

This document is a work in progress. As such it is incomplete and still has errors and omissions. When brought to a state where I cannot easily find any improvements it will form my next document on Complex analysis. Now it looks as if there are going to be so many ideas that I can't just finish it as a paper, it is instead a sort of discussion document.

A strange feature of this study is that as it develops sections get expanded with different material so the section headings get out of date, and it is not easy to get the ideas in the most sensible order and keep it that way. The structure is still obviously not right. Thus there are many places where there are forward references. Comments are welcome. Please send them to john.h.nixon1@gmail.com (see also <https://www.bluesky-home.co.uk> for my other papers and ideas)

Date: 2023-08-30

Towards a Theory of Analytic Functions

Abstract

Multivalued analytic functions (or relations) are defined as mappings of the Riemann Sphere to itself that satisfy the Cauchy-Riemann equations, and are not constrained by artificial boundaries or constraints on values. They are believed to be determined uniquely by their behaviours at all their singular and inversion points, which is a generalisation of a result of the previous study of the algebraic case. The behaviour at these points is determined by simple equations that only make sense in the context of multivalued functions and can describe behaviour near essential singular points as well as simple poles and branch points associated with algebraic functions. Many examples are discussed. It is suggested though not yet proved that the set of analytic functions forms a large algebraic structure that is closed under the operation of taking limits in addition to the operations that give closure to the set of algebraic functions.

The approach will be intuitive and non-technical showing how to handle multi-valued functions in calculations and the topological properties of the surfaces representing them.

Mathematics Subject Classification:

Keywords: analytic functions, complex analysis

1 Introduction

the key differences between this approach and the standard approach to analytic function is (1) Basing all the arguments on the closure of the complex

plane $\overline{\mathbb{C}}$ (the Riemann sphere) instead of the complex plane \mathbb{C} . (2) The different definition of singular points based on topology. (3) The treatment of mappings $\overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ as multivalued functions without restricting their domains.

My earlier work on algebraic functions when considered as multivalued functions $z \rightarrow w$ where z and w are in the Riemann Sphere $\mathbb{C} \cup \{\infty\} = \overline{\mathbb{C}}$ seemed to indicate that their topology determines them uniquely apart from a few parameters. The topology of something is all the properties of it that are not changed by any continuous stretching without breaking and has been described as rubber sheet geometry. More precisely, algebraic functions are determined by the behaviours of these functions at their singular points and their locations.

The main theme of this paper is to investigate how this extends to analytic functions $\overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ that can be multivalued. Many examples are studied then some general theory is developed. This is based on the topological concept of a singular point of a function $f()$, the connection between it and a simple equation satisfied by $f()$ that may be multivalued, and the idea of the *simplest solution* of such an equation. Dealing with multivalued functions and the equations they satisfy can be quite different from the usual right-unique case as the examples show.

The notion of a singular point is slightly changed from my earlier work: the very special function $f : z \rightarrow 1/z$ that motivated the introduction of the point ∞ described in [Nixon2013] (the Riemann Sphere) so as to make it left-unique as well as right-unique, is now not considered to have a singular point because of this. The point $(0, \infty)$ is now called an inversion point of $f()$.

Next follows a result that seems so fundamental that it should be perhaps mentioned here. It is connected with the completion of the complex plane to the Riemann Sphere $\overline{\mathbb{C}}$ and although I am probably not able to express or prove it properly, I present it as a theorem. Consider an analytic function $f : \overline{\mathbb{C}} \rightarrow \overline{\mathbb{C}}$ with a single singular point at z_0 and a circuit described in z that is close to z_0 in $\overline{\mathbb{C}}$. The image of this not a circuit in $f(z)$ that is described just once in the same direction. Suppose there are no other singular points in $f()$ then this can be continuously deformed past ∞ without changing the discrete topology to a small circuit in z at some other point z_1 with a corresponding image in $f(z)$ where again the result is not a circuit described just once in the same direction. This small circuit in z can be made as small as you like while not crossing any singular or inversion point with the same result, and this would imply a singular or inversion point at $z_1 \neq z_0$ of the corresponding type. This contradiction proves that

Theorem 1.1. *An analytic function defined on the Riemann Sphere $\overline{\mathbb{C}}$ cannot have only one singular or inversion point.*

Another important theme, though not yet fully developed, is that analytic functions form a very complex algebraic structure that extends the algebra

of algebraic functions by adding to it a single extra closure operation i.e. the passage to the limit of a sequence of such functions. If something like induction could be done it might provide another way to prove propositions.

The [Encyclopedia of Mathematics] says ‘Each analytic function is an “organically connected whole”, which represents a “unique” function throughout its natural domain of existence.’ and I think this is the approach that should be followed. The layout of the paper is as follows:

Notations and terminology

A description of the closure operations with examples

A motivating very simple example of the equation mentioned above

A look again at algebraic functions and characterising power functions

Examples of analytic functions and characterising their singular and inversion points

General theory singular points and *simplest solutions*

2 Relations in general

Analytic functions are in general multivalued (i.e. are relations) and therefore the general theory of relations must play a major role. Specifically, the concept of \rightarrow on analytic functions which could be read as “could start with” for example the function given by $\sin(z^6)$ could be defined starting from z^2 or from z^3 and then applying another analytic function. This has its origin in relations generally, and for this reason some of its basic properties needed to be established before applying them to analytic functions.

2.0.1 General notations and terminology for relations

Relations generalise the concept of a mapping or function to the multivalued case. For an arbitrary set S a relation on S is a subset of $S \times S$. It will be useful to collect a few results, terminology, and notations involving relations here. The logical symbols / (typed over what it applies to) or $\neg, \forall, \exists, \in, \wedge, \vee, \Rightarrow, \Leftrightarrow$ mean “not”, “for all”, “there exists”, “in”, “and”, “or”, “implies”, and “if and only if” respectively. The Boolean values 0 representing “false” and 1 representing “true” will be used throughout and the following equivalences occur frequently

$$\begin{aligned}
 A \vee B = 0 &\Leftrightarrow A = 0 \wedge B = 0 \\
 A \wedge B = 0 &\Leftrightarrow A = 0 \vee B = 0 \\
 A \vee B = 1 &\Leftrightarrow A = 1 \vee B = 1 \\
 A \wedge B = 1 &\Leftrightarrow A = 1 \wedge B = 1
 \end{aligned}
 \tag{1}$$

for any Boolean variables A and B .

A relation R is left-total if $\forall b \in S \{\exists a \in S[aRb]\}$ and likewise right-total if $\forall a \in S \{\exists b \in S[aRb]\}$. R is left-unique if $\forall b \in S[a_1Rb \wedge a_2Rb \Rightarrow a_1 = a_2]$ and likewise right-unique if $\forall a \in S[aRb_1 \wedge aRb_2 \Rightarrow b_1 = b_2]$. These meaning of these four terms seem to me to be immediately clear.

(In earlier versions of this manuscript I used the terms “one-to-one” instead of “left-unique” and “single-valued” use instead “right-unique”. The terms “left-total” and “right-total” will be used instead of “serial”, and “surjective” or “onto” respectively if needed.)

Apart from the operations of Boolean algebra that apply to all sets, extensive use will be made of composition (denoted by juxtaposition) and inversion. The composition of R_1 with R_2 , R_1R_2 is defined by

$$\forall a, b \in S[aR_1R_2b \Leftrightarrow \exists c \in S[aR_1c \wedge cR_2b]]. \quad (2)$$

Note this is the logical order and corresponds (in the case of functions) to do R_1 then do R_2 i.e. $R_2(R_1(x))$. Then it follows that composition is associative and can be used to define the n th compositional power of a relation R for non-negative integers by $R^0 = I$ (the identity relation I is defined by $\forall a, b \in S[aIb \Leftrightarrow a = b]$), $R^1 = R$, and $R^{n+1} = R^nR$ and extend it to negative integers to give R^{-n} which is defined to be $(R^{-1})^n = (R^n)^{-1}$ which is also easily shown where the inverse of R written as R^{-1} is defined by $\forall a, b \in S[aRb \Leftrightarrow bR^{-1}a]$.

The empty relation where $R = \emptyset$ satisfying $\forall a, b \in S[\neg aRb]$ and similarly $\forall a, b \in S[aRb]$ defines the relation 1 and the negation symbol will also be applied to relations giving their complement so that $R \cup \neg R = 1$ and $R \cap \neg R = \emptyset$. The following properties hold $(R_1 \cup R_2)R_3 = R_1R_3 \cup R_2R_3$. $(R^{-1})^{-1} = R$, $(R_1R_2)^{-1} = R_2^{-1}R_1^{-1}$. There is a general relation on relations denoted by \rightarrow , that has a simple definition based on composition, $R_1 \rightarrow R_2 \Leftrightarrow \exists R_3[R_1 = R_2R_3]$ and could be read as “is more or equally complex as”, which will probably not be clear until analytic functions (that actually are relations) are discussed. Its inverse could be denoted by \leftarrow means “is simpler or equally complex as”. Any relation R satisfies

$$R \rightarrow I \quad (3)$$

even if R is the empty relation \emptyset . It is clearly reflexive ($R \rightarrow R$) and transitive i.e. $(R_1 \rightarrow R_2) \wedge (R_2 \rightarrow R_3) \Rightarrow (R_1 \rightarrow R_3)$.

2.0.2 Roots and rooted sets of relations

Suppose a set T of relations on S is such that if $R_1 \rightarrow R$ and $R \in T$ then $R_1 \in T$. Then T will be defined as a rooted set of relations. This is true if $T = \emptyset$ because both sides of the implication are false, which is true. Also $root(T) \subseteq T$ is defined such that no member R of $root(T)$ is such that there

is a member $R_1 \in T$ such that $R \rightarrow R_1$ and $R \neq R_1$ unless $R_1 \rightarrow R$. This is expressed as

$$R \in \text{root}(T) \Leftrightarrow R \in T \wedge [\nexists R_1 \in T ((R \rightarrow R_1) \wedge (R \neq R_1) \wedge (R_1 \nrightarrow R))]. \quad (4)$$

The set $\text{root}(T)$ could be approached by starting from T and removing relations R satisfying $\exists R_1 \in T ((R \rightarrow R_1) \wedge (R \neq R_1) \wedge (R_1 \nrightarrow R))$ until no more can be removed which will in principle give $\text{root}(T)$. There will be a chain of relations R starting from each member of T , each satisfying \rightarrow to the next one that can only end if the final R cannot be removed according to this criterion and is therefore in $\text{root}(T)$. Because \rightarrow is transitive, any relation in T is related by \rightarrow to a relation in the root, i.e.

$$\forall R_1 \in T \{ \exists R \in \text{root}(T) [R_1 \rightarrow R] \}. \quad (5)$$

Equation (4) can be written as

$$R \in \text{root}(T) \Leftrightarrow R \in T \wedge \forall R_1 \in T ((R \nrightarrow R_1) \vee (R_1 \rightarrow R)). \quad (6)$$

because $R_1 = R \Rightarrow R_1 \rightarrow R$. This can also be written as

$$R \in \text{root}(T) \Leftrightarrow R \in T \wedge \forall R_1 \in T [R \rightarrow R_1 \Rightarrow R_1 \rightarrow R]. \quad (7)$$

2.0.3 Some other properties of roots and rooted sets of relations

From this it is clear that $\text{root}(T)$ could consist of any number of disconnected parts not related to each other by \rightarrow such that within each part every relation is related to every other one by \rightarrow and an important property of $\text{root}(T)$ is the number of such parts denoted by $\#(\text{root}(T))$. A singly-rooted set T is defined as a rooted set T such that $\#(\text{root}(T)) = 1$. Another way to state this is to say that the relation \leftrightarrow defined by $(R_1 \leftrightarrow R_2) \Leftrightarrow (R_1 \rightarrow R_2) \wedge (R_2 \rightarrow R_1)$ is an equivalence relation and the set $\text{root}(T)$ in general consists of $\#(\text{root}(T))$ equivalence classes. It is obvious from this that in general a rooted set is the union of the singly rooted sets corresponding to each of these equivalence classes. [Note these concepts could be applied in the context of any relation in place of \rightarrow , if it is extended by adding related pairs to the minimal extent possible to make it reflexive and transitive, but this will probably not be needed here.]

Let T_1 and T_2 be rooted sets then T_1 and T_2 satisfy the following that holds for any rooted set T : $\forall R, R_1 [((R \in T) \wedge (R_1 \rightarrow R)) \Rightarrow R_1 \in T]$. Then by combining these with \wedge it follows that, whatever the relations R and R_1 are, $R \in (T_1 \cap T_2) \wedge (R_1 \rightarrow R) \Rightarrow R_1 \in (T_1 \cap T_2)$ i.e. the intersection of two rooted sets is a rooted set. Likewise $R \in (T_1 \cup T_2) \wedge (R_1 \rightarrow R) \Rightarrow R_1 \in (T_1 \cup T_2)$ and so $T_1 \cup T_2$ is also a rooted set. This however tells us nothing about the the root sets involved.

Now suppose that T_1 and T_2 are singly rooted so that there exist relations, say R_1 and R_2 such that $R_1 \rightarrow R_2$ and $R_2 \rightarrow R_1$ and act as roots for T_1 and T_2 respectively, then $\forall R \in T_1(R \rightarrow R_1)$ and $\forall R \in T_2(R \rightarrow R_2)$. Then the question is whether the intersection of these is empty or not, and if not, is it singly or multiply rooted?

***** needed? Applying (7) to the case $T = T_1 \cap T_2$ gives

$$R \in \text{root}(T_1 \cap T_2) \Leftrightarrow R \in (T_1 \cap T_2) \wedge \forall R' [R' \in (T_1 \cap T_2) \Rightarrow ((R \rightarrow R') \Rightarrow (R' \rightarrow R))] \quad (8)$$

and this last predicate will be denoted by $P(R)$ which can be rewritten as $P(R) = \forall R' [(R' \rightarrow R_1) \wedge (R' \rightarrow R_2) \wedge (R \rightarrow R') \Rightarrow (R' \rightarrow R)]$. Then the question above reduces to whether or not (8) has any non-equivalent solutions for R i.e. whether or not

$$\forall R_3 \forall R_4 [(P(R_3) \wedge P(R_4)) \Rightarrow R_3 \leftrightarrow R_4], \quad (9)$$

and in particular $T_1 \cap T_2$ is singly-rooted if and only if (9) holds. From

$$\forall R_3 \forall R_4 [(P(R_3) \wedge P(R_4)) \Rightarrow R_3 \rightarrow R_4] \quad (10)$$

R_3 and R_4 can be exchanged (being dummy variables) and the expression rearranged again to give

$$\forall R_3 \forall R_4 [P(R_3) \wedge P(R_4) \Rightarrow R_3 \leftarrow R_4] \quad (11)$$

which when taken together are equivalent to (9), therefore (9) is equivalent to (10). to here*****

There is a related concept which is the root R_3 of the smallest rooted set that includes T_1 rooted by R_1 and T_2 rooted by R_2 . If $R_1 \rightarrow R_2$ the result is just $R_3 = R_2$ or any relation R_3 satisfying $R_3 \leftrightarrow R_2$ i.e. any relation equivalent to R_2 . Therefore it will be henceforth assumed that $R_1 \rightarrow R_2$ and by symmetry $R_2 \rightarrow R_1$. Returning to the main idea, R_3 satisfies

$$(R_1 \rightarrow R_3) \wedge (R_2 \rightarrow R_3) \wedge \nexists R_4 [(R_1 \rightarrow R_4) \wedge (R_2 \rightarrow R_4) \wedge (R_4 \rightarrow R_3) \wedge (R_3 \rightarrow R_4)]. \quad (12)$$

Such an R_3 of course satisfies

$$(R_1 \rightarrow R_3) \wedge (R_2 \rightarrow R_3) \quad (13)$$

and there are always solutions for R_3 to (13) because from (3) I is always a solution. Expression (12) is the defining condition for an extreme solution for R_3 of (13) because (13) must be satisfied and there must not exist any solution for R_4 in (12) that satisfies an extra condition (other than (13)) in relation to R_3 which is $(R_4 \rightarrow R_3) \wedge (R_3 \rightarrow R_4)$. This suggests approaching a solution to

(12) by starting from $R_3 = I$ and repeatedly asking the question for your R_3 whether or not

$$\exists R_4[(R_1 \rightarrow R_4) \wedge (R_2 \rightarrow R_4) \wedge (R_4 \rightarrow R_3) \wedge (R_3 \nrightarrow R_4)] \quad (14)$$

is true. If (14) is true, replace R_3 any of the set of values of R_4 that satisfy (14) and repeat, and if not, the current value of R_3 satisfies (12). This process must terminate (at least when the basic set S on which all the relations are defined is finite) and if not a sequence that converges in some sense is expected, leading to a solution of (12) because

$$\left. \begin{array}{l} R_1 \\ R_2 \end{array} \right\} \rightarrow R_n \rightarrow R_{n-1} \rightarrow \dots R_4 \rightarrow R_3 \rightarrow I \quad (15)$$

and

$$\left. \begin{array}{l} R_1 \\ R_2 \end{array} \right\} \leftarrow R_n \leftarrow R_{n-1} \leftarrow \dots R_4 \leftarrow R_3 \leftarrow I. \quad (16)$$

In expression (16) it cannot be the case that $R_n \rightarrow R_1$ because then $R_2 \rightarrow R_n \rightarrow R_1$ so $R_2 \rightarrow R_1$ which has already been ruled out. Likewise $R_n \nrightarrow R_2$. Here it has been assumed that the solution sought is R_n . Therefore the existence of solutions for R_3 of (12) will be assumed which will be approached by the sequence

$$I, R_3, R_4, \dots R_n \dots \quad (17)$$

where in expressions (15),(16), and (17), R_3 refers to the original value of R_3 , the next ones being R_4, R_5 etc..

