives are zero, and so, between jumps, it is constant. Evaluating both sides of equation (6) at $x = y = 0$, we get 0 on both sides, so we conclude that here, the difference is indeed zero, so that equation (6) is indeed (locally) an identity. However, if we do the same at $x = 1.2$, $y = 0.9$, we get

$$1.608873152\ldots \stackrel{?}{=} -1.532719501\ldots:$$

clearly not valid, and indeed wrong by π.

5.1 $\mathbf{R} \to \mathbf{R}$

It is not obvious why there is a problem here, since arctan is a continuous, differentiable function, indeed bijection, $(-\infty, \infty) \to (-\pi/2, \pi/2)$. Indeed, we can define $\arctan(-\infty) = -\pi/2$, $\arctan(\infty) = \pi/2$ to get a bijection $[-\infty, \infty] \to [-\pi/2, \pi/2]$.

Unfortunately, $[-\infty, \infty]$ is not the right domain for this sort of analysis. One way of seeing this is to observe that, although $(-\infty, \infty) = \mathbf{R} \subset \mathbf{C}$, the analytic completion of $\mathbf{C}$ is the one-point completion $\mathbf{C} \cup \{\infty\}$, and $[-\infty, \infty] \not\subset \mathbf{C} \cup \{\infty\}$.

In fact, arctan has a branch cut[2] at infinity. When $x = \infty$ (or $y = \infty$), this might seem to cause a problem, since $\frac{x+y}{1-xy}$ tends to $\frac{-1}{y}$. However, the problem this causes is masked by the another, more serious, problem.

It is possible for $\frac{x+y}{1-xy}$ to pass through ∞ even when both x and y are finite, namely when $xy = 1$. If $\frac{x+y}{1-xy}$ goes from "large and positive" to "large and negative", then π is subtracted from the value of its arctan, and *vice versa*. We need to add a correction term, which can be done in various ways, e.g.[3]

$$\arctan(x) + \arctan(y) = \arctan\left(\frac{x+y}{1-xy}\right) + \begin{cases} \pi & x > 0,\ xy > 1 \\ 0 & x \geq 0,\ xy \leq 1 \\ 0 & x < 0,\ xy < 1 \\ -\pi & x < 0,\ xy \geq 1 \end{cases}. \tag{19}$$

[2]Some people refer to this a jump discontinuity, which of course it is, but it is better to think of it as a branch cut, since it could be moved elsewhere by another choice of branch cut: indeed, as pointed out in [5], there is no agreement on where the branch cut for the closely related function arccot should go.

[3]We are assuming a single ∞, with $\arctan(\infty) = \frac{\pi}{2}$.

Table 2: Cylindrical Decomposition of (x, y)-plane

x	xy	dimension	correction
>0	>1	2	π
>0	$=1$	1	0
>0	<1	2	0
$=0$	—	1	0
<0	>1	2	$-\pi$
<0	$=1$	1	$-\pi$
<0	<1	2	0

Table 3: 5 Regions of the (x, y)-plane

	R_1	R_2	R_3	R_4	R_5
xy	<1	$=1$	$=1$	>1	>1
x	—	>0	<0	>0	<0

Equation (9) is in one sense only another way of saying the same thing, except that it is coupled more clearly to the problem of "overflow" in adding the two arctan terms, and the boundary cases are dealt with consistently.

Equation (19) divides the (x, y)-plane into three regions (the two entries with zero correction term define one region, open on the lower-left boundary and closed on the top-right boundary. If we proceed algorithmically, by cylindrical algebraic decomposition [4], we in fact decompose the place into seven regions, as given in table 2. The technique of clustering [2] would reduce this, ideally to three regions as in equation (19), but certainly to five regions, which we will describe as follows.

One interesting question is the extent to which equation (19) can be reconstructed automatically. Evaluation at $x = y = 0$ shows that equation (6) is true there, and therefore throughout R_1. Similarly, evaluation at $x = y = 1$ shows that it is true there, and so throughout R_2. Evaluation at $x = y = -1$ shows that equation (6) is false, and needs a correction factor of $-\pi$ at that point, and therefore throughout R_3. By hand, evaluation at $x = y = 2$ shows that equation (6) is false, and needs a correction factor of π at that point, and therefore throughout R_4. However, it seems impossible to persuade Maple (release V.5) to simplify $2 \arctan 2 + \arctan \frac{4}{3}$ to π[4]. *Mutatis mutandis*, the same remarks apply to $x = y = -2$ and R_5. Of course, the author has chosen values of x and y at which the arctan function has a simple expression, and how could an algorithm do that?

