Lecture Notes in Mathematics

Number 2

RUDIMENTS OF RIEMANN SURFACES

B. Frank Jones, Jr.

Houston, Texas
1971

RUDIMENTS OF RIEMANN SURFACES

B. Frank Jones, Jr.

Rice University

1971

PREFACE

In the spring semester, 1969, I taught a course at Rice University on Riemann surfaces. The students were primarily seniors who had taken one semester of complex variables and had been exposed at least to the language of general topology. I made detailed lecture notes at the time, and this volume contains those notes with minor changes.

The purpose of the course was to introduce the various ideas of surfaces, sheaves, algebraic functions, and potential theory in a rather concrete setting, and to show the usefulness of the concepts the students had learned abstractly in previous courses. As a result, I discussed the material carefully and leisurely, and for example did not even attempt to discuss the notions of covering surface, differential forms, Fuchsian groups, etc. Therefore, these notes are quite incomplete. For comprehensive treatments of the subject, please consult the bibliography.

I gratefully acknowledge some of the standard books which I consulted, especially M. H. Heins' _Complex Function Theory_, G. Springer's _Introduction to Riemann Surfaces_, and H. Weyl's _The Concept of a Riemann Surface_. Also, I relied heavily on L. Bers' lecture notes, _Riemann Surfaces_, and especially on his Lectures 15-18.

One of the students was Joseph Becker, to whom I owe special thanks. He helped and prodded me over and over and gave me tremendous encouragement.

Thanks also go to the typists, Janet Gordon, Kathy Vigil, and Barbara Markwardt, and to Rice University for publishing the notes.

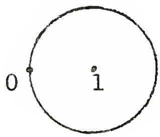
Houston, July 12, 1971

CONTENTS

is an extension of f. But this is not the kind of difficulty that we wish to consider.

Rather, the basic problem is that of multiple-valued "functions." Phrased in terms of continuations, there is not always a largest region to which a holomorphic function can be extended. As an example let $D = \{z: |z-1|<1\}$ and $f(z)$ = principal determination of

$$\log z$$



$$= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(z-1)^n}{n} ,$$

defined on D. Of course, we also have $f(z) = \log|z| + i \arg z$, where arg z is between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Now f can be extended to a holomorphic function on the plane $\mathbb{C}$ with the negative real axis removed, the extension being $\log|z| + i \arg z$, where $-\pi < \arg z < \pi$. But there are other regions which can be considered as largest regions of extension; e.g., the plane $\mathbb{C}$ with the positive imaginary axis removed and the extension being $\log|z| + i \arg z$, where $-\frac{3\pi}{2}<\arg z<\frac{\pi}{2}$.

It is admittedly frequently useful to "cut" the plane $\mathbb{C}$ along a line from 0 to ∞ as visualized in the above cases, and to consider there a single-valued "branch" or "determination" of log z, and such a technique is exploited e.g. in contour integrals.

But from the point of view of this course the cutting of $\mathbb{C}$ really enables one to evade the issue, which is namely how can one speak of log z and face up to its multiple-valuedness in a fearless way. And the same question for other functions. The answer given by Riemann is that the plane $\mathbb{C}$ is too deficient to admit such functions, so we consider other surfaces where functions can be defined which are single-valued and still exhibit the essential behavior of (in our example) log z.

Let us now consider an explicit method for building such a surface for log z. Take an infinite sequence of planes minus the origin, which are to be considered as distinct; call them $\mathbb{C}_n'$, where n is any integer. On $\mathbb{C}_n'$ define a function f_n by

$$f_n(z) = \log|z| + i \arg z + 2n\pi i ,$$

where $-\pi < \arg z \leq \pi$. Now we "glue" the planes $\mathbb{C}_n'$ in a reasonable way. This "gluing" is tantamount to defining a topology on the union of the (disjoint) sets $\mathbb{C}_n'$. To define this topology we shall describe a neighborhood basis of each point. For a point $z \in \mathbb{C}_n'$ which does not lie on the negative real axis a neighborhood basis shall consist of all open disks in $\mathbb{C}_n'$ with center at z. If $z \in \mathbb{C}_n'$ and z is a negative real number, a neighborhood basis shall consist of all sets

$$\{w \in \mathbb{C}_n' : |w-z|<\epsilon, \text{ Im } w \geq 0\} \cup \{w \in \mathbb{C}_{n+1}' : |w-z|<\epsilon, \text{ Im } w<0\} ,$$

where $0 < \epsilon < |z|$.