Here two quite different arguments show that, for a relation amongst relations, \rightarrow only satisfying reflexivity and transitivity, it is not in general true that either of (1) the intersection of two rooted sets, and (2) the smallest rooted set containing the two rooted sets, are rooted by a single relation up to equivalence. These propositions may be true in general if extra properties are given to \rightarrow deriving from its relationship with composition of relations.

It is not true that in general

$$\exists R_1[(R_1 \rightarrow R_2) \wedge (R_1 \rightarrow R_3)] \Rightarrow (R_2 \rightarrow R_3) \vee (R_3 \rightarrow R_2) \quad (18)$$

which is suggested by composition of relations and reading \rightarrow as “starts with”, but the expression on the left can be written as $\exists P_1 \exists P_2 [R_2 P_1 = R_3 P_2]$ and is trivially true because it is satisfied by $P_1 = P_2 = \emptyset$. Note that equality of relations is the same as logical equivalence often written as \Leftrightarrow .

In order to answer other questions the general technique was used an example of which is the factorisation of the relation 1 i.e. $\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ on a set S with 2 elements and matrix notation was used for relations with aRb represented

by the entry in row a and column b where 0 means false and 1 means true. For convenience the conjunction \wedge will be abbreviated by juxtaposition (as is composition of relations) in the following because no confusion can result. The following is the exhaustive search giving the tree of possibilities for the 8 boolean variables such that the result is the relation 1. Starting from

$$1 = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae \vee bg & af \vee bh \\ ce \vee dg & cf \vee dh \end{bmatrix} \quad (19)$$

gives following where all the conditions on the Boolean variables a to h are given above the arrows:

$$\left\{ \begin{array}{l} \xrightarrow{a=0} \begin{bmatrix} bg & bh \\ ce \vee dg & cf \vee dh \end{bmatrix} \xrightarrow{b=1} \begin{bmatrix} g & h \\ ce \vee dg & cf \vee dh \end{bmatrix} \xrightarrow{g=h=1} \begin{bmatrix} 1 & 1 \\ ce \vee d & cf \vee d \end{bmatrix} \rightarrow \\ \left\{ \begin{array}{l} \xrightarrow{d=0} \begin{bmatrix} 1 & 1 \\ ce & cf \end{bmatrix} \xrightarrow{c=e=f=1} 1 \\ \xrightarrow{d=1} 1 \end{array} \right. \\ \xrightarrow{a=1} \begin{bmatrix} e \vee bg & f \vee bh \\ ce \vee dg & cf \vee dh \end{bmatrix} \rightarrow \left\{ \begin{array}{l} \xrightarrow{e=0} \begin{bmatrix} bg & f \vee bh \\ dg & cf \vee dh \end{bmatrix} \xrightarrow{b=d=g=1} \begin{bmatrix} 1 & f \vee h \\ 1 & cf \vee h \end{bmatrix} \rightarrow \\ \left\{ \begin{array}{l} \xrightarrow{h=0} \begin{bmatrix} 1 & f \\ 1 & cf \end{bmatrix} \xrightarrow{c=f=1} 1 \\ \xrightarrow{h=1} 1 \end{array} \right. \\ \xrightarrow{e=1} \begin{bmatrix} 1 & f \vee bh \\ c \vee dg & cf \vee dh \end{bmatrix} \rightarrow \\ \left\{ \begin{array}{l} \xrightarrow{f=0} \begin{bmatrix} 1 & bh \\ c \vee dg & dh \end{bmatrix} \xrightarrow{b=d=h=1} \begin{bmatrix} c \vee g \\ \rightarrow \end{bmatrix} 1 \\ \xrightarrow{f=1} \begin{bmatrix} 1 & 1 \\ c \vee dg & c \vee dh \end{bmatrix} \rightarrow \\ \left\{ \begin{array}{l} \xrightarrow{c=0} \begin{bmatrix} 1 & 1 \\ dg & dh \end{bmatrix} \xrightarrow{d=g=h=1} 1 \\ \xrightarrow{c=1} 1 \end{array} \right. \end{array} \right. \end{array} \right. \end{array} \right. \quad (20)$$

This results in all the combinations of $\{a, b, c, d, e, f, h\}$ that lead to (19) being satisfied. Here the only interest is in all values $\{a, b, c, d\}$ so duplications can occur and result is the following 11 relations (9 distinct ones) X that satisfy

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow X$ with blanks indicating either 0 or 1 at that point in the array:

$$\begin{aligned} & \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right] \left[\begin{array}{cc} 0 & 1 \\ & 1 \end{array} \right] \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right] \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right] \left[\begin{array}{cc} 1 & \\ 0 & 1 \end{array} \right] \left[\begin{array}{c} 1 \\ 1 \end{array} \right]. \text{ i.e.} \\ & \left\{ \begin{array}{l} \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ & 1 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 1 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right] \\ \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right] \end{array} \right\} \end{aligned} \quad (21)$$

Similarly

$$\left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right] \rightarrow \left\{ \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right] \right\} \quad (22)$$

Similarly by swapping rows and columns of the matrices of the relations, corresponding to renaming the elements of the set S on which the relations are defined gives

$$\left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right] \rightarrow \left\{ \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right] \right\}. \quad (23)$$

Similarly

$$\left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right] \rightarrow \left\{ \left[\begin{array}{cc} 0 & 1 \\ 1 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 1 & 1 \end{array} \right] \right\} \quad (24)$$

and

$$\left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right] \rightarrow \left\{ \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right] \right\} \quad (25)$$

These results suggest that columns and rows can independently be swapped on both sides of \rightarrow relationships. To show this note that $R \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = R'$ is R

with its columns swapped, and similarly $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R = R''$ is R with its rows

swapped and $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ which is the identity relation I .

So $R = AX$ can be written as $R' = A \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} X \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ i.e. $R' = A'Y$ where Y is X with its rows and columns swapped. This shows $R \rightarrow A$

implies $R' \rightarrow A'$. Similarly $R = AX$ implies $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} R = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} AX$ i.e. $R'' = A''X$, which shows that $R \rightarrow A$ implies $R'' \rightarrow A''$.

Continuing with the characterisation of \rightarrow relations on a set S where $\#(S) = 2$ gives

$$\left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array} \right] \rightarrow \left\{ \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 1 & 0 \end{array} \right], \left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 0 & 0 \end{array} \right], \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right], \left[\begin{array}{cc} 1 & 1 \\ 0 & 1 \end{array} \right] \right\} \quad (26)$$

together with the row swapped version. Also

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \rightarrow \left\{ \begin{array}{l} \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{array} \right\} \quad (27)$$

together with its column swapped version.

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rightarrow \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \quad (28)$$

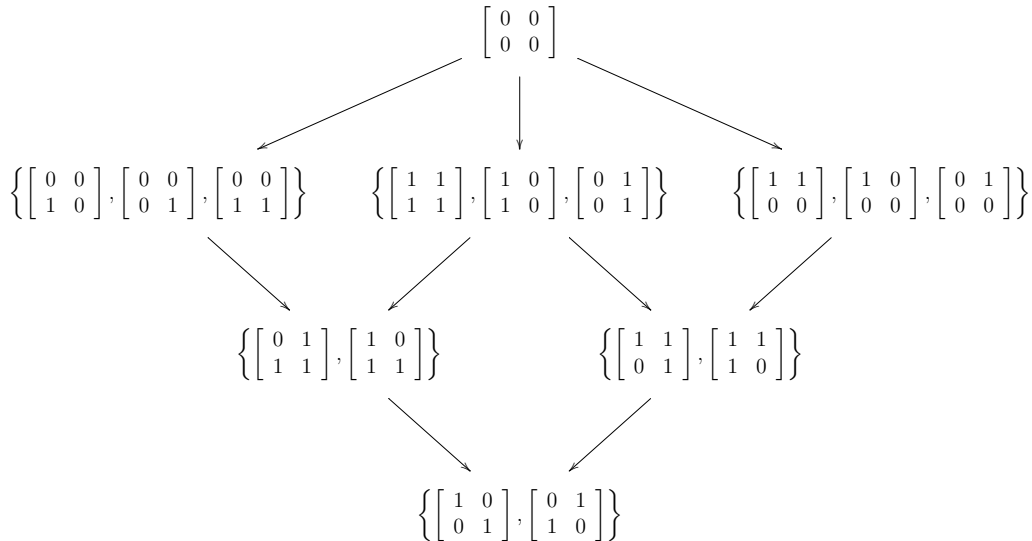
$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \rightarrow \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right\} \quad (29)$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow \left\{ \begin{array}{l} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \\ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \end{array} \right\} \quad (30)$$

together with its row swapped and column swapped versions i.e. 4 results and

$$\emptyset = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \rightarrow \text{every relation on } S \quad (31)$$

because the other relation can always be chosen as \emptyset . This completes the characterisation of \rightarrow where $\#(S) = 2$. These results can be gathered together, omitting \rightarrow relations that follow by reflexivity and transitivity from the given ones as follows:



2.1 The general case where $\#(S) = n$

2.1.1 Alternative descriptions of composition of relations in matrix form and an important theorem

This section involves other ways of describing the equation

$$A = BC \quad (32)$$

where A, B and C are relations on a set S of n elements. The purpose of this is to make it easier to do the computations.

In components (32) is

$$A_{ij} = \exists k [B_{ik}C_{kj}] \quad (33)$$

where the range of all free indices (i.e. ones without being explicitly quantified over by an \exists or \forall) in all equations is by default the set S i.e. $\forall \text{ index} \in S$ that can be taken as $\{1, 2, \dots, n\}$ should ideally be added around these statements, but this is often left implicit in many parts of mathematics.

The simplest case of (32) is when B or C is a permutation. Suppose first B is a permutation i.e. $B_{ij} = \begin{cases} 0 & p(i) \neq j \\ 1 & p(i) = j \end{cases}$ where p is a permutation on the integers from 1 to n , then in (33) for each i, k is unique such that B_{ik} is true, so this expression $(BC)_{ij}$ simplifies to $C_{p(i),j}$ i.e. row i of BC is the same as row $p(i)$ of C . Now suppose that B_{ij} has more than one 1 in its i th row. Each of these is a value of k for which B_{ik} is true therefore the result of the i th row of (33) is the \vee combination of each of these results i.e. row i of BC is the \vee combination of rows of C with those rows determined by where the i th row of B has the value 1. In this way the matrix for B gives the \vee combinations of the rows of C that appear in the rows of A . All of this can of course be argued with the columns and rows reversed and shows how the columns of C determine the \vee combinations of the columns of B that appear in the columns of A , specifically the j th column of A is formed from the columns of B with the j th column of C indicating which of the columns of B to be included i.e. those where $C_{kj} = 1$. These facts make it surprisingly quick to carry out the computations and are related to the concept of $*$ in the next section.

Every one of the n row vectors of $AB = I_n$ has a single 1 in a different location and is a \vee combination of a subset of the n row vectors of B . Therefore all these vectors that contribute to that row vector of AB must have a single 1 or none each and must in fact be just that row vector itself. Thus the row vectors of AB must be the same set as the row vectors of B taken in some order in the matrix for B .

The order in which these row vectors of B are taken and the \vee combinations involved are determined by the rows of A which also have just one 1 in each row because only one row vector of B is involved in each case. This proves

that all the rows of A and B have just one 1 each. Because these two sets each have no repetitions, the columns of A and B also have these properties and the relations A and B are permutations that are inverse to each other i.e. they are both left-total and right-total, and left-unique and right-unique.

Theorem 2.1. *If the basic set S on which relations defined is finite with n elements, $AB = I_n$ implies A and B are inverses to each other i.e. $A = B^{-1}$, and A and B are permutations on S .*

2.1.2 The operations: obtaining a basis (b) and closure under \vee (*)

For dealing with matrices representing relations, one has to first deal with vectors each component of which is a Boolean variable (having values 0 and 1 representing “false” and “true” respectively). Let U_n denote the set of n vectors that are all 0 except for one component which is 1. When constructing the composition of relations, these Boolean vectors are combined componentwise with the “or” i.e. \vee operation which allows any Boolean vector of length n to be obtained, the set of which is denoted by U_n^* . For any subset of a set A of members of U_n^* , they can be combined with \vee which is commutative and associative. Therefore just this subset of A involved needs to be specified to determine the result. The operation $*$ in general is defined such that the set of all such distinct results for different subsets of A is denoted by A^* . This gives an augmented set of Boolean vectors that includes the original set i.e. $A^* \supseteq A$ because each singleton set from A i.e. having only a single member generates just that single Boolean vector and 0. A^* can be defined as starting with just all the Boolean vectors in A , and is closed under the “or” or \vee operation of Boolean vectors. (It could be thought of as analogous to a vector space generated by a set of vectors that it contains.) Under $*$, the empty set will give rise to only the Boolean vector 0 that is all 0’s, so $0 \in A^*$.

Any set of distinct Boolean vectors A (i.e. for any $A \in 2^{U_n^*}$) can be associated with a tree, which is constructed by associating nodes with Boolean vectors such that a node s is connected to a set of nodes at a lower level, if and only if the corresponding vector to s in A is the \vee combination of the corresponding vectors in A to the nodes at the lower level. All such \vee combinations of the vectors in A that are also in A must be found and included in the tree. The operation b is defined such that $b(A)$ is the set of all the Boolean vectors at the lowest level of the tree generated from A . This can be stated alternatively by saying that $b(A)$ is the unique subset of A such that every member of $b(A)$ cannot be expressed as the \vee combination of more than one distinct members of A .

