The Challenges of Multivalued “Functions”

James Davenport

University of Bath

5 July 2010

Notation

$\mathbf{P}(A)$ denotes the power set of the set A.
For a function f, we write $\mathrm{graph}(f)$ for $\{(x, f(x)) : x \in \mathrm{dom}(f)\}$ and $\mathrm{graph}(f)^T$ for $\{(f(x), x) : x \in \mathrm{dom}(f)\}$.

Convention

Where an underspecified object, such as $\sqrt{x}$, occurs more than once in a formula, the same *value, or interpretation, is meant at each occurrence.*

For example, $\sqrt{x} \cdot \frac{1}{\sqrt{x}} = 1$ for non-zero x, even though one might think that one root might be positive and the other negative. More seriously, in the formula for the roots of a cubic $x^3 + bx + c$,

$$\frac{1}{6}\sqrt[3]{-108\,c + 12\sqrt{12\,b^3 + 81\,c^2}} - \frac{2b}{\sqrt[3]{-108\,c + 12\sqrt{12\,b^3 + 81\,c^2}}},$$

the two occurrences of $\sqrt{12\,b^3 + 81\,c^2}$ are meant to have the same value, similarly $\sqrt[3]{-108\,c + 12\sqrt{12\,b^3 + 81\,c^2}}$.

Notation continued

We use the notation $A \stackrel{?}{=} B$ to denote what is normally given in the literature as an equality with an $=$ sign, but where one of the purposes of this paper is to question the meaning of that, very overloaded, symbol.
We will often need to refer to polar coordinates for the complex plane.

Notation

We write $\mathbf{C} \equiv X \mathop{\times}_{\mathrm{polar}} Y$ *for such a representation*
$z = re^{i\theta} : r \in X \wedge \theta \in Y$.

z (and variants) is complex, x and y are real (often $z = x + iy$).

Examples

As statements about functions, we consider the following.

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

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

$$\log z_1 + \log z_2 \stackrel{?}{=} \log z_1 z_2. \tag{3}$$

$$\arctan x + \arctan y \stackrel{?}{=} \arctan\left(\frac{x+y}{1-xy}\right). \tag{4}$$

(1) is valid for $\Re(z) > 0$, also for $\Re(z) = 0$, $\Im(z) > 0$.

(2) is valid everywhere, despite the resemblance to (1).

(3) is valid with $-\pi < \arg(z_1) + \arg(z_2) \leq \pi$.

(4) is valid, even for *real* x, y, only when $xy < 1$.

That's curious: arctan is nice

(as a real-valued function, at least).

In fact there is a "branch cut at infinity", since $\lim_{x\to+\infty} \arctan x = \frac{\pi}{2}$, whereas $\lim_{x\to-\infty} \arctan x = -\frac{\pi}{2}$ and $xy = 1$ therefore falls on this cut of the right-hand side of (4).

This is also the branch cut that many symbolic integrators (used to) fall over.

The (Bourbakist) Theory

In principle, (pure) mathematics is clear.

> *On dit qu'un graphe F est un graphe fonctionnel si, pour tout x, il existe au plus un objet correspondant à x par F (I, p. 40). On dit qu'une correspondance $f = (F, A, B)$ est une fonction si son graphe F est un graphe fonctionnel, et si son ensemble de départ A est égal à son ensemble de définition $\mathrm{pr}_1 F$ [pr_1 is "projection on the first component"].* [Bourbaki, Ensembles]

So for Bourbaki a function includes the definition of the domain and codomain, and is *total* and *single-valued*. We will write $(F, A, B)_{\mathcal{B}}$ for such a function definition.