The answer is that it need not. Since we know that $\arctan(x) + \arctan(y) \in \operatorname{Arctan}\left(\frac{x+y}{1-xy}\right)$, any error has to be a multiple of π. So a (sufficiently careful) numerical evaluation is all that is needed.

[4] Of course, direct simplification of this would use equation (6), which would spoil the point, but it ought to be doable via complex logarithms.

5.2 $\mathbf{C} \to \mathbf{C}$

The definition of $\arctan x$ is [1, 5]

$$\frac{1}{2i}\left(\log(1+ix) - \log(1-ix)\right).$$

Hence its branch cuts in the complex plane are along $1 + ix \in (-\infty, 0]$, i.e. $x \in [i, i\infty)$, and along $1 - ix \in (-\infty, 0]$, i.e. $x \in (-i\infty, -i]$. These do not disconnect the complex plane, so the only special cases are along them. The same is true of $\arctan y$ and $\arctan\left(\frac{x+y}{1-xy}\right)$. There is still the critical locus $xy = 1$, as in the real case.

We now have to determine how these interact. Write $x = x_R + ix_I$ etc. Then we have to consider various interesections.

- The intersection of the two branch cuts for $\arctan x$ and $\arctan y$, This is clearly possible, since the two constraints are independent.
- The intersection of the branch cuts for $\arctan\left(\frac{x+y}{1-xy}\right)$ with the branch cuts for $\arctan x$. If one solves for the two (real) equations resulting from $\Re\left(\frac{x+y}{1-xy}\right) = 0$ and $x_R = 0$, one is left with, *inter alia*, the resultant

$$-3y_R y_I^2 - y_I^6 y_R - 3y_I^2 y_R^5 - 3y_I^4 y_R^3 - y_R^7 + y_R + y_R^3 + 2y_R^3 y_I^2 - y_R^5 + 3y_I^4 y_R. \quad (20)$$

  The only solutions of this have $|y_R| \le 1$ and $|y_I| \le 1$, and therefore are not on the branch cut for $\arctan y$, except that when $y_R = 0$, y_I is unconstrained. Hence it is in fact possible to be on all the branch cuts of arctan simultaneously, say with $x = y = 2i$, and $\frac{x+y}{1-xy} = \frac{-4}{3}i$.
- The intersection of the branch cuts for $\frac{x+y}{1-xy}$ with the branch cuts for $\arctan x$. This is identical to the previous case.
- The intersection of the branch cut for $\arctan x$ with the critical locus $xy = 1$. It then follows that y_R has to be zero (since $x_I y_R = 0$), so the point is also on the branch cut of $\arctan y$, and possibly on the branch cut for $\arctan\left(\frac{x+y}{1-xy}\right)$.
- The intersection of the branch cut for $\arctan\left(\frac{x+y}{1-xy}\right)$ with the critical locus $xy = 1$ cannot happen.

In sum, the situation here is very complicated, but can be reduced to a finite number of regions which need to be tested. It in fact turns out that the move to the complex plane does not, in fact, add any further complication.

The unwinding number approach tells us easily that the correction is

$$-\pi\mathcal{K}\Bigg(\ln(1+ix) - \ln(1-ix) + \ln(1+iy) - \ln(1-iy) + \quad (21)$$
$$\ln\left(1 + \frac{i(x+y)}{1-xy}\right) - \ln\left(1 - \frac{i(x+y)}{1-xy}\right)\Bigg)$$

but there seems to be no easy simplification of this.

6 Strategies for simplifying elementary expressions

Traditionally, computer algebra systems, and indeed humans, are bad at simplifying expressions containing the (inverse) elementary functions. Humans fail to spot special cases, often with serious consequences [10, pp. 178–9], and computer algebra writers are torn between:

- being careful, as in Maple's `simplify(...)`, and therefore failing to make correct simplifications such as $\sqrt{1-z}\sqrt{1+z} \to \sqrt{1-z^2}$;
- using the full range of algebraic rules regardless of correctness, as in Maple's `simplify(...,symbolic)`, and therefore making incorrect simplification such as $\sqrt{z^2} \to z$.