The binary operation \vee can be naturally extended to any pair of sets of Boolean vectors A and B to generate the set of Boolean vectors $A \vee B$ such

that each one is a \vee combination of a pair of vectors one from each set:

$$\forall A, B \in 2^{U_n} [c \in A \vee B \Leftrightarrow \exists a \in A [\exists b \in B [c = a \vee b]]]. \quad (34)$$

2.1.3 Some properties of $*$ and b

As just stated,

$$A = B \Rightarrow b(A) = b(B). \quad (35)$$

It was noted above that

$$0 \in A^*. \quad (36)$$

Can the number of members of b be greater than n ? From above it is also immediate that for any sets of Boolean vectors A and B ,

$$b(A) \subseteq A \subseteq A^* \text{ and} \quad (37)$$

$$(A^*)^* = A^* \text{ and } b(b(A)) = b(A). \quad (38)$$

It is also obvious that

$$A \subseteq B \Rightarrow A^* \subseteq B^*. \quad (39)$$

Also

$$b(A) \subseteq b(B) \Rightarrow A^* \subseteq B^* \quad (40)$$

is almost obvious. To prove this, just add some vectors to $b(A)$ to get $b(B)$ for some B , then any vector which is a \vee combination of $b(A)$ i.e. is in A^* must also be a \vee combination of $b(B)$ in which the extra vectors play no part. Another pair of results is

$$b(A^*) = b(A) \text{ and } (b(A))^* = A^* \quad (41)$$

Proof. The first part is immediate. It is also immediate that $(b(A))^* \supseteq A^*$ because adding vectors to $b(A)$ that are in A cannot generate any new vectors in $e()$ because they are already known to be \vee combinations of $b(A)$ and from (37) and (39) $(b(A))^* \subseteq A^*$. \square

The remaining 4 results are negative but were included because it could be thought naively that these are true as well: $A^* \subseteq B^* \not\Rightarrow b(A) \subseteq b(B)$.

A counter example to prove this is $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

Similarly it is easy to see that $A \subseteq B \not\Rightarrow b(A) \subseteq b(B)$. Also

$$b(A) \subseteq b(B) \not\Rightarrow A \subseteq B. \quad (42)$$

For example if $A = \{0, 1, 2, 1 \vee 2\}$ and $B = \{1, 2, 3, 2 \vee 3, 1 \vee 2 \vee 3\}$ where $1, 2, 3$ are distinct vectors then $b(A) = \{1, 2\}, b(B) = \{1, 2, 3\}$, $A^* = A$ and $B^* = \{0, 1, 2, 3, 1 \vee 2, 1 \vee 3, 2 \vee 3, 1 \vee 2 \vee 3\}$. This example also serves to prove that $A^* \subseteq B^* \not\Rightarrow A \subseteq B$ because $1 \vee 2 \in A$ but $1 \vee 2 \notin B$. The 6 similar results can be collected together in the following diagram

$$\begin{array}{ccc}
 A \subseteq B & \xrightarrow[\not\Rightarrow]{\not\Leftarrow} & b(A) \subseteq b(B) \\
 \searrow[\Rightarrow] & & \swarrow[\Leftarrow] \\
 & & A^* \subseteq B^* \\
 \swarrow[\not\Leftarrow] & & \searrow[\not\Rightarrow]
 \end{array}
 \tag{43}$$

The intersection $A = R_1^* \cap R_2^*$ is closed under \vee i.e. $(X \in A) \wedge (Y \in A) \Rightarrow X \vee Y \in A$ because R_1^* and R_2^* are also closed under \vee . Suppose $R \in T_1 \Leftrightarrow R \rightarrow R_1$ and $R \in T_2 \Leftrightarrow R \rightarrow R_2$. Then

$$R \in T_1 \cap T_2 \Leftrightarrow (R^* \subseteq R_1^*) \wedge (R^* \subseteq R_2^*) \Leftrightarrow R^* \subseteq R_1^* \cap R_2^* \tag{44}$$

which is a set closed under \vee which is B^* for some set of vectors B because B could be $R_1^* \cap R_2^*$ itself for example, or a smaller set and $b(B)$ can be found uniquely from this. Then this can be used to define a binary operation on the sets of Boolean vectors which in its simplest form is

Definition 2.2. $A \oplus B = b(A^* \cap B^*)$.

From this it is easy to show that \oplus is both commutative and associative.

2.1.4 Characterising \rightarrow in terms of the matrix representation of relations

Closely related to this result is the situation when $T_1 \rightarrow T$ only. This is illustrated in the following diagram where the symbol \vee on an arrow indicates that the node at the head of the arrow is formed from the contents of the node at the tail of the arrow by combining some elements with the \vee operation. Note here the unfortunate notation in which $T_1 \rightarrow T$ is equivalent to $T \xrightarrow{\vee} T_1$. In the following diagram the annotation on the right refers to the values of b .

$$\begin{array}{ccc}
 T & \xleftarrow{\vee} & b(T) & \text{Not minimal for } T_1 \\
 \vee \downarrow & & \downarrow \vee & \\
 T_1 & \xleftarrow{\vee} & b(T_1) & \text{minimal for } T_1
 \end{array}$$

From this it is clear that no member of $b(T_1)$ can be outside $(b(T))^*$ justifying

that all members of $b(T_1)$ are \vee combinations of members of $b(T)$ as indicated. That is every member of $b(T_1)$ is in $(b(T))^*$ and so $(b(T_1))^* \subseteq (b(T))^*$ because the result of $*$ is closed under \vee . Conversely $(b(T_1))^* \subseteq (b(T))^*$ implies $T_1 \subseteq (b(T))^*$ because $T_1 \subseteq (b(T_1))^*$ i.e. every member of T_1 is a \vee combination of a special subset of T i.e. $T_1 = TB$ for some relation B i.e. $T_1 \rightarrow T$. Combining this with (41).2 proves that

Theorem 2.3. $T_1 \rightarrow T$ if and only if $T_1^* \subseteq T^*$.

By combining this with its converse shows that $T_1 \leftrightarrow T \Leftrightarrow T_1^* = T_2^* \Leftrightarrow b(T_1^*) = b(T_2^*) \Leftrightarrow b(T_1) = b(T_2)$ using (35) and (41) i.e. the following theorem is proved.

Theorem 2.4.

$$T_1 \leftrightarrow T \Leftrightarrow b(T_1) = b(T). \quad (45)$$

***** this part may not be strictly needed to page 17 but might be useful
Returning to the equations

$$T_1 = TA \text{ and } T = T_1B \quad (46)$$

where A and B are arbitrary relations, that determine the relations on relations $T_1 \rightarrow T$ and $T \rightarrow T_1$ respectively, it is desired to characterise these relations (singly and together) in terms of the matrix representations of T and T_1 .

In general for relations on a set S with $\#(S) = n$, the relation on relations \rightarrow is defined by $R \rightarrow A \Leftrightarrow \exists X[R = AX]$ for some relation X on S where the composition AX of relations A and X has components in matrix notation given by $(AX)_{ij} = \exists k(A_{ik}X_{kj})$

Now it is clear that a necessary condition for the vectors of T to be recoverable from the vectors of T_1 by the transformation (46).2 is that the set of unique vectors at the lowest level of the tree for the column vectors of T , $f(T)$, be present in T_1 because if they are not all present, they cannot be reconstructed from the columns of T_1 . On the other hand if they are all present, they can be put in their correct places by an appropriate B and kept uncombined (apart from with a zero vector) together with all the other nodes that are known \vee combinations of the members of $f(T)$.

Similarly $(RP)_{ij} = \exists k[R_{ik}P_{kj}]$. Here k has to be $p^{-1}(j)$ otherwise P_{kj} is false. Therefore $(RP)_{ij} = R_{i,p^{-1}(j)}$. This shows that the relation RP is R with its columns permuted. $R \rightarrow A \Leftrightarrow \exists X[R = AX] \Leftrightarrow \exists X[PR = (PA)X] \Leftrightarrow PR \rightarrow PA$ i.e. R with its rows permuted $\rightarrow A$ with its rows permuted in the same way. Likewise $R \rightarrow A$ is equivalent to $\exists X[RP = AXP = AP(P^{-1}XP)] \Leftrightarrow \exists(P^{-1}XP)[RP = AP(P^{-1}XP)] \Leftrightarrow RP \rightarrow AP$ i.e. R with its columns permuted $\rightarrow A$ with its columns permuted in the same way, which are straightforward generalisations of the results above for the case $\#(S) = 2$.

Let T_1 be an ordered set of n Boolean vectors of length n , i.e. their members are taken in a particular order and repetition is allowed. Then the relation corresponding to T_1 (obtained by stacking these as column vectors to form a matrix) satisfies $T_1 = TA$ where column j of A is all zeros except 1 at row $k = p(j)$ if the k th column of T (obtained similarly from all the distinct Boolean vectors in T_1) appears at column j of T_1 for $1 \leq j \leq n$. If there are some repeated columns in T_1 , the number of columns of T used will be less than n and $p()$ will have some repeated values. T is then padded out to make it an $n \times n$ array with columns (index k) that are all zero. In symbols

$$T_{1ij} = \exists k [T_{ik} A_{kj}] = \exists k [T_{ik} \delta_{k,p(j)}] = T_{i,p(j)} \exists k (p(j) = k) \quad (47)$$

where the Kronecker delta has been used in the obvious way to represent Boolean values i.e.

$$\delta_{ij} = \begin{cases} i = j & \text{true} \\ i \neq j & \text{false} \end{cases} \quad (48)$$

In this way the unique columns of T_1 are put into T which is made square by padding it out at the right with columns that are zero and from $T_1 = TA$ it follows that $T_1 A^{-1} = T A A^{-1}$.

Suppose first that all the columns of T_1 are distinct and $p()$ is a permutation on $\{1, 2, \dots, n\}$ then

$$(A A^{-1})_{ik} = \exists j [A_{ij} A^{-1}_{jk}] = \exists j [\delta_{i,p(j)} \delta_{k,p(j)}] = \delta_{ik} \quad (49)$$

Therefore

$$\exists j [T_{1ij} A^{-1}_{jk}] = \exists j [T_{ij} \delta_{jk}] = T_{ik} \quad (50)$$

i.e. $T_1 A^{-1} = T$ thus both $T_1 \rightarrow T$ and $T \rightarrow T_1$. Also from $T_1 = TA$ it follows that $T_1 B = TAB$, so if B can be found such that

$$AB = I \text{ i.e. } \exists j [A_{ij} B_{jk}] = \delta_{ik} \quad (51)$$

then $T_1 B = T$ and

$$T_1 \rightarrow T \Rightarrow T \rightarrow T_1. \quad (52)$$

Consider the case where A has a column with more than one value equal to 1 so suppose column l has values

$$A_{i_1,l} = A_{i_2,l} = 1 \text{ with } i_1 \neq i_2. \quad (53)$$

From (51) it follows that $\exists j \neq l [A_{ij} B_{jk}] \vee A_{i,l} B_{lk} = \delta_{ik}$ so using (53) $\exists j \neq l [A_{i_1,j} B_{jk}] \vee B_{lk} = \delta_{i_1,k}$ (implying for $k \neq i_1$ $B_{lk} = 0$) and $\exists j \neq l [A_{i_2,j} B_{jk}] \vee B_{lk} = \delta_{i_2,k}$ (implying for $k \neq i_2$ $B_{lk} = 0$). Therefore $B_{lk} = 0$ for all $k \in \{1, 2, \dots, n\}$ and

$$\exists j \neq l [A_{ij} B_{jk}] = \delta_{ik}. \quad (54)$$

Consider the $n - 1$ vectors B_{jk} with $j \neq l$. Some combinations of these with \vee is row 1 given by δ_{1k} therefore one of these vectors is δ_{1k} therefore the rows involved in this can only be δ_{1k} or 0 and any other vector will have a 1 somewhere else therefore giving another 1 in the result. This argument works for any i in $\{1, 2, \dots, n\}$ therefore all n vectors δ_{ik} must be in the set. But there are only $n - 1$ of these. Therefore in this case no solution B of (51) is possible.

Theorem 2.5. $T \rightarrow T_1$ and $T_1 \rightarrow T$ (which is written more briefly as $T \leftrightarrow T_1$) if and only if $f(T) = f(T_1)$ where $f(M)$ is the minimal set of unique column vectors in the the matrix representation of the relation M that allows all the other columns of M to be constructed using \vee .

An example of this construction will make this clear. Suppose

$$T = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad T_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \quad (55)$$

then $T \leftrightarrow T_1$ because both parts of (46) are satisfied with

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (56)$$

and, given only the value for A , and it would be awkward to verify the existence of an appropriate relation B if Theorem 2.5 was not known. The relation B is determined by the fact that the columns of T and T_1 are given by $\{c \vee d, a, c, b, a, a \vee b, d, c \vee d\}$ and $\{a, b, c, d, a \vee d, b \vee c \vee d, c \vee d, b \vee c\}$ respectively, and $b(T) = b(T_1) = \{a, b, c, d\}$ can be expressed in matrix form

as

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (57)$$

Note that $A \neq B^{-1}$ and $AB \neq I$ because the first row of AB is zero and B is not unique because its first column could be replaced by $[00000010]$. A special case of this relationship (not true in this case) is when $AB = I$ and $A = B^{-1}$ when the columns of A and B are permutations of each other.

***** end

There is a kind of inverse of a relation A among relations denoted by A' and defined by

$$R_1 A' R_2 = R_1^{-1} A R_2^{-1} \quad (58)$$

for *any* such relation A and clearly $(A')' = A$.

2.2 Analytic Functions

For compositional powers of functions eg $f(f(z)) = f^{o2}(z)$ defined as for relations in general, the symbol o was used because it is sometimes used to indicate composition and it distinguishes the inverse from a reciprocal in the context of numerical functions and relations.

By considering the example when R_1 is $z \rightarrow z^2$, R_2 is $z \rightarrow z^3$ and we are expecting the intersection to be given by the root function $z \rightarrow z^6$ (R_3).

Returning to analytic functions there is a kind of discreteness in them which is exemplified by the fact that there does not appear to be a function $f()$ such that $z^4 \rightarrow f() \rightarrow z^2$ and $z^2 \rightarrow f() \rightarrow z^4$.

Consistency of the use of “simplest” with \rightarrow in this document.

Roughly this includes is any function, single or multivalued, that can be expressed by a formula that does not depend on splitting the complex variable z into parts e.g. real and imaginary or modulus-argument etc. or is the solution of any problem defined using calculus involving such functions. See the closure operations below. They are differentiable and therefore infinitely many times differentiable in the extended sense (including ∞) wherever they are defined. They have no boundaries. The phrase analytic relations could be used because they can be multivalued, but I will stick to using the term analytic functions because of its common use. The term “analytic” is used because it is hoped that these functions will be closely related to complex analytic functions as this term is usually used.

The function $\exp()$ plays a very special role. It uniquely satisfies $\exp(0) = 1$ and $\exp'(z) = \exp(z)$. It satisfies $\exp(z) = e^z$ whenever $z \in \mathbb{Z}$ and e is the base of natural logarithms. $\exp(x)$ is equal to the positive real value of e^x for other real x and is $e^x(\cos(y) + i \sin(y))$ when $z = x + iy$ thus there is a distinction between e^z and $\exp(z)$ with only the former being multivalued for non-integer and finite values of z . However due to the common usage that these are the same, if there is not likely to be an ambiguity e^z will be used when more properly $\exp(z)$ should be used.

Together with its inverse $\ln()$, $\exp()$ can be used to define the general exponent function by

$$a^b = \exp(b \ln(a)). \quad (59)$$

To show that this in general has the correct number of values (q when $b = p/q$ with $p \in \mathbb{Z}$, $q \in \mathbb{N}$, $q > 0$ and $\gcd(p, q) = 1$), let $n \in \mathbb{N}$ with $0 \leq n \leq q - 1$. Upon dividing np by q let $np = sq + r$ where $r \in \mathbb{N}$ and $0 \leq r \leq q - 1$, and $s \in \mathbb{Z}$. Then the mapping $k : n \rightarrow np \bmod q$ is a permutation of the integers $Q = \{0, 1, \dots, q - 1\}$. The mapping $k()$ is left-unique because $n_1 p \bmod q = n_2 p \bmod q \Leftrightarrow (n_1 - n_2)p = tq \Leftrightarrow n_1 = n_2$ where $t \in \mathbb{Z}$ and $0 \leq n_1, n_2 \leq q - 1$. The last step is because the gcd of p and q is 1 and $q \nmid p$ it follows that $q | (n_1 - n_2)$ and $1 - q \leq n_1 - n_2 \leq q - 1$. Therefore the fractional part of $\{np/q\}$ for $n \in Q$ is $\{n/q\}$ for $n \in Q$ in a different order and the sets $\exp(2\pi i np/q)$ and $\exp(2\pi i n/q)$ where $n \in Q$ are the same but in a different order. Therefore the set of values of $\exp(b \ln(a))$ for one particular value of $\ln(a)$ is $\exp(\frac{p}{q}(\ln(a) + 2\pi i n))$ where $b = p/q$ and is $\exp(\frac{p}{q} \ln(a)) \exp(2\pi i n/q)$ therefore the expression $\exp(b \ln(a))$ has all q values and no others and can be used to define a^b .