It is then easily checked that the set $S = \bigcup_{n=-\infty}^{\infty} C'_n$ becomes a topological space with a neighborhood basis for each point of S being described as above. Also, if f is the function from S to $\mathbb{C}$ which equals f_n on C'_n for each n, then f becomes a <u>continuous</u> function on S. Indeed, it suffices to check continuity at points $z \in C'_n$ which are negative real numbers. In the semidisk in C'_n depicted above $f = f_n$ takes values close to $\log|z| + i\pi + 2n\pi i$, and in the semidisk in C'_{n+1} $f = f_{n+1}$ takes values close to $\log|z| - i\pi + 2(n+1)\pi i$, so in the whole neighborhood of z f is close to $\log|z| + i\pi + 2n\pi i = f(z)$. Thus, f is continuous.

Thus, we have succeeded in defining a set S which carries a single-valued function f which obviously is closely related to log z. We shall later point out the essential feature of S which allows us to call it a Riemann surface (definition to be given in Chapter II).

We remark that it is easy to visualize S as a collection of planes glued together as indicated and forming in $\mathbb{R}^3$ an infinite spiral. The next surface we construct will not have so simple a form.

For this construction consider the function $z^{1/m}$, where m is an integer ≥ 2. Since each $z \neq 0$ has m distinct m^{th} roots, this is a multiple-valued "function." In order to treat this function consider distinct copies of the plane minus the origin, $C_1', C_2', \ldots, C_m'$. Define a function f_n on C_n' by the formula:

$$\text{if } z = re^{i\theta}, \quad r>0, \quad -\pi<\theta\leq\pi ,$$
$$f_n(z) = r^{1/m}e^{i\theta/m}e^{2i\pi(n-1)/m} .$$

Let $T = \bigcup_{n=1}^{m} C_n'$ and define a topology on T exactly as before, except that a neighborhood basis of a negative real number $z \in C_m'$ is treated a little differently. The same situation obtains as in the figure on p. 4, with C_{m+1}' replaced by C_1'. Note that an attempt to visualize T as a spiral in R^3 is doomed, since the "top" level C_m' has to be glued to the "bottom" level C_1' along their negative real axes, and this without crossing any of the intermediate levels $C_2', \ldots, C_{m-1}'$ and also without crossing the seam where C_1' is joined to C_2' (in case m = 2). As before, define a function f on T by the formula $f = f_n$ on C_n'. As in the figure on p. 4, if $z \in C_n'$ is a negative real number, then in the semidisk in C_n' f takes values close to $|z|^{1/m}e^{i\pi/m}e^{2i\pi(n-1)/m}$, and in the semidisk in C_{n+1}' f takes values close to $|z|^{1/m}e^{-i\pi/m}e^{2i\pi n/m}$, so f stays close to $|z|^{1/m}e^{i\pi(2n-1)/m} = f(z)$ in a neighborhood of z. And this holds even if n = m, in which case C_{n+1}'

is replaced by C_1'. Thus, f is continuous on T and gives a reasonable representation of $z^{1/m}$.

Now a very interesting addition can be made to T. Namely, consider each C_n' to have its origin replaced, but with the origins in each C_n representing a single point to be added to T. Thus, consider $T \cup \{0\}$ (the original set with one point added) and let a neighborborood basis of 0 consist of sets of the form

$$\{0\} \cup \bigcup_{n=1}^{m} \{z \in C_n' : |z| < \varepsilon\}$$

for $0 < \varepsilon < \infty$. Extend f by $f(0) = 0$. Then f is again continuous on $T \cup \{0\}$. In the very same way, the point ∞ can be added. Let

$$\tilde{T} = T \cup \{0\} \cup \{\infty\} ,$$

let a neighborhood basis of ∞ consist of sets of the form

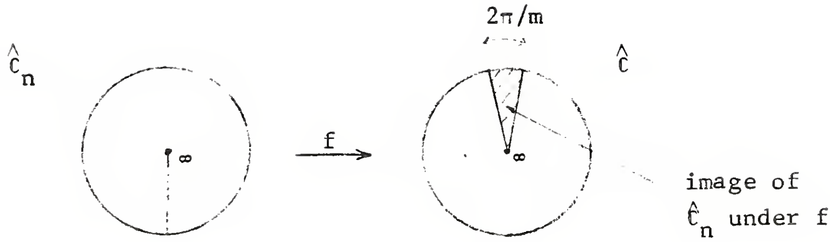
$$\{\infty\} \cup \bigcup_{n=1}^{m} \{z \in C_n' : |z| > \frac{1}{\varepsilon}\} ,$$