We now describe two strategies (at the current state of development, it would not be accurate to describe them as algorithms), based on the views outlined in sections 2.2 and 2.3 and illustrated by the examples above.

6.1 The Unwinding Number Approach

This approach, outlined in [3, 6], works as follows.

While a reduction exists
    Perform a reduction
        Generally introducing unwinding numbers
    Remove as many unwinding numbers as possible
See if case-by-case analysis removes unwinding numbers

For example, the equivalent of equation (11) would be

$$\sqrt{x}\sqrt{y} = \sqrt{xy}(-1)^{K(\log x + \log y)}. \tag{22}$$

An illustration of this method would be the following proof from [6].

Lemma 1 *Whatever the value of z,*

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

Proof. It is sufficient to show that the unwinding number term in equation (22) is zero. Whatever the value of z, $1+z$ and $1-z$ have imaginary parts of opposite signs. Without loss of generality, assume $\Im z \geq 0$. Then $0 \leq \arg(1+z) \leq \pi$ and $-\pi < \arg(1-z) \leq 0$. Therefore their sum, which is the imaginary part of $\ln(1+z) + \ln(1-z)$, is in $(-\pi, \pi]$. Hence the unwinding number is indeed zero.

Note the manual reasoning required to remove the unwinding number term that equation (22) introduced.

6.2 The "Regions of Validity" Approach

This approach[5], used informally in the previous sections, works as follows.

Let E' be the multi-valued version of E.
While a reduction of E' exists
    reduce E' to yield a new E'
Let E'' be the single-valued version of the reduced E'.
Analyse $\mathbf{C}^n$ to determine corrections required to $E \stackrel{?}{=} E''$.

This method would prove lemma 1 the following way

Proof. $E' := \mathrm{Sqrt}(1-z)\,\mathrm{Sqrt}(1+z)$. This reduces to $\mathrm{Sqrt}(1-z^2)$, so we know that $\sqrt{1-z}\sqrt{1+z} \in \mathrm{Sqrt}(1-z^2)$. Hence we have to analyse $\sqrt{1-z}\sqrt{1+z} \stackrel{?}{=} \sqrt{1-z^2}$. The relevant branch cuts are as follows.

For $\sqrt{1-z}$: $1-z \in (-\infty, 0]$, so $z \in [1, \infty)$.

For $\sqrt{1+z}$: $1+z \in (-\infty, 0]$, so $z \in (-\infty, -1]$.

For $\sqrt{1-z^2}$: $1-z^2 \in (-\infty, 0]$, so $z^2 \in [1, \infty)$, and $z \in [1, \infty)$ or $z \in (-\infty, -1]$.

These branch cuts do not disconnect the complex plane, so the three regions are $[1, \infty)$, $(-\infty, -1]$ and the rest of the plane.

$[1, \infty)$ Take $z = 3$. Then $\sqrt{1-z}\sqrt{1+z} = \sqrt{-2}\sqrt{4} = 2\sqrt{2}i$. $\sqrt{1-z^2} = \sqrt{-8} = 2\sqrt{2}i$, so no correction is needed.

$(-\infty, -1]$ Take $z = -3$. Then $\sqrt{1-z}\sqrt{1+z} = \sqrt{4}\sqrt{-2} = 2\sqrt{2}i$. $\sqrt{1-z^2} = \sqrt{-8} = 2\sqrt{2}i$, so no correction is needed.

rest of plane Take $z = i$. Then $\sqrt{1-z}\sqrt{1+z} = \sqrt{1-i}\sqrt{1+i} = \sqrt{2}$ (clearly the correct norm, and the arguments cancel). $\sqrt{1-z^2} = \sqrt{2}$, so again no correction is necessary.

Note that, while it might appear that some intelligence was used to select the sample values, any sample value s such that $\sqrt{1-s^2} \neq -\sqrt{1-s^2}$ would do, as long as we evaluated $\sqrt{1-s}\sqrt{1+s}$ to enough precision to be sure which of the two values it was taking.

7 Strategies to Algorithms?

In this section, we explore what would be necessary to convert the strategies outlined above into complete (and possibly even efficient) algorithms.