A peculiar consequence of dealing with multivalued expressions is an ambiguity that can arise when doing calculations that involve them. Consider the following paradox which is probably one of the simplest examples of its kind:

$$e^{i\pi} = -1 \Rightarrow 2\pi i = 2 \ln(-1) = \ln((-1)^2) = \ln(1) = 0! \quad (60)$$

While forgetting that $\ln()$ is multivalued it is too easy to carry out calculations like this and arrive at absurd conclusions. If for each instance of $\ln()$ it is remembered that any multiple of $2\pi i$ can be added to a result of $\ln()$ to give another value of the function, the following results is obtained:

$$\begin{aligned} 2 \ln(-1) &= 2(\pi i + 2n_1 \pi i) \\ \ln((-1)^2) &= 2n_2 \pi i \end{aligned} \quad (61)$$

where n_1, n_2 are arbitrary integers. The equality of these looks much better but is still clearly wrong because the LHS can never be 0 but the RHS can be with $n_2 = 0$, but all the values of the LHS are included in the RHS so that if z is one of the values in the LHS it is one of the values in the RHS.

There are two points where the logic is faulty in (60): first replacing $2\pi i$ by the multivalued expression $2\ln(-1)$ and secondly making the replacement $2\ln(-1)$ by $\ln((-1)^2)$ which as shown is not exactly the same set of values. Actually the substitution $\ln(a^b) = \ln(\exp(b\ln(a))) \rightarrow b\ln(a)$ is valid in the sense that any value of $\ln(a^b)$ i.e. $\ln(a^b) + 2n\pi i$ for any particular value of $\ln(a^b)$ with $b = p/q \in \mathbb{Q}$ and $n, p, q \in \mathbb{Z}$, can be written using (59) as $b\ln(a) + 2\pi i(n_1 + bn_2)$ with $n_1, n_2 \in \mathbb{Z}$ for some particular value of $\ln(a)$ i.e. is a member of the set $b\ln(a) \exp(2\pi ir/q)$ for $r \in \mathbb{N}$ satisfying $0 \leq r \leq q - 1$ but the reverse substitution is not always valid as the example in (61) shows.

The expression $\ln(a^b)$ is also problematic if $b \in \mathbb{Z}$ because it can be thought of as b instances of a or a^{-1} multiplied together and then the $\ln()$ taken, and the question arises if a is itself multivalued do those instances have to be the same value or can they be different values? Thus there is more than one possible interpretation of the same expression. A very simple example is what is the value of $(\pm 1) + (\pm 1)$? If these two instances have to be the same the result is ± 2 otherwise 0 can be included. The general principle it seems to me is to take note of when two or more instances of the same multivalued expression occur in a formula have a common origin meaning they have to have the same value, otherwise they are independent. Doing so will give the maximally informative result i.e. having the least number of possible values. In the above examples this means preferring the form $b\ln(a)$ to $\ln(a^b)$ implicitly requiring the a 's to be the same. Thus in general, the version of a formula having the smallest number of possible values is to be preferred after any other conditions in the problem have been taken into account. This accords with the idea that derived equations usually give just necessary conditions to solve a problem which might not be sufficient, then the results have to be checked against the original problem to see if the solutions are valid. Therefore in example (61), the LHS is to be preferred over the RHS. I'm not how to systematically avoid these problems except to question the meaning of an ambiguous expression where it first appears. They are probably the reason that multivalued functions are not often discussed.

A blanket assumption will be used throughout which is that functions are analytic unless otherwise stated i.e. the existence of derivatives of all functions mentioned which implies that the Cauchy Riemann equations are satisfied for each point $(z, f(z))$ except possibly for the singular points. The point ∞ was added to the complex plane to get the Riemann Sphere so that functions always have a value. This works for algebraic functions where continuity and differentiability hold for a function $f()$ even if $f()$ and its derivative go to ∞ there for example $z \rightarrow z^{-p/q}$ for $p, q \in \mathbb{N}$ at $z = 0$. However this does have some unusual consequences for example $z \rightarrow \exp(1/z^2)$ at $z = 0$ which is 0 and ∞ because $\exp(\infty)$ is 0 and ∞ (this follows from $e^z = e^x e^{iy}$ if $z = x + iy$ where if x and y approach ∞ with x/y is constant, the result is 0 if $x \rightarrow -\infty$ and ∞

if $x \rightarrow \infty$). These are examples of essential singular points for non-algebraic functions where the number of terms in the series about the singular points is infinite (Laurent series for finite singular points and power series for singular points at ∞).

A singular point at (z, w) is finite iff $z \neq \infty$

3 Closure operations

The set of algebraic functions includes the identity function $z \rightarrow z$ and constant functions $z \rightarrow c$ for any $c \in \overline{\mathbb{C}}$ and is closed under the following unary and binary operations on functions.

1. union
2. composition
3. inversion
4. addition
5. subtraction
6. multiplication
7. division
8. differentiation

except that the inverse of the constant function does not exist. The subtraction operation is merely the addition of a negative and so is not strictly required. The inclusion of division is needed to ensure that the special function $z \rightarrow 1/z$ is included.

The arithmetic operations just refer to the operations $f(z) = g_1(z) * g_2(z)$ for defining $f()$ in terms of $g_1()$ and $g_2()$ where $g_1()$ and $g_2()$ are analytic functions, then so will $f()$ where $*$ is $+$, $-$, \times or \div . Likewise the derivative $f'(z)$ is an analytic function if $f(z)$ is.

The absence of integration as a closure operation for algebraic functions suggests the extension of these ideas to include it as an operation that gives closure. This requires the familiar functions $\ln()$ and $\exp()$ to be included and some functions with singular points that are not poles or branch points known as essential singular points [Churchill et al.].

However including instead the limit of a sequence of functions can replace including derivatives and integrals. Differentiation does not need to be included as a closure operation because a derivative is the limit $f'(z) = \lim_{h \rightarrow 0} \frac{f(z+h)-f(z)}{h}$ of a difference that is already included. Also an integral is just the limit of a sum

$$\int_a^b f(z)dz = \lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{i=0}^{n-1} f\left[a + i \left(\frac{b-a}{n}\right)\right] \right\} \quad (62)$$

which is already included. Note that the limit of a set of continuous functions can be discontinuous in the real domain, and this extends to evaluating an analytic function on a path in $\overline{\mathbb{C}}$ that goes through a singular point that arises as a result of the limit taken. Therefore the closure operations to define the set of analytic functions are as follows

1. union
2. composition
3. inversion
4. addition
5. subtraction
6. multiplication
7. division
8. limit

that naturally fall into three categories, the first three involve only sets and relations, the second set of four are the arithmetic operations, and finally the limit operation allows all operations of calculus to operate within this algebra.

The main conceptual difficulty with my approach compared with the standard approach to complex analysis is how to deal with multivaluedness. The obvious first step is to define composition and inversion as for binary relations in general. When working with multivalued functions, the equivalent of the function value is now a set of values and equality between relations is of course the equality between the two sets of values. This has consequences when manipulating equations with multi-valued analytic functions.

Perhaps this simplest closure operation is that of union. A union is simply the union of the two sets of pairs (z, w) defining each of the functions in the union. The concept of a union was not mentioned much in my previous paper. The simplest example of a union is when $f(z) = (z^2)^{1/2}$ which is the union of z and $-z$ which consists of pairs (z, z) and $(z, -z)$ for all z in $\overline{\mathbb{C}}$.

Related to “union” is the concept of a component. A component will be a single analytic surface i.e. an analytic function that itself could be multivalued. The number of components an analytic function has will be an important property of it. Generally, only solutions of equations which consist of a single component are likely to be of interest. If a set of single components each satisfy an equation of the type considered here, then so does their union. Unless otherwise stated an arbitrary function will refer to a single component. The operation of extracting all the components from a union will probably be needed.

An analytic function can be a union of smoothly differentiable components that each consist of a single continuum of points $(z, f(z)) \in \overline{\mathbb{C}} \times \overline{\mathbb{C}}$ provided there is an extension of the notion of differentiation from \mathbb{C} to $\overline{\mathbb{C}}$. Finite and countable unions will surely be needed.

Singular points specified by $(z, f(z))$ are points in the analytic surface where a small circuit round z is *not* mapped into a small circuit round $f(z)$ in the Riemann Sphere. They are not to be confused with points where $f(z)$ is ∞ though these may often coincide. Another way to state this is to say a singular point $(z, f(z))$ is any point about which for all neighbourhoods N of $(z, f(z))$ in $\overline{\mathbb{C}} \times \overline{\mathbb{C}}$ however small, the graph of $f()$ intersected with N is not topologically equivalent to an open disk. In such a case one value of z will correspond to more than one value of $f(z)$ or vice versa in N .

4 A simple example

Probably the simplest example of the type of equation mentioned above is $f(z) = -f(z)$. For right-unique equations this of course means $f(z) = 0$, but because $f()$ can be multivalued it just means that whenever $(z, f(z))$ is in the multi-sheeted analytic surface i.e. the graph of $f()$, then so is $(z, -f(z))$. The inverse of this clearly satisfies

$$f(z) = f(-z) \tag{63}$$

and this is satisfied by $f_1(z) = z^2$ and by $f_2(z) = z^4$ and in fact any function of z^2 . Suppose the condition is required to be an inequality unless equality is explicitly required, then in the above case

$$f(z_1) = f(z_2) \Leftrightarrow z_1 = \pm z_2. \tag{64}$$

This eliminates z^4 from being a solution because then $f(z_1) = f(z_2) \Leftrightarrow z_1 = \pm z_2$ or $z_1 = \pm iz_2$. Now the question is does (64) (which implies (63)) have the solution set $f(z) = a + bz^2$ for arbitrary constants a and b . Is this the same as requiring $f()$ to have the minimum number of singular points?

5 Another look at algebraic functions

The topology of an algebraic function clearly must involve the behaviour at points that are not regular points where the behaviour is non-trivial. A simple way to describe this is to imagine a small circle described around the point (z_0, w_0) within the surface. Imagine it so small that no other points with irregular behaviour are included. If this can be done it will have projections down to both the z and w planes and if the circuit is complete ending where it started, the projections will be circuits around z_0 and w_0 respectively described p and q times, or for non-algebraic functions, either p or q may be infinite if the corresponding circuit never joins up again. Such points (z_0, w_0) with either p or q not equal to 1 are singular points and if $p = q = 1$ the point is a regular

or non-singular point. Another kind of thing that can happen is when (z_0, w_0) is at the intersection of two or more surfaces, which again implies (z_0, w_0) is a singular point. In general a singular point is where in a small region surrounding it, the function surface(s) cannot be stretched so that it becomes flat.

Using the methods I developed earlier [Nixon2013] to locate singular points for algebraic functions, suppose $w = z^{p/q}$ where $p, q \in \mathbb{N}$ then $w^q = z^p$ and $P = w^q - z^p = 0$ and $\partial P/\partial z = -pz^{p-1} = 0 \Rightarrow z^{p-1} = 0$ which is false if $p = 1$. If $p > 1$ then $z = 0$ and $w = 0$. Also $\partial P/\partial w = qw^{q-1} = 0 \Rightarrow w^{q-1} = 0$ which is false if $q = 1$. If $q > 1$ then $w = 0$ and $z = 0$. Therefore all finite singular points are at $(0, 0)$ provided $p > 1$ or $q > 1$ with the transformation $w^* = 1/w, z^* = 1/z$ giving the other one at $z^* = 0, w^* = 0$ i.e. (∞, ∞) . Now suppose $p > 0$ and $q < 0$ then the same argument gives that all singular points are at $(0, \infty)$ or $(\infty, 0)$. In many examples of algebraic functions I have studied, it is easy to miss a singularity with either z or w being ∞ in addition to the finite singular points. It is later proved that no analytic function can have just one singular point.

Consider $f(z) = z^{1/q}$ where q is an integer. Rather than describing this behaviour simply by saying that it is expressed by a “winding number”, near the branch point at $z = 0$, the idea is to relate $f(z)$ to f evaluated at the “next” branch of the function obtained by tracking $f(z)$ continuously once round a small circle surrounding $z = 0$ described in the anticlockwise direction until the same point z is reached. This circuit in z will have to be described q times to get back to the same value of $f(z)$. This is because if $f(z) = z^{1/q} = r^{1/q}e^{i\theta/q}$ then $f(z)^q = z = re^{i\theta}$ with $0 \leq \theta \leq 2\pi q$. Let $g_1(z) = e^{2\pi i/q}z$ where q is a positive integer. Then $g_1(f(z)) = e^{2\pi i/q}r^{1/q}e^{i\theta/q} = r^{1/q}(e^{2\pi i}e^{i\theta})^{1/q} = r^{1/q}e^{i\theta/q} = f(z)$.

In fact this equation, being an equation for a multivalued function, represents the equality of the two sets of values each being q in number, and the equation generates a permutation of those q values. Equality of the sets of values will be implied whenever an equality occurs between two multivalued expressions. This is a simple example of equations which now have to be treated differently because the expressions are multivalued. This relationship is a better way of describing this situation because it just involves the right-unique function $g_1()$ and no mention of topological concepts that are not so easy to make precise. However $f(z) = z^{1/q}$ is clearly not the only solution of

$$f(z) = e^{2\pi i/q}f(z) \quad (65)$$

(for example $f(z) = az^{1/q}$ or $f(z) = z^{p/q}$). Consider what can be said about the single component solutions of (65) in general. Raising (65) to the power q gives the tautology $f^q = f^q$ so there is nothing that can be said about f^q except that it is also a single component, so every single component solution of (65) is the q th root of some analytic function regardless of its other singularities i.e.

$f(z) = h(z)^{1/q}$ for an arbitrary function h is the general solution of (65). Any such function has a q -fold branch point at all points where $f = h = 0$, and satisfies (65) because $(e^{2\pi i/q})^q = 1$. Relaxing the condition of a single component, any union of the form $\{f(z).e^{2\pi ij/q}$ for $0 \leq j < q\}$ obviously satisfies (65) and has q components. It is this example that motivated the introduction of the concepts of a union and the components of an analytic function. As a simple example, is $f(z) = z + 1$ a solution of (65)? No because it is not the q -th root of an analytic function. It is not the q th root of $(z + 1)^q$ which is the union $\{e^{2\pi ij/q}(z + 1)$ for $0 \leq j \leq q - 1\}$. The result of this can be written more succinctly as follows.

Lemma 5.1. *Suppose a multivalued function $g()$ satisfies*

$$f(z) = e^{2\pi i/q} f(z) \tag{66}$$

for some $q \in \mathbb{N}$ then this is equivalent to

$$f(z) = h(z)^{1/q} \tag{67}$$

for some function $h()$.

Note: if the last step in computing $h()$ is to apply $z \rightarrow z^q$, all q values in (67) must still be used giving a union of q components.

Proof. From (66) interpreted as an equation for a set of values it can be iterated q times to come to the original starting value i.e. (66) determines the set $e^{2\pi in/q} f(z)$ for all n such that $0 \leq n \leq q - 1$, so $g(z)^q$ is a single value $h(z)$ say, for each original starting value $f(z)$, and $h(z)$ can be anything i.e. any solution of (66) is of the form

$$f(z) = h(z)^{1/q} \tag{68}$$

for some function $h()$ and any function of this form satisfies (66). □

In this context a concept arises, made precise later, which could be called the *simplest solution* of an equation. This term will always be italicised to indicate this meaning. As will be shown later, the singular point(s) associated with an equation like (66) is/are given by where this equation is treated as an equation for a single valued quantity i.e. where $f(z) = 0$ or ∞ . Now suppose that $f()$ satisfies no conditions other than (66) hence there are no other equations determining behaviour at the singular point(s) at $f(z) = 0, \infty$. Therefore there can be no conditions affecting behaviour of $h()$ near the point(s) where $f(z) = h(z) = 0, \infty$ i.e. no need for singular point(s) there, and no other singular points for $h()$ because they would imply other singular points for $f()$. Therefore the *simplest solution* could be defined such that $h()$ has no singular

points so $h(z) = a + bz$ and $f(z) = (a + bz)^{1/q}$. This can be summarised by saying that the *simplest solution* of (66) is $f(z) = (a + bz)^{1/q}$.