Bourbaki meets (analytic) reality

The natural domains of definition of analytic functions are simply connected open sets, generally referred to as "$\mathbf{C}^n$ with branch cuts". The table maker, or programmer, abhors "undefined", and extends definitions to the whole of $\mathbf{C}^n$ by making the values on the branch cut 'adhere' to one side or the other, extending a definition from D, a slit version of $\mathbf{C}^n$, to the whole of $\mathbf{C}^n$.
Rather than just writing $\mathbf{C}^n$ for the domain, we will explicitly write $\overline{D}$ to indicate that it is an extension of the definition with domain D.

The multivalued view

Analysts sometimes take a completely multivalued view, as here, discussing our exemplar (3).

> *The equation merely states that the sum of one of the (infinitely many) logarithms of z_1 and one of the (infinitely many) logarithms of z_2 can be found among the (infinitely many) logarithms of $z_1 z_2$, and conversely every logarithm of $z_1 z_2$ can be represented as a sum of this kind (with a suitable choice of $\log z_1$ and $\log z_2$).*
>
> *[Caratheodory pp. 259–260] (our notation)*

Here we essentially have $\left(\mathrm{graph}(\exp)^T, \mathbf{C}, \mathbf{P}(\mathbf{C})\right)_{\mathcal{B}}$.
But, of course, this is far from being a surjection!

The Riemann surface view

This is closely related to the multivalued view. This can be seen as $\left(\mathrm{graph}(\exp)^T, \mathbf{C}, \mathcal{R}_{\log z}\right)_{\mathcal{B}}$, where $\mathcal{R}_{\log z}$ signifies the Riemann surface corresponding to the function $\log z$.
The Riemann surface view is discussed in [Bradfordetal2002, Section 2.4], which concludes

> *Riemann surfaces are a beautiful conceptual scheme, but at the moment they are not computational schemes.*

The additional structure imparted by $\mathcal{R}_{\log z}$ (over that of $\mathbf{P}(\mathbf{C})$) is undoubtedly very useful from the theoretical point of view, and provides a global setting for the next, essentially local, view.

The branch view: [Cartan1973]

p. 32 "The mapping $y \mapsto e^{iy}$ induces an isomorphism ϕ of the quotient group $\mathbf{R}/2\pi\mathbf{Z}$ on the group $\mathbf{U}$. The inverse isomorphism ϕ^{-1} of $\mathbf{U}$ on $\mathbf{R}/\pi\mathbf{Z}$ associates with any complex number u such that $|u| = 1$, a real number which is defined up to the addition of an integral multiple of 2π; this class of numbers is called the argument of u and is denoted by $\arg u$." In our notation this is $\left(\mathrm{graph}(\phi)^T, U, \mathbf{R}/2\pi\mathbf{Z}\right)_{\mathcal{B}}$.

p. 33 "We define

$$\log t = \log |t| + i \arg t, \tag{5}$$

which is a complex number defined only up to addition of an integral multiple of $2\pi i$." In our notation this is $((5), \mathbf{C}, \mathbf{C}/2\pi i\mathbf{Z})_{\mathcal{B}}$.

p. 33 "For any complex numbers t and t' both $\neq 0$ and for any values of $\log t$, $\log t'$ and $\log tt'$, we have

$$\log tt' = \log t + \log t' \pmod{2\pi i}." \tag{6}$$

p. 33 "So far, we have not defined $\log t$ as a *function* in the proper sense of the word".

p. 61 "$\log z$ has a branch in any simply connected open set which does not contain 0."

So any given branch would be $(G, D, I)_{\mathcal{B}}$, where D is a simply connected open set which does not contain 0, G is a graph obtained from one element of the graph (i.e. a pair $(z, \log(z))$ for some $z \in D$) by analytic continuation, and I is the relevant image set.

An 'Applied' view

Applied mathematics is sometimes less unambiguous.

> *...when we say that $f(x)$ is a function of x in some range of values of x we mean that for every value of x in the range one or more values of $f(x)$ exist. ... It will usually also be required that the function shall be single-valued, but not necessarily.* [JeffreysJeffreys1956, p. 17]

So for these authors, a function might or might not be multivalued.