[5] This approach is similar to the approach "how to decide where two analytic expressions describe the same function" of [10].

7.1 The Unwinding Number Approach

A prime requirement of this method is a complete set of *correct* transformation rules. Thus, instead of the incorrect equation (11), we need the correct equation (22). While not conceptually very difficult, to my knowledge no such complete list has been built up. It would have to include rules for "incorrect"[8] applications of inverse functions, so that the incorrect $z \stackrel{?}{=} \log \exp z$ becomes the correct $z = \log \exp z + 2\pi i \mathcal{K}(z)$, and similarly the incorrect equation (3) becomes the correct equation (16). The often-quoted, but incorrect[6], equation

$$\arcsin z \stackrel{?}{=} \arctan \frac{z}{\sqrt{1-z^2}} \tag{23}$$

has to be replaced by

$$\arcsin z = \arctan\left(\frac{z}{\sqrt{1-z^2}}\right) + \pi\mathcal{K}(-\ln(1+z)) - \pi\mathcal{K}(-\ln(1-z)). \tag{24}$$

Another obstacle, as pointed out in [6] is the tendency for expressions involving $\mathcal{K}$ to proliferate: in the course of proving equation (24), they had to handle equations such as the following.

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

The third problem is that of understanding the result, which in this case means reducing

$$-\pi\mathcal{K}\left(2\ln\left(iz+\sqrt{1-z^2}\right)\right) + \pi\mathcal{K}\left(\ln\left(1+i\frac{z}{\sqrt{1-z^2}}\right) - \ln\left(1-i\frac{z}{\sqrt{1-z^2}}\right)\right).$$

to $\pi\mathcal{K}(-\ln(1+z)) - \pi\mathcal{K}(-\ln(1-z))$. This may well result in a difficult case-by-case analysis, so there may be no useful overall picture. See formula (21) for another example.

7.2 The "Regions of Validity" Approach

In this approach, the first problem is also that of finding a "correct" set of equations, in this case replacing the incorrect equation (11) by the equality of sets in equation (22). While not conceptually very difficult, to my knowledge no complete list of such equations has been built up. It would have to include rules for "incorrect"[8] applications of inverse functions, so that the incorrect $z \stackrel{?}{=} \log \exp z$

[6] With our definitions. Derive's definition of arctan is to make this true, but this means that it deviates from [1].

becomes the correct $z \in \operatorname{Log} \exp z$, and similarly the incorrect equation (11) becomes the correct equation (10).

One problem, though, is that not all simplification rules remain set equalities, as in equation(12). Provided the strategy only expands E', then no real harm is done, but it must never contract it. There is also a problem with equations such as $z \in \operatorname{Sqrt}(z^2)$. It does not seem to be trivial to make an algorithm out of this.

Handling the decompositions of $\mathbf{R}^{2n}$ (as we see $\mathbf{C}^n$) is also difficult. The existing technology of cylindrical algebraic decomposition [4] has the following defects when applied to the problem at hand.

- As we saw in section 5, it does not handle directly descriptions such as the branch cut for log, i.e. $\Re(z) < 0 \wedge \Im(z) = 0$. Instead one has to introduce the two planes $\Re(z) = 0$ and $\Im(z) = 0$, even though $\Re(z) = 0$ is irrelevant off $\Im(z) = 0$.
- As has been remarked by many authors, even for the original purposes of quantifier elimination, cylindrical algebraic decomposition is *too* powerful, since it solves not only the question at hand, but also all other questions[7] relating to the same set of equations.
- The algorithm produces far too many regions. This can partly be solved by the technique of clustering [2], but this is a retrospective technique (produce too many, then reduce), and there is no work on how this interacts with the first problem.

7.3 Putting it all together

Assuming that we eventually want an identity relating single-valued functions, what is the simplest was of expressing it? Equation (23), viz.

$$\arcsin z = \arctan\left(\frac{z}{\sqrt{1-z^2}}\right) + \pi\mathcal{K}(-\ln(1+z)) - \pi\mathcal{K}(-\ln(1-z)).$$

is hardly simple, but is it any better as

$$\arcsin z = \overline{\arctan\left(\frac{z}{\sqrt{1-z^2}}\right)}? \tag{25}$$

The latter equation does not easily imply that, almost everywhere, $\arcsin z = \arctan\left(\frac{z}{\sqrt{1-z^2}}\right)$, whereas, once used to reading formulae involving $\mathcal{K}$, the former does.