The changed definitions of singular points and the new definition of an inversion point require some well-known theorems to be rephrased.

This is because “analytic” in the textbooks should be replaced by “right-unique analytic and finite” in the terminology of this paper. This would make the statements of theorems like the Cauchy integral formula slightly more cumbersome. Also “entire” means “right-unique, analytic, and without any singular points except possibly at $z = \infty$ ”. If a function is bounded i.e. $|f(z)| < k$ for some $k > 0$ for all $z \in \mathbb{C}$ then by continuity, it cannot be ∞ at any point in $\overline{\mathbb{C}}$ including at ∞ itself. Therefore Liouville’s theorem can be expressed as

Theorem 5.2. *If $f()$ is right-unique analytic, finite at every point $z \in \overline{\mathbb{C}}$ and without a singular point at any point $z \in \mathbb{C}$ then $f()$ is constant $\in \mathbb{C}$.*

Now suppose that $f()$ is right-unique, analytic and none of its values are equal to $w \in \mathbb{C}$ at any point $z \in \overline{\mathbb{C}}$ and $f()$ has no singular points with $z \in \mathbb{C}$. Then $\frac{1}{f(z)-w}$ is everywhere finite because $f(z)$ cannot approach w arbitrarily closely (for otherwise at the limit point it would equal w and $\overline{\mathbb{C}}$ is compact so includes all its limit points) and analytic and has no singular points for $z \in \mathbb{C}$ by Theorem 5.2, $\frac{1}{f(z)-w} = c \in \mathbb{C}$, therefore $f(z)$ is constant $\in \overline{\mathbb{C}}$. This proves that

Theorem 5.3. *Every right-unique analytic function $f()$ without any singular points where $z \in \mathbb{C}$ reaches every value $f(z) \in \overline{\mathbb{C}}$ for some $z \in \overline{\mathbb{C}}$ unless $f()$ is a constant $\in \overline{\mathbb{C}}$.*

Suppose a single component analytic function maps p values each to the same q values $\in \overline{\mathbb{C}}$. Does every single component analytic function have to be like this, with p or q allowed to be ∞ ? In the two set of values, each member of a set is equivalent to any other member.

An analytic function $f()$ has a single component if and only if for every pair of points P_1 and P_2 in $\overline{\mathbb{C}} \times \overline{\mathbb{C}}$ in the graph of $f()$ there is a continuous and analytic curve starting at P_1 and ending at P_2 at each point being in the graph of $f()$ and not including any singular point of $f()$ i.e. every such point is connected not via singular points to every other such point within the graph of $f()$.

For a multivalued single component analytic function $f()$ it is possible to have a circuit in which the z value is returned to but w comes back to a different value. That gives rise to an equation of type (108). As the circuit is reduced in size, at some points the final value reached will suddenly change and will eventually will suddenly equal the original value. It suddenly changes where the curve crosses a singular point of which there can be many. Having found all

the singular points and their associated equations relating the function values, it should be possible to, by following any combination of the paths in any order allowing repetition, each of which is associated with a single singular point, to get from say (z_1, w_1) to any other point (z_1, w) in the graph of $f()$. This would indicate that all the equations of type (108) have been found. It is possible (see for example (97)) that there is a pair (or perhaps more) of singular points that are associated with the same transformation (108) or its inverse.

Similarly there can be circuits that return the w to the same value but z returns to a different value. This gives rise to an equation of type (107) and is equivalent to doing the same thing for $f^{o-1}()$. There could be a finite or a countably or uncountably infinite number of singular points. See for example (93) with solution (95) that has uncountably many singular points on the unit circle. For the case when the number of singular points is finite or countably infinite, this leads to the graph of $f()$ being described as a set of collections of points say $z_1, z_2 \dots z_p, w_1, w_2 \dots w_q$ such that every one of the z 's is mapped to all of the w 's in every collection. Away from singular points, all the z 's are distinct and so are all the w 's. Therefore the positive integers p, q are constants for the function $f()$, but either could be ∞ . It may be useful to define the signature of an analytic function to be say $\{(p_1, q_1), (p_2, q_2), \dots\}$ where each of the pairs corresponds to one component of the function.

Theorem 5.4. *every analytic function reaches every value $f(z) \in \overline{\mathbb{C}}$ for some $z \in \overline{\mathbb{C}}$ unless $f()$ is a constant $\in \mathbb{C}$.*

Proof. This follows from the corresponding property of algebraic functions ($P(z, w) = 0$ always has a solution for z given w for any bivariate polynomial P) and the fact that analytic functions are continuous and are limits of sequences of algebraic functions which are all continuous. \square

An interesting case occurs if the point that is the solution of such an equation approaches, under the limit, a singular point of the limit function. For example if the limit function is $f(z) = \exp(1/z)$ and the solutions approach $z = 0$ as would happen if $w = 0$. This works because $f(0)$ is 0 and ∞ i.e. both these values are attained by $f()$.

Lemma 5.5. *An analytic function with no singular points and no inversion points is a linear function.*

Proof. The absence of a singular point at $z = \infty$ for a function $f()$ implies a neighbourhood of ∞ (a large circle in the complex plane but a small circle in the Riemann Sphere) in z maps in a left-unique manner locally to a neighbourhood in w say centred on $w_0 = f(\infty)$. If $w_0 \neq \infty$ then $1/z \approx a(w - w_0)$ for very large $|z|$ therefore $dw/dz = -1/az^2 \rightarrow 0$ as $z \rightarrow \infty$. Similarly if $w_0 = \infty$ a small neighbourhood in $1/z$ about 0 maps to a small neighbourhood round $1/w$ at 0

so $1/z \approx b/w$ therefore $dw/dz = b$ at (∞, ∞) and if there are no singular points and no inversion points anywhere in $w(z)$ then dw/dz is also everywhere finite and analytic there, so by Liouville's theorem (see for example [Churchill et al.]) dw/dz is constant so $w = a + bz$ where a and b are constants. \square

Theorem 5.6. *An analytic function with no singular points is a bilinear function given by $f(z) = \frac{a+bz}{c+dz}$.*

Proof. Let $f(z) = w$ be an analytic function with no singular points. Then apply a bilinear function $b(\cdot)$ to w such that $b(f(0)) = 0, b(f(1)) = 1, b(f(\infty)) = \infty$. This can be done uniquely (see [Churchill et al.] section 33). Then by Lemma 6.11 $b(f(\cdot))$ has no singular points and maps, $0 \rightarrow 0, 1 \rightarrow 1$, and $\infty \rightarrow \infty$. Also $b(f(\cdot))$ can have no inversion point because if some finite point $z_0 \rightarrow \infty$ then $b(f(\cdot))$ would be not left-unique there and would have a singular point there by Lemma 6.5 contradicting the assumption. Therefore $b(f(\cdot))$ satisfies the condition of Lemma 5.5 and must be a linear function i.e. $b(f(z)) = \alpha + \beta z$ and $f(z) = b^{-1}(\alpha + \beta z)$ which is also a bilinear function. \square

5.0.1 Characterising power functions

Lemma 5.7. *If $p \in \mathbb{N}$ where $p > 1$ then*

$$f(z) = f(e^{2\pi i/p} z) \quad (69)$$

for all $z \in \overline{\mathbb{C}}$ for some analytic function $f(\cdot)$ if and only if

$$f(z) = h(z^p) \quad (70)$$

for all $z \in \overline{\mathbb{C}}$ where $h(\cdot)$ is some other analytic function.

Note: if the first step in computing $h(z)$ is to apply $z \rightarrow z^{1/p}$, all p values must be included, giving a result which is a union of p components.

Proof. Equation (69) implies all p values $e^{2\pi i j/p} z$ for $0 \leq j \leq p-1$ have the same value of g and z^p is the same for all these. Also the distinct sets $\{z, e^{2\pi i/p} z, e^{4\pi i/p} z, \dots, e^{(p-1)\pi i/p} z\}$ for all $z \in \overline{\mathbb{C}}$ have the union which is $\overline{\mathbb{C}}$ and are disjoint. Thus any solution of (69) on the Riemann Sphere $\overline{\mathbb{C}}$ is of the form (70) and any function of this form satisfies (69) because $f(e^{2\pi i/p} z) = h((e^{2\pi i/p} z)^p) = h((e^{2\pi i/p})^p z^p) = h(z^p) = f(z)$. \square

Again an argument motivating the concept of the *simplest solution* follows. An extra condition on $f(\cdot)$ is obviously connected with an extra condition on $h(\cdot)$ and vice versa because of the relationship (70). Therefore there is no extra condition on $f(\cdot)$ is equivalent to saying that there is no extra condition on $h(\cdot)$. There are no conditions on $h(\cdot)$ at the moment, therefore this argument

gives rise to the notion of the *simplest solution* of an equation such as (69) in which, because its general solution involves an arbitrary function $h()$ with no conditions placed on it giving rise to singular points, $h()$ will be assumed to have no singular points and therefore be a linear function. The points z at which (69) requires a singular point are when the two function arguments coincide i.e. $z = e^{2\pi i/p}z$ giving $z = 0$ and ∞ . An extra condition on $f()$ modifying the behaviour surrounding the singular points at $z = 0, \infty$ will require an extra condition on $h()$ also requiring a singular points at $z = 0, \infty$. A singular point in $h()$ at any finite point $z_0 \neq 0$ implies $f()$ has singular points at all finite points $z_0^{1/p} \neq 0$. Therefore the *simplest solution* of (69) is $f(z) = a + bz^p$.

A related example is $f(z) = (z - z_0)^p$ where p is a positive integer. Here the only finite singular point is at $(z_0, 0)$. Introducing the variable s by $s = z - z_0$, and $f^*(s)$ by $f^*(s) = f(z) = s^p$ then $f^*(s)$ satisfies (69). Therefore expressing this in terms of f using the chain of equalities

$$f(z) = f^*(s) = f^*(e^{2\pi i/p}s) = f^*(e^{2\pi i/p}(z - z_0)) = f(e^{2\pi i/p}(z - z_0) + z_0) \quad (71)$$

i.e. $f()$ satisfies

$$f(z) = f(g_2(z)) \text{ where } g_2(z) = e^{2\pi i/p}(z - z_0) + z_0. \quad (72)$$

This relationship just involves the right-unique function $g_2()$.

Suppose a multivalued function satisfies

$$f(z) = e^{2\pi i/q} f(e^{2\pi i/p}z) \quad (73)$$

where $p, q \in \mathbb{N}$ then this is equivalent to $f^*(z) = f^*(e^{2\pi i/p}z)$ where now $f^*(z) = (f(z))^q$ or equivalently (by Lemma 5.7) $f^*(z) = h(z^p)$ i.e.

$$f(z) = (h(z^p))^{1/q} \quad (74)$$

for some function $h()$ and the *simplest solution* of (73) is

$$f(z) = (az^p + b)^{1/q}. \quad (75)$$

This function has finite singular points at $((-b/a)^{1/p}, 0)$ so if in addition $f(z)$ has no finite singular point other than at $(0, 0)$ then $b = 0$ and $f(z) = (az^p)^{1/q}$.

If there are other singularities, equations like (69) and (70) will not necessarily be exact but only asymptotically correct as the corresponding singular point is approached. For example in (75) if $z = (-b/a)^{1/p} + \epsilon$ then $f(z)$ can be expanded as a power series in ϵ in which terms higher than the first contribute so that the asymptotic behaviour near $((-b/a)^{1/p}, 0)$ is affected by the singular point at $(0, 0)$.

The ideas in Equations (72) and (66) can be combined by considering the solutions of

$$f(z) = e^{2\pi i/q} f(e^{2\pi i/p}(z - z_0) + z_0). \quad (76)$$

Introducing the new variable $s = z - z_0$ and the new function $f^*(s) = f(s + z_0)$ then (76) becomes

$$f^*(s) = e^{2\pi i/q} f^*(e^{2\pi i/p} s) \quad (77)$$

whose general solution is $f^*(s) = (h(s^p))^{1/q}$ i.e. therefore the general solution of (76) is $f(z) = [h((z - z_0)^p)]^{1/q}$.

As would be expected (and is justified later) the singular point(s) of $f()$ are given by

1. where the argument of the q -th root i.e. $h((z - z_0)^p) = 0$ or ∞
2. where $(z - z_0)^p$ is a singular point of $h()$
3. where $z - z_0$ is a singular point of the p -th power function which is at 0 and at ∞ so $z = z_0, \infty$.

For the case where $h()$ is the identity function, the second singular point no longer exists and the first and third of these singular points coincide at $z = z_0, \infty$ and $f(z) = (z - z_0)^{p/q}$ and the winding number ratio is $q : p$ in the earlier description.

6 General theory singular points

All the types of singular point so far found are of the types $q : p$ representing the winding number ratio where p and q are positive integers have no common factors. These are all the types of singular points for algebraic functions. In the cases where p and q are finite, a singular point (z_0, w_0) is a point about which if a path is traced from the starting point back to itself q times in the z plane this corresponds to a path in the w plane described p times back to itself.

The most general form of equations such as (63), (65), (69), (73) and (76) that describe the behaviour in the neighbourhood of a singular point seems to be

$$f(z) = g_1(z, f(g_2(z))) \quad (78)$$

in which g_1 has direct z dependence in addition to its dependence on $f()$. The meaning of (78) where $g_2()$ is the identity function is that there is an associated singular point (z_0, w_0) which is the point about which if a path in the z plane is followed to its starting point and if the function value is followed continuously, the values of the function at each end of the path are related by (78). This is the case where $q = 1$. As will be shown, the singular point is also a point

where the number of function values changes and w_0 is given by the different values of the function $w_0 = f(z_0)$ being equal. This can be used to determine (z_0, w_0) .

There is another version of this to describe the situation where $p = 1$. In this case $g_1()$ is the identity function and the roles of z and $w = f(z)$ are reversed. There is then a point (z_0, w_0) about which if a continuous path is traced in the w plane back to itself then the corresponding values of z are related by (78). The equality of these values determines the value z_0 .

In addition to these cases, for non-algebraic functions it is possible to have $q = \infty$. In this case the value of w is never returned to its original value. Probably the simplest example is $w = f(z) = \ln(z)$ the inverse of the complex exponential function. This is equivalent to $z = \exp(w) = \exp(w) \cdot \exp(2\pi i) = \exp(w + 2\pi i)$. Therefore $w + 2\pi i = \ln(z)$ and equation (78) is satisfied for $f() = \ln()$ and $g_2(z) = z + 2\pi i$ and $g_1(z, f) = f$. Therefore the singular points are given by $z = z + 2\pi i$ from lemma 6.9 which implies $z = \infty$. This resolves the paradoxical situation with Theorem 1.1 and Lemma 5.5 and the fact that the exponential function has no finite singular points (they are at $z = \infty$ with $w = 0$ and ∞). As in the examples above $g_1()$ and $g_2()$ are right-unique and the singular point of $f()$ is at $z = 0$.

6.1 Definition and properties of singular points

In all these definitions, a neighbourhood of a point $(z, w) \in \overline{\mathbb{C}} \times \overline{\mathbb{C}}$ is an open set containing (z, w) in the cartesian product topology.

These results depend on general properties of mappings right-unique versus multi-valued, and left-unique versus many-to-one. These properties can be defined such that they are local to a particular point as follows.