The table-maker's point of view

This is essentially also the computer designer's point of view, be it hardware or software. From this point of view, it is necessary to specify how to compute $f(x)$ for any given x, irrespective of any "context", and return a single value, even though, in the text accompanying the tables, we may read "only defined up to multiples of $2\pi i$" or some such.

(1) If we substitute $z = -2$, we obtain $\sqrt{-3}\sqrt{-1} \stackrel{?}{=} \sqrt{3}$, which is false, so the statement is not universally true.

(2) It is impossible to refute this statement.

(3) If we take $z_1 = z_2 = -1$, we obtain $\log(-1) + \log(-1) \stackrel{?}{=} \log 1$, i.e. $i\pi + i\pi \stackrel{?}{=} 0$, so the statement is not universally true.

(4) If we take $x = y = \sqrt{3}$, we get $\frac{\pi}{3} + \frac{\pi}{3} \stackrel{?}{=} \frac{-\pi}{3}$, so the statement is not universally true.

Differential Algebra

A completely different point of view of view is the differential-algebraic one. Here $\sqrt{1-z}$ is an object whose square is $1-z$, formally definable as w in $\mathbf{C}(z)[w]/(w^2-(1-z))$. Similarly $\log z$ is a new symbol θ such that $\theta' = 1/z$, and so on for other elementary expressions.

(1) The left-hand side is $vw \in K = \mathbf{C}(z)[v,w]/(v^2-(z-1), w^2-(z+1))$, and the right-hand side is $u \in \mathbf{C}(z)[v,w]/(u^2-(z^2-1))$. But to write the equation we have to express $u^2-(z^2-1)$ in K, and it is no longer irreducible, being $(u-vw)(u+vw)$. Depending on which factor we take as the defining polynomial, the equation

$$vw = u \tag{1'}$$

is either true or false, and we have to decide which. Once we have decided which, the equation becomes trivially true (or false). The problem is that, with the standard interpretations (which of course takes us *outside* differential algebra), the answer is "it depends on which value of z you have".

(2) The analysis is identical up to the standard interpretations, at which point it transpires that, for the standard interpretations, $vw = u$ is true for all values of z. But, of course, this is what we were trying to prove in the first place.

(3) Here we define θ_1 such that $\frac{\partial \theta_1}{\partial z_1} = \frac{1}{z_1}$ (and $\frac{\partial \theta_1}{\partial z_2} = 0$), θ_2 such that $\frac{\partial \theta_2}{\partial z_2} = \frac{1}{z_2}$ (and $\frac{\partial \theta_2}{\partial z_1} = 0$) and θ_3 such that $\frac{\partial \theta_3}{\partial z_1} = \frac{z_2}{z_1}$ and $\frac{\partial \theta_3}{\partial z_2} = \frac{z_1}{z_2}$. If we then consider $\eta = \theta_1 + \theta_2 - \theta_3$, we see that

$$\frac{\partial \eta}{\partial z_1} = \frac{\partial \eta}{\partial z_2} = 0 \qquad (3'),$$

which implies that η "is a constant".

(4) Again, the difference between the two sides "is a constant".

Differential Algebra: “Constants”

We have said “is a constant”, since the standard definition in differential algebra is that a constant is an object all of whose derivatives are 0. Of course, this is related to the usual definition by the following.

Proposition

A differentiable function $f : \mathbf{C}^n \to \mathbf{C}$, all of whose first derivatives are 0 in a connected open set D, takes a single value throughout D, i.e. is a constant in the usual sense over D.

The difference between the two can be seen in these “corrected” versions of (3) and (4), where the choice expressions are the “constants”.