If we have an analysis in terms of cases, is there any algorithmic reconstruction in terms of $\mathcal{K}$? None is known, and it is worth pointing out that one can

[7] To be accurate, all other questions with the same order of variables in the quantifiers, though the quantifiers themselves, and the Boolean combination of the polynomials, may be different.

also produce "nonsense" justifications in terms of $\mathcal{K}$, e.g. equations (17) and (18) can be combined to give the correct, but unhelpful

$$\log \frac{1}{z} = -\log z - 2\pi i \mathcal{K}(\overline{\log z}). \tag{26}$$

8 Conclusion

There are several forms of "identities" relating complex-valued (or even real-valued) elementary functions.

1. Those that are true, and easily provable to be so. This class has not had much mention in this paper, but, in particular, all identities involving forward functions (exp, trigonometric and hyperbolic functions, raising to integer powers) only are correct, as are those in which inverse functions are applied correctly[8].

2. Those that are true, but whose naïve proof is false. A good example of this is Lemma 1, where the naïve proof by multiplying out square roots needs a lot of help to make it correct. These are the ones that computer algebra systems ought to get *reliably*, but currently cannot.

3. Those that are "almost true". Equations (5) and (4) are good examples of this. In a numerical context, these can effctively be taken as true, since the chance of landing on an exception value is very small, and an adaptive solver will increase the sampling density near the exceptional value to reduce its impact. In some numerical contexts, signed zeros (see section 2.1) will obviate the problem. In computer algebra, there is no such excuse: the answer given is wrong. However, not warning the user that there is an "almost good enough" simplification is unhelpful, to say the least. As we saw, sometimes these can be re-written by means of a "double conjugate" trick, as in equation (25).

4. Those that are only partially true, i.e. only true on some subset, of measure less than one, of the argument range. Equations (3) and (6) are good examples of this.

Computer algebra systems currently give no help with the third and fourth type of "identity", or in distinguishing them from the second. The techniques outlined in this paper are, it is hoped, a start in filling this gap.

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.

[8] For example, tan is many–one, hence Arctan is one–many, and therefore arctan needs to be applied carefully. In particular, $\tan(\arctan x) = x$ is valid, but $\arctan(\tan x) = x$ is not.

[2] Arnon,D.S. & McCallum,S., Cylindrical Algebraic Decomposition II: An Adjacency Algorithm for the Plane. *SIAM J. Comp.* **13** (1984) pp. 878–889.

[3] Bradford,R.J., Corless,R.M., Davenport,J.H., Jeffrey,D.J. & Watt,S.M., Reasoning about the Elementary Functions of Complex Analysis. To appear in *Annals of Mathematics and Artificial Analysis.*

[4] Collins,G.E., Quantifier Elimination for Real Closed Fields by Cylindrical Algebraic Decomposition. Proc. 2nd. GI Conference Automata Theory & Formal Languages (Springer Lecture Notes in Computer Science 33) pp. 134–183.

[5] Corless,R.M., Davenport,J.H., Jeffrey,D.J. & Watt,S.M., "According to Abramowitz and Stegun". *SIGSAM Bulletin* **34** (2000) 2, pp. 58–65.

[6] Corless,R.M., Davenport,J.H., Jeffrey,D.J., Litt,G. & Watt,S.M., Reasoning about the Elementary Functions of Complex Analysis. Artificial Intelligence and Symbolic Computation (ed. John A. Campbell & Eugenio Roanes-Lozano), Springer Lecture Notes in Artificial Intelligence Vol. 1930, Springer-Verlag 2001, pp. 115–126.

[7] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey & D.E. Knuth, On the Lambert W function. *Advances in Computational Mathematics* $\underline{5}$ (1996) pp. 329–359.

[8] Corless,R.M. & Jeffrey,D.J., The Unwinding Number. *SIGSAM Bulletin* **30** (1996) 2, issue 116, pp. 28–35.

[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] Risch,R.H., Algebraic Properties of the Elementary Functions of Analysis. *Amer. J. Math.* **101** (1979) pp. 743–759.