Definition 6.1. *The function $f()$ is locally left-unique at $P = (z, f(z))$ if and only if there is a neighbourhood N of P such that for every pair $(z_1, f(z_1))$ and $(z_2, f(z_2))$ in N , $z_1 \neq z_2 \Rightarrow f(z_1) \neq f(z_2)$.*

and likewise

Definition 6.2. *The function $f()$ is locally right-unique at $P = (z, f(z))$ if and only if there is a neighbourhood N of P such that for every pair $(z_1, f(z_1))$ and $(z_2, f(z_2))$ in N , $f(z_1) \neq f(z_2) \Rightarrow z_1 \neq z_2$.*

Definition 6.3. *$f()$ has a singular point P at $(z, f(z))$ if and only if for all neighbourhoods N of P there exists $(z_1, f(z_1)) \in N$ and $(z_2, f(z_2)) \in N$ such that either $[z_1 \neq z_2 \text{ and } f(z_1) = f(z_2)]$ or $[z_1 = z_2 \text{ and } f(z_1) \neq f(z_2)]$.*

An equivalent statement of this is to require this condition only for all neighbourhoods of a specified neighbourhood of P however small it is. This makes it clearer that the condition is a local property of the behaviour at P .

Definition 6.4. *This is the same as saying the condition that needs to be satisfied for the absence of a singular point of the function $f()$ at the point P , $(z, f(z))$ is that there exists a neighbourhood N of P such that*

$$\forall (z_1, f(z_1)), (z_2, f(z_2)) \in N [z_1 = z_2 \Leftrightarrow f(z_1) = f(z_2)] \quad (79)$$

i.e. $f()$ is left-unique and right-unique within N .

Now it is easy to show that

Lemma 6.5. *A function $f()$ has a singular point at $P = (z, f(z))$ if and only if $f()$ is either not locally left-unique there or $f()$ is not locally right-unique there.*

Proof. It is only necessary to choose the neighbourhood that is the intersection of the two neighbourhoods in definitions 6.1 and 6.2 and take the negation of the result. \square

Next follows a pair of trivial yet confusing lemmas that show

Lemma 6.6. *If $f()$ is an analytic function that is a solution of (108) then $f()$ has singular points at every point (z, w) that is a solution of $w = g_1(z, w)$ where $w = f(z)$.*

This is quite confusing because the word “solution” is being used in different contexts and the same equation (108) is being used in two different ways, one to determine $f()$ and the other once $f()$ is fixed, to determine the set of singular points of $f()$.

Proof. Let P be such a point then for every neighbourhood of P there will be points where w is arbitrarily close but not equal to $g_1(z, w)$. Thus by the second option of definition 6.3, P is a singular point of $f()$. \square

And likewise

Lemma 6.7. *if $f()$ is an analytic function that is a solution of (107) then $f()$ has singular points at every point that satisfies $z = g_2(z)$.*

Proof. likewise using the first option in 6.3 \square

Definition 6.8. *$f()$ has an inversion point at $(z, f(z))$ if and only if $f(z) = \infty$.*

It is possible for a singular point to also be an inversion point e.g. $f(z) = z^{-2}$ at $z = 0$. An example of an inversion point that is not a singular point is $f(z) = z^{-1}$ at $z = 0$ because this function is everywhere right-unique and left-unique.

The definition used in my earlier paper on algebraic functions [Nixon2013] includes inversion points with the singular points, and inversion points were not considered as a separate category. The reason for separating them out is for consistency in definition 6.3 that now works even if $f(z) = \infty$ where a neighbourhood of ∞ is as would be expected on the Riemann Sphere i.e. a region of the complex plane outside of a finite connected region defined by a single boundary.

A topological argument involving moving $f(z_0)$ to ∞ where z_0 is a singular or inversion point suggests that the direction of traversal of $f(z)$ round a circuit surrounding $(z, f(z))$ (P) is the same as that of the corresponding circuit z for any point P in the graph of $f()$ except when $f(z_0) = \infty$ when it is reversed as the result of this circuit crossing ∞ .

Lemma 6.9. *In definition 6.3 the location of the singular point(s) is determined by $z_1 = z_2$ and $f(z_1) = f(z_2)$.*

Proof. If a singular point P for $f()$ is due to $f()$ not being left-unique, in all neighbourhoods of P , there exists $(z_1, f(z_1))$ and $(z_2, f(z_2))$ such that $z_1 \neq z_2$ and $f(z_1) = f(z_2)$. Clearly if z_1 and z_2 , which are related, are forced to satisfy $z_1 = z_2$ then this has special significance. In fact this determines the location of the singular point or points. This is because if a region surrounding where z_1 and z_2 are forced to be equal is excluded, then surrounding such a point P a sufficiently small neighbourhood N exists such that because of $z_1 \neq z_2$ not both of z_1 and z_2 can be in N and the condition for a singular point fails. Therefore the singular points can only be at points P given by $z_1 = z_2$ where this is the only solution of $f(z_1) = f(z_2)$ which also holds at P . The existence of points where this is not true arbitrarily close to P will guarantee that P is a singular point. Likewise this will work if the singular point is due to $f()$ not being right-unique i.e. if $f()^{o-1}$ is not left-unique by swapping the roles of z and $f(z)$. Thus singular points are where the number of function values changes. \square

The following results relate singular behaviour to the operations of inversion, composition, addition and multiplication, and union.

Lemma 6.10. *$(z, f(z))$ is a singular point of $f()$ if and only if $(f(z), z)$ is a singular point of $f^{o-1}()$.*

Lemma 6.11. *Composition with a function $h()$ that is analytic and has no singular point at a particular location implies that the singular/non-singular status of $f()$ is the same as that of $h(f())$ and $f(h())$ each at the corresponding point.*

Proof. Suppose $h()$ is analytic and has no singular point at $(z_1, h(z_1))$ then there is a neighbourhood N_1 of $(z_1, h(z_1))$ such that for every pair $(z_2, h(z_2))$

and $(z_3, h(z_3)) \in N_1$, $z_2 = z_3 \Leftrightarrow h(z_2) = h(z_3)$. Then $f()$ has no singular point at $(h(z_1), f(h(z_1)))$ if and only if there is a neighbourhood N of $(h(z_1), f(h(z_1)))$ such that for every pair $(z_4, f(z_4))$ and $(z_5, f(z_5))$ in N , $z_4 = z_5 \Leftrightarrow f(z_4) = f(z_5)$. Let N_2 be the image of N (with a typical point being (x_1, x_2)) under the mapping $k()$ defined by $x_1 \rightarrow h^{o-1}(x_1)$, $x_2 \rightarrow x_2$. This mapping is left-unique and right-unique because $h()$ is. Now let N_3 be the subset of N_2 such that the first component of each point is also in N_1 . This will be non-empty because N_1 and N_2 are both neighbourhoods centred on a point with first component z_1 . Then for any pair of points $(z_6, f(h(z_6)))$ and $(z_7, f(h(z_7)))$ in N_3 , $z_6 = z_7 \Leftrightarrow h(z_6) = h(z_7) \Leftrightarrow f(h(z_6)) = f(h(z_7))$. The first equivalence is true because of the property of $h()$ and the second is true because of the property of $f()$. The existence of such a neighbourhood N_3 is precisely the statement that $f(h())$ has no singular point at $(z_1, f(h(z_1)))$. The other half of the theorem will be proved similarly or by considering the inverses of these functions. \square

For a very similar reason

Lemma 6.12. *Adding or multiplying a function by another analytic function without a singular point will not alter the singular/non-singular status of the function at the corresponding point.*

Lemma 6.13. *The only singular points of a union that are not included in one of the separate components is where at least two components intersect.*

These are known as intersection singular points. [needed?

Lemma 6.14. *If $f()$ is right-unique with a singular point at $(z_1, f(z_1))$ then $h(f())$ has a singular point at $(z_1, h(f(z_1)))$.*

Proof. Because $f()$ is right-unique, the second option in definition 6.3 is not possible i.e. for all neighbourhoods N of $(z_1, f(z_1))$ there exists $(z_2, f(z_2))$ and $(z_3, f(z_3)) \in N$ such that $z_2 \neq z_3$ and $f(z_2) = f(z_3)$. If $h()$ is any analytic function then $h(f(z_2)) = h(f(z_3))$ where $h(f())$ is analytic, and if $h()$ is multivalued these sets are the same. Therefore for all neighbourhoods N' , defined as an image of N under $h()$, centred on $(z_1, h(f(z_1)))$ there exists $(z_2, h(f(z_2)))$ and $(z_3, h(f(z_3))) \in N'$ where $z_2 \neq z_3$ and $h(f(z_2)) = h(f(z_3))$ implying $h(f())$ has a singular point at $(z_1, h(f(z_1)))$. \square

]

Lemma 6.15. *If $f()$ has a singular point at $(z_1, f(z_1))$ then $h(f())$ has a singular point at $(z_1, h(f(z_1)))$.*

Proof. $f()$ has a singular point at $(z_1, f(z_1))$ if and only if for all neighbourhoods N of $(z_1, f(z_1))$ there exists $(z_2, f(z_2))$ and $(z_3, f(z_3)) \in N$ such that $z_2 \neq z_3$ and $f(z_2) = f(z_3)$ or $z_2 = z_3$ and $f(z_2) \neq f(z_3)$. If $h()$ is any function, in the first case these conditions can be written as $[z_2 \neq z_3 \text{ and } h(f(z_2)) = h(f(z_3))]$ where if $h()$ is multivalued these sets are the same. In the second case $[z_2 = z_3 \text{ and } h(f(z_2)) \neq h(f(z_3))]$ if $h()$ is left-unique and if not,

, . Therefore for all neighbourhoods N' , defined as an image of N under $h()$, centred on $(z_1, h(f(z_1)))$ there exists $(z_2, h(f(z_2)))$ and $(z_3, h(f(z_3))) \in N'$ where $z_2 \neq z_3$ and $h(f(z_2)) = h(f(z_3))$ implying $h(f())$ has a singular point at $(z_1, h(f(z_1)))$. \square

The importance of this result is that it is not possible to remove a singular point in a right-unique analytic function e.g. $z \rightarrow z^2$ by applying another function to the result. For example applying $z \rightarrow z^{1/2}$ gives the union $z \rightarrow \pm z$ that has an intersection singular point where these components coincide at $(0, 0)$.

7 Examples of non-algebraic analytic functions and their singular and inversion points.

Analysis of behaviour in the neighbourhood of singular points similar to the above can be found for functions of a complex variable that are not algebraic as the following examples show.

Returning to $f(z) = \ln(z)$, it satisfies $f(z) = f(z) + 2\pi i$. Conversely $f(z) = f(z) + 2\pi i$ implies, taking the exp of both sides, the identity $\exp(f(z)) = \exp(f(z) + 2\pi i) = h(z)$ say, for some analytic function $h(z)$ which is completely arbitrary because this imposes no condition on $h()$, therefore in general $f(z) = \ln(h(z))$. The singular point(s) of $f()$ are only where $h(z) = 0$ or ∞ and at points z that are singular points of $h()$. At minimum there are singular points of $f()$ only where $h(z) = 0$ or ∞ when $h(z) = a + bz$ so that $h()$ has no singular points. This implies $z_0 = -a/b$ or ∞ and the only fixed singular point is at $z_0 = \infty$ with the other one having an arbitrary location, and the singular point by 6.9 has w_0 given by the solution of the single-value equation $w_0 = w_0 + 2\pi i$ which is $w_0 = \infty$. Therefore the singular points of $\ln()$ are at $(0, \infty)$ and (∞, ∞) and those of its inverse $\exp()$ are at $(\infty, 0)$ and (∞, ∞) .

Consider $w = (\ln(z))^2$. Can a similar analysis for this be done? We have $w = (\ln(z) + 2\pi i)^2$ then (78) is satisfied with $g_1(z, f) = (f^{1/2} + 2\pi i)^2$ and $g_2(s) = s$. Note that $g_1()$ is now not right-unique. Another analysis of this sort comes from $(\ln(z))^2 = (-\ln(z))^2 = (\ln(z^{-1}))^2$ i.e. Equation (78) with $g_1(z, f) = f$ and $g_2(z) = z^{-1}$, which shows that if in equation (78) either of $g_1()$ or $g_2()$ is not right-unique, this analysis may not be unique.

Consider $f(z) = z \ln(z)$, then

$$f(z) = f(z) + 2\pi iz. \quad (80)$$

This can be represented in terms similar to (78) with single valued $g_1()$ and $g_2()$ but this time g_1 has direct z dependence in addition to its dependence on $f()$ and $g_1(z, f) = 2\pi iz + f$ and $g_2(z) = z$. Conversely from (80), dividing by z and taking the exponential gives the tautology $\exp(f(z)/z) = \exp(f(z)/z)$, therefore this function can be any analytic function say $h(z)$. Therefore $f(z)/z = \ln(h(z))$ and $f(z) = z \ln(h(z))$. The singular points of $f()$ are at any point where $h(z) = 0$ or ∞ or at any point that is a singular point of $h()$. This gives at minimum, where $h(z) = a + bz$ with $b \neq 0$, singular points at $z = -a/b$ and $z = \infty$.

It seems paradoxical to say that $z \ln(h(z))$ is the general solution of (80) because (80) just states that whatever the multivalued function $f(z)$ is, if it has any value w at some point z , then at that point it also has the values $w + 2\pi in z$ for all $n \in \mathbb{Z}$. In fact $z \ln(h(z))$ can be any analytic function $f()$ provided $h(z) = \exp(\frac{f(z)}{z})$ and (80) holds in the multivalued sense. Nevertheless the use of the term “general solution” in this and other cases does seem convenient.

Suppose $f(z) = (\ln(z))^k$. Introduce the auxiliary function $g_2(z) = z^p$ then $f(g_2(z)) = (\ln(z^p))^k = p^k f(z)$ so (78) holds with $g_1(z, f) = fp^{-k}$, and Lemma 5.7 characterises $g_2()$. Alternatively, if only $f(g_2(z)) = p^k f(z)$ and $g_2(z) = g_2(e^{2\pi i/p} z)$ then this is a set of defining equations for $f()$ involving two instances of (78) and linear functions only, one to characterise $g_2()$ and the other to define $f()$.

8 The relationship between $g_2()$ and the type of singular points of $f()$ satisfying (107)

Consider the role played by $g_2()$ and its derivatives at an intersection point z_1 which is a solution of $g_2(z) = z$. This as will be seen controls to leading order the behaviour of $f(z)$ in the neighbourhood of the singular point at z_1 provided $f(z)$ satisfies (107) where $g_2()$ is as in (107). First consider an arbitrary value of $g_2'(z_1)$. For $z \approx z_1$, $g_2(z) \approx g_2(z_1) + (z - z_1)g_2'(z_1) = z_1 + (z - z_1)g_2'(z_1)$ therefore $f(z) \approx f(z_1 + (z - z_1)g_2'(z_1))$. Put $z = z_1 + \delta$ and treating this as an equality then $f(z_1 + \delta) = f(z_1 + \delta g_2'(z_1))$. A change of variable can now be made so as to relate this equation to $f(z) = f(z) + 2\pi i$ with its known solution. Let $w = \ln(\delta) = \ln(z - z_1)$ and the new function f^* by $f^*(w) = f(z)$ then $f^*(w) = f^*(w + \ln g_2'(z_1))$. Now let $w = \alpha t$ and $f^+(t) = f^*(w) = f(z)$ then $f^+(t) = f^+\left(t + \frac{\ln g_2'(z_1)}{\alpha}\right)$. Then choose α so that $\ln(g_2'(z_1))/\alpha = 2\pi i$ i.e.

$\alpha = \frac{\ln(g_2'(z_1))}{2\pi i}$ then $f^+(t) = h(\exp(t))$ i.e.

$$f(z) = f^*(w) = h(\exp(w/\alpha)) = h\left((z - z_1)^{\frac{2\pi i}{\ln(g_2'(z_1))}}\right). \quad (81)$$

This is the asymptotic behaviour of $f()$ for z close to z_1 where $h()$ is an arbitrary analytic function. This works provided $g_2'(z_1) \neq 0$ which is clearly a special case.