Equations with constants

$$\log z_1 + \log z_2 = \log z_1 z_2 + \begin{cases} 2\pi i & \arg z_1 + \arg z_2 > \pi \\ 0 & -\pi < \arg z_1 + \arg z_2 < \pi \\ -2\pi i & \arg z_1 + \arg z_2 < -\pi \end{cases} \tag{3''}$$

$$\arctan x + \arctan y = \arctan\left(\frac{x+y}{1-xy}\right) + \begin{cases} \pi & xy > 1, x > 0 \\ 0 & xy < 1 \\ -\pi & xy > 1, x < 0 \end{cases} \tag{4''}$$

Equation ($4''$) appears as such, *with* the correction term, as [Apostol1961 p. 205, ex. 13].

The pragmatic (meta-)view

So, which view actually prevails?
The answer depends on the context, but it seems to the current author that the view of *most* mathematicians, *most* of the time, is a blend of “local” and “differential algebra”.
This works because the definitions of differential algebra give rise to power series, and therefore, given “suitable” initial conditions, the *expressions* of differential algebra can be translated into *algorithms* expressed by power series, which “normally” correspond to *functions* in some open set around those initial conditions.
Whether this is an ‘adequate’ open set is a more difficult matter.

A textbook example [Apostol1961 p. 189]

$$\int \frac{2^{\sqrt{x}}}{\sqrt{x}} \mathrm{d}x = \frac{2^{1+\sqrt{x}}}{\log 2} + C \tag{7}$$

(We ignore any problems posed by "$\log 2$".)
The proof given is purely in the setting of differential algebra, despite the fact that the source text is entitled *Calculus*.
Translated into that language, we are working in $\mathbf{C}(x, u, \theta)$ where $u^2 = x$ and

$$\theta' = (\theta \log 2)/2u. \tag{8}$$

(We note that equation (8) implicitly gives effect to Convention 1, in that θ' represents $\left(2^{\sqrt{x}} \log 2\right) /2\sqrt{x}$ where the two occurrences of $\sqrt{x}$ represent the same object.) Similarly the right-hand side is $\frac{2\theta}{\log 2} + C$. Note that, having introduced $2^{\sqrt{x}}$, $2^{1+\sqrt{x}}$ is not legitimate in diferential algebra, since the Risch Structure Theorem will tell us that there is a relationship between θ and an η standing for $2^{1+\sqrt{x}}$, viz. that η/θ is constant.

A different point of view

It is possible to take a different approach to these functions, and say, effectively, that "each use of each function symbol means what I mean it to mean at that point". This is completely incompatible with the table-maker's, or the computer's, point of view, but has its adherents, and indeed uses.

A classic example of this is given in [Henrici1974 pp. 294–8]. He considers the Joukowski map $f : z \mapsto \frac{1}{2}\left(z + \frac{1}{z}\right)$ and its inverse $f^{-1} : w \mapsto w + \sqrt{w^2 - 1}$, in two different cases. If we regard these functions as $(f, D, D')_{\mathcal{B}}$ and $\left(f^{-1}, D', D\right)_{\mathcal{B}}$, the cases are as follows.

(i): $D = \{z : |z| > 1\}$. Here $D' = \mathbf{C} \setminus [-1, 1]$. The problem with f^{-1} is interpreting $\sqrt{w^2 - 1}$ so that $|w + \sqrt{w^2 - 1}| > 1$.

(ii): $D = \{z : \Im(Z) > 0\}$. Here $D' = \mathbf{C} \setminus ((-\infty, -1] \cup [1, \infty))$, and the problem with f^{-1} is interpreting $\sqrt{w^2 - 1}$ so that $\Im(w + \sqrt{w^2 - 1}) > 0$.

We require f^{-1} to be injective, which is a problem, since in both cases $w \mapsto w^2$ is not.

Hence the author applies (1) formally (though he does not say so explicitly), and writes

$$f^{-1}(w) = w + \sqrt{w+1}\sqrt{w-1}. \tag{9}$$

(i) Here he takes both $\sqrt{w+1}$ and $\sqrt{w-1}$ to be uses of the square-root function from [AbramowitzStegun1964], viz. $\left(\sqrt{\ },\mathbf{C},\mathbf{C}\equiv\mathbf{R}^{+}\mathop{\times}_{\text{polar}}(-\frac{\pi}{2},\frac{\pi}{2}]\cup\{0\}\right)_{\mathcal{B}}$. We should note that this means that $\sqrt{w+1}\sqrt{w-1}$ has, at least potentially, an argument range of $(-\pi,\pi]$, which is impossible for *any* single-valued interpretation of $\sqrt{w^2-1}$.

(ii) Here he takes $\sqrt{w+1}$ as before, but $\sqrt{w-1}$ to be an alternative interpretation: $\left(\sqrt{\ },\mathbf{C},\mathbf{C}\equiv\mathbf{R}^{+}\mathop{\times}_{\text{polar}}[0,\pi)\cup\{0\}\right)_{\mathcal{B}}$.

In Bourbaki-speak, of course, $(f,D,D')_{\mathcal{B}}$ is a bijection (in either case), so $\left(f^{-1},D',D\right)_{\mathcal{B}}$ exists, and the question of whether there is a "formula" for it is not in the language.

Formalisations of these statements?

Of course, the first question is "which kind of statement are we trying to formalise".

This matters in two sense — which of the views are we trying to formalise, and are we trying to formalise just the statement, or the statement *and* its proof.

The question "which view" seems to be a hard one — when reading a text one often has few clues as to the author's intentions in this area. Nevertheless, let us suppose that the view is given.

The (Bourbakist) Theory

In this view a function is defined by its graph, there is no language of formulae, and the graph of the inverse of a bijective function is the transpose of the graph of the original.
Therefore the task of formalising any such statements is the general one of formalising (set-theoretic) mathematical texts.

The multivalued view

We use capital initial letters to denote the multivalued equivalents of the usual functions, so $\mathrm{Log}(z) = \{w : \exp(w) = z\}$.
Here, an expression like (3) becomes

$$\begin{aligned} &\forall w_3 \in \mathrm{Log}(z_1 z_2) \exists w_1 \in \mathrm{Log}(z_1), w_2 \in \mathrm{Log}(z_2) : w_3 = w_1 w_2 \wedge \\ &\forall w_1 \in \mathrm{Log}(z_1), w_2 \in \mathrm{Log}(z_2) \exists w_3 \in \mathrm{Log}(z_1 z_2) : w_3 = w_1 w_2, \end{aligned} \tag{10}$$

There is significant expansion here, and one might be tempted to write

$$\mathrm{Log}(z_1 z_2) = \mathrm{Log}(z_1) + \mathrm{Log}(z_2) \tag{11}$$

using set-theoretic addition and equality of sets, which looks reassuringly like a multivalued version of (3).

However, there are several caveats.
The correct generalisation of $\log(z^2) = 2\log(z)$ is

$$\mathrm{Log}(z^2) = \mathrm{Log}(z) + \mathrm{Log}(z) \tag{12}$$

(note that Convention 1 does *not* apply here, since we are considering sets of values, rather than merely underspecified values) and *not* $\mathrm{Log}(z^2) = 2\,\mathrm{Log}(z)$ (which, effectively, *would* apply the convention). Also not all such equations translate as easily: the multi-valued equivalent of

$$\arcsin(z) \stackrel{?}{=} \arctan\left(\frac{z}{\sqrt{1-z^2}}\right) \tag{13}$$

is in fact

$$\mathrm{Arcsin}(z) \cup \mathrm{Arcsin}(-z) = \mathrm{Arctan}\left(\frac{z}{\mathrm{Sqrt}(1-z^2)}\right). \tag{14}$$

Conclusion

Translating statements about these functions to the multivalued view is not as simple as it seems, and producing correct *translations can be difficult.*

Multivalued continued

It *might* be possible to define a rewrite with constant expansion (a "de Bruijn factor")by defining new functions such as $\mathcal{AS}(z) = \mathrm{Arcsin}(z) \cup \mathrm{Arcsin}(-z)$, but to the author's knowledge this has not been done, and would probably be a substantial research project.