Now suppose $g_2'(z_1) = 0$ but $g_2''(z_1) \neq 0$. Then $g_2(z) \approx g_2(z_1) + \frac{(z-z_1)^2}{2}g_2''(z_1)$ then $f()$ satisfies $f(z) = f\left(z_1 + \frac{(z-z_1)^2}{2}g_2''(z_1)\right)$. Now put $k(\delta) = f(z_1 + \delta)$ where as before $\delta = z - z_1$ then $k(\delta) = k(\delta^2 g_2''(z_1)/2)$. Introduce $k^*(\cdot)$ by $k(\delta) = k^*(\ln(\delta))$ then $k^*(\ln(\delta)) = k^*(2 \ln \delta + \ln(g_2''(z_1)) - \ln(2))$. Introduce w by $w = \ln \delta$ then $k^*(w) \approx k^*(2w)$ because as $\delta \rightarrow 0$, $|w| \rightarrow \infty$ so the other terms can be asymptotically ignored. Now introduce $k^+(\cdot)$ by $k^+(\ln(x)) = k^*(x)$ then $k^+(\ln w) = k^+(\ln w + \ln 2)$ so $k^+(u) = k^+(u + \ln 2)$ where $u = \ln w$. Now let $t()$ be defined by $t(u\beta) = k^+(u)$ then $t(u\beta) = t(u\beta + \beta \ln(2))$. Choosing β to be $\beta = \frac{2\pi i}{\ln(2)}$ then $t(x) = t(x + 2\pi i)$ from which $t(x) = h(\exp(x))$. Undoing all these transformations now shows that $t(x) = t(u\beta) = k^+(u) = k^+(\ln(w)) = k^*(w) = k(\delta) = f(z_1 + \delta) = f(z)$ and $h(\exp(x)) = h(\exp(\beta u)) = h(\exp(\beta \ln(w))) = h(w^\beta) = h([\ln(z - z_1)]^\beta)$ so finally

$$f(z) = h\left([\ln(z - z_1)]^{\frac{2\pi i}{\ln 2}}\right) \quad (82)$$

where this result will only be asymptotically correct as $z \rightarrow z_1$. Note that $g_2''(z_1)$ is not involved.

From (81) $g_2'(z_1) = 1$ is obviously also a special case needing separate treatment. Then $g_2(z) \approx z + \frac{(z-z_1)^2}{2}g_2''(z_1)$ and the equation to be solved is $f(z) = f\left(z + \frac{(z-z_1)^2}{2}g_2''(z_1)\right)$. Putting $z = z_1 + \delta$ and introducing $f^*(\delta) = f(z_1 + \delta)$ gives

$$f^*(\delta) = f^*\left(\delta + \frac{\delta^2}{2}g_2''(z_1)\right). \quad (83)$$

Introduce the new variable k by $k\left(\delta + \frac{\delta^2}{2}g_2''(z_1)\right) - k(\delta) = \Delta$ so that the iteration of (83) is transformed to an arithmetic progression, then for small δ , $\frac{\delta^2}{2}g_2''(z_1)k'(\delta) = \Delta$ which can be integrated and inverted to give $\delta = -\frac{2\Delta}{kg_2''(z_1)}$.

Then $f^*\left(\frac{-2\Delta}{kg_2''(z_1)}\right) = f^*\left(\frac{-2\Delta}{kg_2''(z_1)} + \frac{2\Delta^2}{k^2g_2''(z_1)}\right)$. Introducing $f^+(k) = f^*(\delta)$ this can be written in terms of $f^+(\cdot)$ as $f^+(k) = f^+\left(\frac{\frac{-2\Delta}{g_2''(z_1)}}{\left(\frac{-2\Delta}{kg_2''(z_1)} + \frac{2\Delta^2}{k^2g_2''(z_1)}\right)}\right)$ which simplifies to $f^+(k) = f^+\left(\frac{k^2}{k-\Delta}\right) \approx f^+(k + \Delta)$. Let $g()$ be given by $g(l) = f^+(k)$ where

$k = l/\alpha$ then $g(l) = g(l + \alpha\Delta)$ and choosing $\alpha\Delta = 2\pi i$ then $g(l) = h(\exp(l))$ where $h(\cdot)$ is arbitrary and this implies

$$f(z) = h\left(\exp\left(-\frac{4\pi i}{g_2''(z_1)(z - z_1)}\right)\right) \quad (84)$$

asymptotically as $z \rightarrow z_1$.

This result can be generalised as follows. Suppose $g_2'(z_1) = 1$ and $g_2^{(n)}(z_1) = 0$ for $2 \leq n \leq m - 1$ and $g_2^{(m)}(z_1) \neq 0$ for $m \geq 2$. Then $g_2(z) = z + \frac{(z-z_1)^m}{m!}g_2^{(m)}(z_1) + O(z-z_1)^{m+1}$. In terms of $f^*(\cdot)$ and δ as above, $f(z) = f(g_2(z))$ becomes $f^*(\delta) = f^*\left(\delta + \frac{\delta^m g_2^{(m)}(z_1)}{m!} + O(\delta^{m+1})\right)$. This can be iterated and if k is chosen such that $k\left(\delta + \frac{\delta^m g_2^{(m)}(z_1)}{m!}\right) = k(\delta) + \Delta$ which can be approximated by $k'(\delta)\frac{\delta^m g_2^{(m)}(z_1)}{m!} = \Delta$ which integrates to $k(\delta) = \frac{-\Delta m!}{(m-1)\delta^{m-1}g_2^{(m)}(z_1)}$, then the iteration is an arithmetic progression and $f^*(\delta) = f^+(k) = f^+(k + \Delta)$. Therefore similarly to the above,

$$f(z) = h\left(\exp\left(\frac{-2\pi i m!}{(m-1)g_2^{(m)}(z_1)(z - z_1)^{m-1}}\right)\right) \quad (85)$$

asymptotically as $z \rightarrow z_1$.

9 Some interesting examples

Another example is

$$f(z) = (f(z))^{1/2} \quad (86)$$

with a singular point where $f(z) = 0$, which is a special case of (105) in which $g_1(\cdot)$ is not single valued. Taking natural logarithms twice gives

$$\ln \ln(f(z)) = \ln(1/2) + \ln \ln(f(z)) \quad (87)$$

and so

$$\frac{2\pi i}{\ln(2)} \ln \ln(f(z)) = -2\pi i + \frac{2\pi i}{\ln(2)} \ln \ln(f(z)) \quad (88)$$

so

$$\exp\left(\frac{2\pi i}{\ln(2)} \ln \ln(f(z))\right) \quad (89)$$

is arbitrary so call it $h(z)$ then

$$f(z) = \exp\left(\exp\left(\frac{\ln(2)}{2\pi i} \ln(h(z))\right)\right). \quad (90)$$

The function $f()$ can only have a singular or inversion point where $h()$ has singular or inversion point(s) or where $h(z) = 0$ or ∞ so $f(z) = 0$ or ∞ . This log-like singularity from (86) is characterised by the equations

$$\begin{aligned} g_1(z) &= -g_1(z) \\ f(z) &= g_1(f(z)) \end{aligned} \quad (91)$$

for the multivalued functions $f()$ and $g_1()$, where $g_1()$ is the *simplest solution*. If $f()$ is also the *simplest solution* then

$$f(z) = \exp \left(\exp \left(\frac{\ln(2)}{2\pi i} \ln(a + bz) \right) \right) \quad (92)$$

where a and b are constants.

Next consider

$$f(z) = f(z^2)/2. \quad (93)$$

This is a special case of (105) in which the condition for a singular point is more complicated than for (107) for which the condition for a singular point would give

$$z = g_2(z) = z^2 \quad (94)$$

determining more than one such point i.e. $z = 0, 1$. The effect of the extra factor of 2 complicates this a bit but this is still clearly true. Because $g_2()$ is not left-unique, (94) relates new singular points to other points already known to be singular points. In this example the singular points are dense on the unit circle because these are points for which $z^{(2^k)} = 1$ for arbitrarily large k . It follows that $f(z) = f(z^2)/2 = f(z^4)/4 = \dots f(z^{(2^k)})/2^k$ so if $z = re^{i\theta}$, $f(re^{i\theta}) = f((re^{i\theta})^{2^k})/2^k$ for all $k > 0$. For fixed r and θ suppose $\theta + 2\pi p \approx 2^k \theta$ where $p, k \in \mathbb{N}$ then $f(r^{(2^k)} e^{i\theta}) \approx 2^k f(re^{i\theta})$. Putting $R = r^{(2^k)}$ gives $f(Re^{i\theta}) \approx \frac{\ln(R)}{\ln(r)} f(re^{i\theta})$. The log dependence on R behaviour at large R and the positions (z) of the singular points may suggest the following formula

$$f(z) = \int_0^{2\pi} d\theta \log_2 |z - e^{i\theta}|. \quad (95)$$

for a solution of (93). Its proof is as follows

$$\begin{aligned} f(z^2) &= \int_0^{2\pi} d\theta \log_2 |z^2 - e^{i\theta}| = \int_0^{2\pi} d\theta \log_2 (|z + e^{i\theta/2}| |z - e^{i\theta/2}|) \\ &= \int_0^{2\pi} d\theta \log_2 |z + e^{i\theta/2}| + \int_0^{2\pi} d\theta \log_2 |z - e^{i\theta/2}| \\ &= 2 \int_0^{\pi} d\theta \log_2 |z + e^{i\theta}| + 2 \int_0^{\pi} d\theta \log_2 |z - e^{i\theta}| \\ &= 2 \int_{\pi}^{2\pi} d\theta \log_2 |z + e^{i(\theta-\pi)}| + \dots \\ &= 2 \int_{\pi}^{2\pi} d\theta \log_2 |z - e^{i\theta}| + \dots \\ &= 2 \int_0^{2\pi} d\theta \log_2 |z - e^{i\theta}| = 2f(z) \end{aligned} \quad (96)$$

This example has really peculiar properties because $f(z)$ is ∞ on the unit circle and this appears to isolate the function into two regions that can behave somewhat independently because (94) is satisfied for $f()$ replaced by $af()$ where $a \in \mathbb{C}$ and clearly any two different values of a can be chosen inside and outside the unit circle and the solutions can be described as having a natural boundary on the unit circle. [This doesn't work for finite prescribed values because if finite values are prescribed on any closed contour the Cauchy integral formula determines a function that is everywhere analytic and finite, uniquely inside it, but does it work for the outside region?] This is an example that divides $\overline{\mathbb{C}}$ into two domains of holomorphy [Encyclopedia of Mathematics] that overlap only on the unit circle.

Next follows an intriguing example where the condition for a singular point (an equation of the type (107)) determines two of them and the solutions found satisfy an additional equation of the type (105). Suppose $g_2(z) = \frac{a+bz}{c+z}$. Then $g_2(z) = z$ is a quadratic equation with solutions say z_1 and z_2 such that $z_1 + z_2 = b - c$ and $z_1 z_2 = -a$ and $g_2(z)$ can be written as $g_2(z) = \frac{-z_1 z_2 + bz}{b - z_1 - z_2 + z}$. However in this case, $g_2()$ is left-unique and single valued so only two singular points arise as a result of (107) which becomes in this case

$$f(z) = f\left(\frac{bz - z_1 z_2}{b - z_1 - z_2 + z}\right). \quad (97)$$

Therefore by lemma 6.7 solutions of (97) have singular points at z_1 and z_2 .

Using methods similar to those used in deriving (81) it possible to formally derive

$$f(z) = h_k \left(\sum_{n \in \mathbb{Z}} c_n \exp \left(\frac{2\pi i (\ln(z - z_k) + 2n_1 \pi i)}{\ln\left(\frac{z_1 - b}{z_2 - b}\right) + 2n\pi i} \right) \right) \quad (98)$$

for $k = 1, 2$ where $h_1()$ and $h_2()$ are arbitrary functions. By trial and error, the following are possible solutions of (97):

$$f(z) = \left(c_n \frac{z - z_1}{z - z_2} \right)^s \quad (99)$$

where $s = \frac{2\pi i}{\ln\left(\frac{z_1 - b}{z_2 - b}\right) + 2n\pi i}$ and $n \in \mathbb{Z}$. It is easy to show that

$$\frac{g_2(z) - z_1}{g_2(z) - z_2} = \frac{(z_1 - b)(z_1 - z)}{(z_2 - b)(z_2 - z)}. \quad (100)$$

Therefore

$$f(g_2(z)) = \left(c_n \frac{z - z_1}{z - z_2} \right)^s \left(\frac{b - z_1}{b - z_2} \right)^s. \quad (101)$$

The extra factor is $\left(\frac{b-z_1}{b-z_2}\right)^s$ can be written (including all its possible values) as

$$\exp(s \ln(t)) = \exp\left(\frac{\ln(t) \times 2\pi i}{\ln(t) + 2n\pi i}\right) = \exp\left(\left(\frac{\ln(t) + 2n_1\pi i}{\ln(t) + 2n\pi i}\right) \times 2\pi i\right) = E_{n_1, n} \tag{102}$$

where $n_1, n \in \mathbb{Z}$ for some specific value of $\ln(t)$ and where $t = \frac{b-z_1}{b-z_2}$. Increasing n_1 by 1 adds $\frac{2\pi i \times 2\pi i}{\ln(t) + 2n\pi i}$ to the argument of $\exp()$ multiplying the whole expression by $\exp\left(\frac{-4\pi^2}{\ln(t) + 2n\pi i}\right)$ and $E_{n, n} = 1$. From these it follows that $E_{n_1, n} = \exp\left(\frac{4\pi^2(n-n_1)}{\ln(t) + 2n\pi i}\right)$. Therefore

$$f(g_2(z)) = \left(c_n \frac{z - z_1}{z - z_2}\right)^{\frac{2\pi i}{\ln\left(\frac{z_1-b}{z_2-b}\right) + 2n\pi i}} \exp\left(\frac{4\pi^2(n - n_1)}{\ln\left(\frac{b-z_1}{b-z_2}\right) + 2n\pi i}\right). \tag{103}$$

From (107)

$$f(z) = \exp\left(\frac{2\pi i}{\ln(t) + 2n\pi i} \times \ln\left(c_n \frac{z - z_1}{z - z_2}\right)\right). \tag{104}$$

Taking this continuously round a small circuit C_1 anticlockwise round z_1 given by $z = z_1 + \epsilon e^{i\theta}$ for $0 \leq \theta \leq 2\pi$ where ϵ is a very small positive real number gives $f(z) = \exp\left(\frac{2\pi i}{\ln(t) + 2n\pi i} \ln\left(\frac{c_n e^{i\theta}}{z - z_2}\right)\right) = \exp\left((\ln(c_n) + i\theta - \ln(z - z_2)) \frac{2\pi i}{\ln(t) + 2n\pi i}\right)$. The difference over the path C_1 of the argument of $\exp()$ is $\frac{2\pi i \cdot 2\pi i}{\ln(t) + 2n\pi i}$ so the factor associated with doing C_1 is $\exp\left(\frac{-4\pi^2}{\ln(t) + 2n\pi i}\right)$ i.e. $f(z)$ satisfies $f(z) = f(z) \exp\left(\frac{-4\pi^2}{\ln(t) + 2n\pi i}\right)$. This can be applied to write (103) as (99) verifying the assumed form of $f()$ though this is probably not its most general form. Doing the same thing for a small circuit C_2 anticlockwise round z_2 gives the equivalent result $f(z) = f(z) \exp\left(\frac{4\pi^2}{\ln(t) + 2n\pi i}\right)$.

10 Simplest solutions of the equations defining singular points

***** This section seems as if there are some very important results to be found but it needs quite a lot of work yet *****

*'s indicate likely theorems that have not yet been proved.

Let the binary relation \rightarrow on analytic functions be defined by $f() \rightarrow g() \Leftrightarrow$ there exists an analytic function $h()$ such that $f() = h(g())$. Then the relation \rightarrow that points towards the simpler function is reflexive and transitive. Also

Theorem 10.1. *If $f() \rightarrow g()$ and $g() \rightarrow f()$ then $f(z) = \frac{a+bg(z)}{c+dg(z)}$ for some finite constants $a, b, c, d \in \mathbb{C}$.*

Proof. Suppose $f() \rightarrow g()$ and $g() \rightarrow f()$ then $f() = h_1(g())$ and $g() = h_2(f())$ for some analytic functions $h_1()$ and $h_2()$, and therefore $f() = h_1(h_2(f()))$ i.e. $h_1(h_2()) = I()$ which has no singular point. By Theorem 5.6 $h_1()$ can have no singular point and is therefore a bilinear function and so $f(z) = \frac{a+bg(z)}{c+dg(z)}$. \square

Suppose a set S of analytic functions is such that if $f() \in S$ then $h(f()) \in S$. Then this set is determined by the set $R \subseteq S$, the root functions, such that for any analytic function $f()$ in S there exists a member $g() \in R$ such that $f() \rightarrow g()$. Such a set will be called a rooted set. Suppose a single root function $r()$ acts a root for S i.e. $\forall f() \in S [f() \rightarrow r()]$. Suppose another function $r_1()$ also has this property, then $\forall f() \in S [f() \rightarrow r_1()]$ and in particular $r() \rightarrow r_1()$. Likewise $r_1() \rightarrow r()$. Then by Theorem 10.1 $r(z) = \frac{a+br_1(z)}{c+dr_1(z)}$. Such a rooted set will be called singly-rooted. Thus the root functions associated with a singly rooted set are related by a bilinear transformation.