It would be tempting to wonder about the difficulties of translating proofs, but, other than his and his colleagues', the author has only seen proofs which work by reduction modulo $2\pi i$, and therefore do not generalise to equations like (14), for which the author only knows his own (unpublished) proof.

As has been said, we see little hope for formalising the more general 'Riemann surfaces' version of this view.

The branch view

In this view, a function is defined locally, in a simply-connected open set, and the statements made informally in mathematics are true in this interpretation if they are true in the informal sense.
The first real problem comes in determining the side conditions, such as "not containing 0". For a *fixed* vocabulary of functions, such as the elementary functions (which can all be derived from exp and log) this can probably be achieved, but when new functions can be introduced, it becomes much harder.
The second problem is to determine what such a suitable open set is, and whether one can be found which is large enough for the mathematical setting envisaged. This is often equivalent to the problem of finding a suitable path, and the challenges are really those of formalising traditional analysis.

The table-maker's point of view: statements

For the elementary functions, there is an effective methodology, implicit in [AS] and made explicit in [Corlessetal2000].

1. Choose a branch cut $\mathcal{X}$ for log, and this defines the value of $\log(z)$ for $z \in \mathbf{C} \setminus \mathcal{X}$ by integration from $\log 1 = 0$.
2. Choose an *adherence rule* to define the value of log on $\mathcal{X}$.
3. For each other function f, choose an expression for $f(x)$ in terms of log. Several such choices, and none are perfect.
4. As a consequence of step 3, write down the various simplification rules that f (and other functions) must satisfy.
5. If one is unhappy with these results, return to step 3. **Do not** attempt to rewrite the rules — this leads to inconsistencies, with which tables have been (and alas continue to be) bothered over the years.

Conclusion

In the table-maker's view, statements *about multi-valued functions, if correct, are the same as usually stated. However, they may require amplification, as in (4") versus (4). At least naïvely, such expansion may be unbounded.*

The table-maker's point of view: proofs

Proofs are a trickier matter. As far as the author knows, such proofs were generally not published before the days of computer algebra, though the table-makers certainly had intuitive understandings of them, at least as regards real variables.

Conclusion

Producing (formal) proofs of such statements is a developing subject, even in the context of a computer algebra system. Converting them into an actual theorem prover is a major challenge. Unfortunately, cylindrical algebraic decomposition, as used here, does not seem to lend itself to being used as an 'oracle' by theorem provers.

Differential Algebra

The translation from *well-posed* statements of analysis into this viewpoint is comparatively easy. There are, however, two significant problems.

1. "Well-posed", in the context of differential algebra, means that every extension that purports to be transcendental really is, *and* introduces no new constants. Hence every simplification rule essentially reduces to "correct up to a constant", and beyond here differential algebra does not help us, as seen with $(4'')$.
2. There is no guarantee that the expressions produced by differential algebra, when interpreted as functions, will be well-behaved. [LazardRioboo1990]

The pragmatic view

The fundamental problem with the pragmatic view is that it is a hybrid, and the formalisms are different. Indeed, it is precisely this difference that causes most of the problems in practice.
The pragmatist takes formulae produced in the differential-algebraic viewpoint, and interprets them in the branch viewpoint. In the branch viewpoint, every integral is continuous, but there is no guarantee of this remaining true in the hybrid view unless special care is taken.

Conclusion

The pragmatist's view, while useful, is indeed a hybrid, and great care must be taken when translated from one viewpoint to the other.

Conclusions

We need to

1. Work out which view(s) we are adopting
   * Difficult in retrospect
2. Formalise the view(s)
   * Sometimes difficult
3. Formalise the translation between the views
   * *Always* difficult