From Theorem 10.1 any root function $k()$ is unique up to a bilinear function or transformation (also known as a Möbius transformation or a linear fractional transformation) i.e. $k_1(z) = \frac{a+bk(z)}{c+dk(z)}$ so a root function is actually a set of functions each member of which is related to any other member like this for some set of values $a, b, c, d \in \mathbb{C}$ such that $ad - bc \neq 0$. The terminology below will for simplicity refer to this special set just as a single function, the root function.

Lemma 10.2. *Every analytic function is in the set rooted by a left-unique function.*

Proof. If $g()$ is left-unique then $f(z) = f(g^{o-1}(g(z)))$ so $f() \rightarrow g()$. \square

Theorem 10.3. *If $gof() \equiv f(g()) = g(f())$ and $g()$ is left-unique and right-unique then $f() \oplus g() = gof()$.*

Proof. The condition on $g()$ gives $g^{o-1}(g()) = I$ and $g(g^{o-1}()) = I$ and also $f(g()) = g(f())$. Suppose $l(z) = h_1(f(z))$ and $l(z) = h_2(g(z))$ for some arbitrary analytic functions $h_1()$ and $h_2()$. Then $f(z) = f(g^{o-1}(g(z))) = g^{o-1}(g(f(z))) = g^{o-1}(f(g(z)))$. Therefore $f(g^{o-1}(w)) = g^{o-1}(f(w))$ generally [where $w = g(z)$] and $l(z) = h_1(f(g^{o-1}(g(z)))) = h_1(g^{o-1}(f(g(z)))) = h_3(f(g(z)))$ where $h_3() = h_1(g^{o-1}())$. According to the criterion for a root function, $f(g())$ is the required root function for the set of possible analytic functions $l()$. \square

This shows that the general solution is determined by one particular solution. Could $f(z)+g(z)$ be the root function? If so, then $f(z)+g(z) = h_1(f(z))$ for some function $h_1()$. Then $h_1(z) = z+g(f^{o-1}(z))$ formally, and this requires $f^{o-1}(f(z)) = z$ i.e. $f()$ is left-unique. Likewise for $f(z)+g(z) = h_2(g(z))$ showing that this also requires $g()$ to be left-unique. This shows that $f() + g() \in S$ if $f()$ and $g()$ are left-unique. $f() + g()$ is to be the root function also requires that for any function $l()$, $l() \rightarrow f()$ and $l() \rightarrow g() \Rightarrow l() \rightarrow f() + g()$.

Suppose now that $f()$ is not left-unique for example $f(z) = l(z)^2$ then the problem is to find the intersection $h_1(l(\cdot)^2) \cap h_2(g(\cdot))$.

A common type of equation defining behaviour around a singular point is

$$f(z) = g_1(f(g_2(z))) \quad (105)$$

where $g_1()$ and $g_2()$ are right-unique functions. The more general form

$$f(z) = g_1(z, f(g_2(z))) \quad (106)$$

occurs later. Most of the examples above are actually special cases of

$$f(z) = f(g_2(z)) \quad (107)$$

[could this be generalised to $f(z) = f(z, g_2(z))$?] or

$$f(z) = g_1(z, f(z)) \quad (108)$$

which are themselves special cases of (106). Equation (107) can be iterated to give $\forall n \in \mathbb{N}[f(z) = f(g_2^{on}(z))]$ which is equivalent to (107). This can be expressed as

$$\exists n \in \mathbb{N}[z_2 = g_2^{on}(z_1)] \Rightarrow f(z_1) = f(z_2). \quad (109)$$

If $l()$ is any function then from (109) it follows that $\exists n \in \mathbb{N}[z_2 = g_2^{on}(z_1)] \Rightarrow l(f(z_1)) = l(f(z_2))$ i.e. if $f()$ is a solution of (107) so is $l(f())$. If $l()$ is analytic so will be $l(f())$.

Let $f_s()$ be a special solution of (107) that satisfies in addition the converse of (109) i.e.

$$f_s(z_1) = f_s(z_2) \Rightarrow \exists n \in \mathbb{N}[z_2 = g_2^{on}(z_1)]. \quad (110)$$

This introduces the equivalence relation \sim defined by $z_1 \sim z_2 \Leftrightarrow \exists n \in \mathbb{N}[z_2 = g_2^{on}(z_1) \text{ or } z_1 = g_2^{on}(z_2)]$ and states that the values of $f_s()$ are in one to one correspondence with the equivalence classes of \sim . Any solution $f(z)$ of (107) is a function of the equivalence classes i.e. its value is the same for each member of the same equivalence class, (but regarded as a function of the equivalence classes, is not necessarily left-unique), therefore it can be written as a function of a $f_s()$ i.e. $f(z) = h(f_s(z))$. Until now $f()$ and $f_s()$ were tacitly assumed to be right-unique but this does not have to be the case because pairs or sets of values of $f_s()$ will then be in left-unique correspondence with the equivalence classes of \sim . Because the functions $f()$ and $f_s()$ are analytic, $h()$ will be also. This works if $f_s()$ exists.

The extension to multiple simultaneous equations of type (107) will also depend on the analogous existence theorem. Another approach is to replace the equations by the limit of some equations for algebraic functions for which the uniqueness of the solution given all the singular behaviours at each of the singular points has been established.

Because of lemma 6.9, every solution of (107) has a singular point at points z where $z = g_2(z)$ for example $(f_s(z))^{-2}$. This example also has a singular point where $f_s(z) = 0$ and suggests that of all the analytic solutions of (107), the special solutions $f_s()$ that also satisfy (110) have singular points only where $z = g_2(z)$. To prove this suppose $z_1 \neq g_2(z_1)$. The condition for $f_s()$ to have no singular point at $P, (z_1, f_s(z_1))$, using (109) and (110), is that there is a neighbourhood N of P such that for all points $(z_2, f_s(z_2))$ and $(z_3, f_s(z_3)) \in N$, $z_2 = z_3 \Leftrightarrow z_2 \sim z_3$. The last condition reduces to $\exists n \in \mathbb{N}[z_2 = g_2^{on}(z_3)] \Rightarrow z_2 = z_3$. To establish this it is sufficient to choose N so small that if z_3 is included by being sufficiently close to z_1 that none of $g_2(z_3), g_2(g_2(z_3))$ etc. are included i.e. the image of N under $g_2()$ must not overlap N itself. There is another case i.e. $g_2^k(z_1) = z_1$ where $k \geq 2$ where the proposition is also true.

Therefore these special fundamental solutions of (107) that also satisfy (110) will be called the *simplest solutions* of (107). Let $f_s^*()$ be another function that satisfies the conditions on $f_s()$ above then $f_s^*(z) = h^*(f_s(z))$. If $h^*()$ is any function without a singular or an inversion point (i.e. a linear function) then by Lemma (6.11) $f_s^*()$ will satisfy this condition i.e. the set of *simplest solutions* of (107) must include $a + bf_s(z)$ if $f_s(z)$ is included. Can there be any more? Any other such solution must take this form with a different function $h^*()$ that will be nonlinear and must have at least two singular or inversion points somewhere and $h^*(a + bf_s(z))$ must have no singular point except when z satisfies $z = g_2(z)$ for all $a, b \in \mathbb{C}$. This is impossible because the argument of $h^*()$ can then take any value, so this proves

Theorem 10.4. *The set of simplest solutions to (107) i.e. those that also satisfy (110) have singular points only where $z = g_2(z)$ where $g_2()$ is as in (107). This set is the set $a + bf_s(z)$ for arbitrary $a, b \in \mathbb{C}$ if $f_s(z)$ is itself a simplest solution of (107). Any solution to (107) can be written as $h(f_s(z))$ for some simplest solution $f_s(z)$ for some analytic function $h()$.*

It is interesting to note that in the examples the index set \mathbb{N} can sometimes be replaced by a finite set. If $g_2()$ is not a linear function the equation $z = g_2(z)$ that determines the singular point could have many solutions, and $g_2()$ itself could be described by another equation of the type (107) or (108) etc.. In such a case the original equation (107) together with other similar equations to determine $g_2()$ could determine behaviour at a set of singular points simultaneously. In such a case it might be a good idea to try to solve for the singular points and then with $g_2()$ replaced by linear functions that give the same singular points, analyse each separately using the results in Section 8 and then try to reconstruct the original function but note example (93) indicating that in this case an infinite number of singular points can sometimes occur.

A similar argument to that applied to (107) can be applied to (108) giving

the iteration as

$$\forall n \in \mathbb{N}[f(z) = g_1^{on}(z, f(z))] \tag{111}$$

which is equivalent to (108) where $g_1()$ appears n times in $g_1^{on}(z, f(z)) = g_1(z, g_1(z, g_1(z, \dots g_1(z, f(z)) \dots))$.

*****The *simplest solution* satisfies

$$z_1 = z_2 \Rightarrow \exists n \in \mathbb{N}[f(z_1) = g_1^{on}(z_2, f(z_2))] \tag{112}$$

where the justification is similar i.e. (111) gives all the values of the function $f(z)$ that are *determined* by one value of $f(z)$ whereas (112) states that for any value of $f(z)$ at the same point z , it must be one of the values in (111).

Now consider iteration applied to (106) which gives

$$\begin{aligned} f(z) &= g_1(z, g_1(g_2(z), f(g_2^o(z)))) = \dots = \\ &g_1(z, g_1(g_2(z), g_1(g_2^o(z), g_1(g_2^o(z), g_1(g_2^o(z), \dots)) \dots))) = [g_1((), f(g_2()))^{on}(z) \end{aligned} \tag{113}$$

where g_1 appears n times in this expression. Now suppose $g_2^{on}()$ is the identity function : $z \rightarrow z$ then

$$f(z) = g_1(z, g_1(g_2(z), g_1(g_2^o(z), \dots g_1(g_2^{o(n-1)}(z), f(z)) \dots))). \tag{114}$$

This is last expression depends independently on z and $f(z)$ through the functions $g_1()$ and $g_2()$ and can therefore be written as $k(z, f(z))$ i.e. (114) can be written in the form (108) for different $g_1()$ and $g_2()$. Also it is conceivable that (114) for some value of n takes the simpler form (107) again for different $g_1()$ and $g_2()$. In either of these cases the *simplest solution* of the respective iterated form of (106) can be defined as above. If this can be done for both cases the following example suggests this might define the *simplest solution* for (106) itself.

There are many results that can be obtained relating the solution sets of (105) with different values of $g_1()$ and $g_2()$. If (105) holds then the same relationship holds with $f()$ replaced by $k(f(l()))$, $g_1()$ replaced by $k(g_1(k^{o-1}()))$ and $g_2()$ replaced by $l^{o-1}(g_2(l()))$. Making these substitutions gives the same relationship with the function $k()$ applied to both sides and expressed in terms of the independent variable w given by $z = l(w)$. For example suppose $k(z) = az + b$ and $l(z) = cz + d$ then the function $f^*(z) = k(f(l(z))) = af(cz + d) + b$ satisfies $f^*(z) = g_1^*(f^*(g_2^*(z)))$ i.e. (105) with $g_1^*(z) = ag_1((z - b)/a) + b$ and $g_2^*(z) = (g_2(cz + d) - d)/c$.

If in equation (105) $g_1^{o-1}()$ is applied to both sides and the result expressed in terms of the variable $w = g_2(z)$ then the same relationship holds with $g_1()$ replaced by $g_1^{o-1}()$ and $g_2()$ replaced by $g_2^{o-1}()$.

The inverse functions of both sides of Equation (105) again give an equation of the same form showing that f^{o-1} satisfies the equation of the same form but with $g_1()$ replaced by $g_2^{o-1}()$ and $g_2()$ replaced by $g_1^{o-1}()$.

In these general arguments, it has to be borne in mind that $f^{o-1}(f(z))$ could have several components and is not necessarily just the identity function as in section 1.

11 Further thoughts on solutions to equations (105) and (78)

Suppose a general solution of (78) is of the form

$$f(z) = F(h(G(z))) \quad (115)$$

for some fixed functions F and G where $h()$ is an arbitrary analytic function. Now suppose in addition to (78) that $f()$ has no singular points other than those required by (78) including no other conditions that could modify behaviour at the singular point(s) required by (78) i.e. at (z, f) such that

$$\begin{aligned} z &= g_2(z) \\ f &= g_1(z, f) \end{aligned} \quad (116)$$

Next consider the finite singular points of $h()$. Let $z = s$ be such a point then by (115) any point z such that $G(z) = s$ will be a singular point of $f()$. By assumption this cannot happen [unless this z with a value for f satisfies (116)]. An additional condition on $h()$ at the singular points would correspond to an additional condition on $f()$ which is assumed not to happen therefore $h()$ has no finite singular points and therefore $h(z) = a + bz$ and $f(z) = F(a + bG(z))$. This proves that

Theorem 11.1. *If the general solution of (78) is (115) and if $f()$ has no singular points other than those required by (78) including no other conditions that could modify behaviour at the singular point(s) required by (78) then $f(z) = F(a + bG(z))$.*

Consider examples where more than one singular point is analysed like this in detail and then I expect (105) or (78) will only asymptotically hold close to the corresponding singular point.

Consider for this a linear combination (LC) of the minimal solutions for each separate singular point. By minimal I mean solutions that have no other finite singular points (see below). This LC will have precisely the asymptotically defined behaviours at the singular points because all the other terms will not have a singular point at each of them. This I think can be generalised to nonlinear combinations, if the condition of minimality is dropped. i.e. find the general solution of a set of simultaneous asymptotically defined relations about a set of singular points as some arbitrary analytic function of basic solutions to them singly?

Questions about future research

- Is it possible and practical to use the closure properties to prove statements by a kind of induction i.e. if a statement is true for the constant function, and it is true for functions f and g implies it is true for their union, sum, product, $f'()$ (the derivative of $f()$), $\int_0^x f(s)ds$, composition, $f^{\circ-1}$ (inverse of f), etc. then it is true for the whole algebra of functions?
- Does the algebra include the solutions of differential equations? eg $f''(x) + k(x)f(x) = 0$ Perhaps this needs the limit of a sequence of functions to be included if the sequence is already included because any differential equation can be written as an integral equation which can be solved as the limit of an iteration. Then I think differentiation and integration can be removed from the closure operations.
- Can algebraic functions be characterised as those satisfying (106) with linear $g_1()$ and $g_2()$?
- What role is played by the operation of getting the *simplest solution* of (106) with given $g_1()$ and $g_2()$? How should this be done in general?
- Consider the binary operation which is the *simplest common solution* $f()$ of $f(z) = h_1(f_1(z))$ and $f(z) = h_2(f_2(z))$ for fixed $f_1()$ and $f_2()$, but arbitrary $h_1()$ and $h_2()$.
- What about the solutions of different equations say $f(z) = g_1(f(f(z)))$?

References

- [Churchill et al.] Churchill R.V., Brown J.W., Verhey R.F., Complex Variables and Applications, Third Edition, McGraw-Hill Kogakusha 1974
- [Nixon2013] Nixon J., Theory of algebraic functions on the Riemann Sphere *Mathematica Aeterna* Vol. 3, 2013, no. 2, 83-101
<https://www.longdom.org/articles/theory-of-algebraic-functions-on-the-riemann-sphere.pdf>
- [Encyclopedia of Mathematics] *
<https://encyclopediaofmath.org/wiki/Analytic-function>