and let $f(\infty) = \infty$. Then f is a continuous function from $\tilde{T}$ to the extended complex plane (Riemann sphere) $\hat{C}$. Obviously the points 0 and ∞ are in some sense different from the other points in $\tilde{T}$. They are called <u>branch points</u>, and are said to have <u>order</u> m-1.

Although $\tilde{T}$ is somewhat difficult to visualize as situated in R^3, we shall now easily see that it is homeomorphic to the sphere $\hat{C}$! In fact, the mapping $f: \tilde{T} \to \hat{C}$ is a homeomorphism. We have shown that it is continuous;

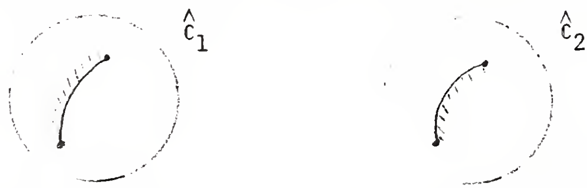
it is onto since every complex number is an m^{th} root; it is 1-1 since different complex numbers definitely have different m^{th} roots and also the same complex number $z \neq 0$ has m distinct m^{th} roots. General topology then shows f^{-1} is continuous since $\tilde{T}$ is compact and $\hat{\mathbb{C}}$ is Hausdorff; but it is quite easy to see directly that f^{-1} is continuous. Indeed, $f^{-1}(z)$ is essentially z^m (positioned on the correct $\mathbb{C}_n'$).

The nature of this homeomorphism and the geometry involved in the construction of $\tilde{T}$ are perhaps better seen when one considers the Riemann sphere $\hat{\mathbb{C}}$ instead of $\mathbb{C}$ as the basic region from which f is to be built. If one regards $\hat{\mathbb{C}}$ as the Euclidean sphere $\{(x,y,z)\colon x^2+y^2+z^2=1\}$ in $\mathbb{R}^3$ by means of stereographic projection and uses m distinct copies $\hat{\mathbb{C}}_1,\ldots,\hat{\mathbb{C}}_m$ with the gluing described above to be done along the meridians corresponding to the negative real axis, then an essentially equivalent surface $\tilde{T}$ is obtained. Now consider the action of the function f. On $\hat{\mathbb{C}}_n$ it is given by the determination f_n of the m^{th} root and maps $\hat{\mathbb{C}}_n$ onto a portion of $\hat{\mathbb{C}}$ cut off by two meridians which correspond to rays in the plane with an included angle of $2\pi/m$. In other words, it "spreads open" the cut in $\hat{\mathbb{C}}_n$ from a hole with 0 opening to a hole with $(1 - \frac{1}{m})2\pi$ opening. Here is a picture, a "top" view looking "down" on the north pole, ∞:

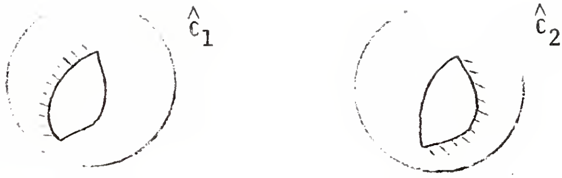


Thus, the image of $\tilde{T}$ under f consists of m "slices" of $\hat{C}$, and the gluing in $\tilde{T}$ shows that these slices of $\hat{C}$ are pieced together in such a way that $\tilde{T}$ is mapped homeomorphically onto $\hat{C}$.

If one is interested only in the topological properties of $\tilde{T}$, then the procedure discussed in the above paragraph can be considerably shortened by ignoring the specific nature of the cuts and of the function f. We illustrate with the case m = 2. Since we shall only discuss topological properties, we replace the cut along a meridian by any old cut on the sphere which looks reasonable, and take two copies of the sphere:



The gluing is to be done in such a way that the shaded areas are to be attached, as are the unshaded areas. The action of the function f is now replaced by a continuous opening of the two holes:



It is then obvious how to attach these spheres with holes; the resulting figure looks like

, a figure which is obviously homeomorphic to a sphere.

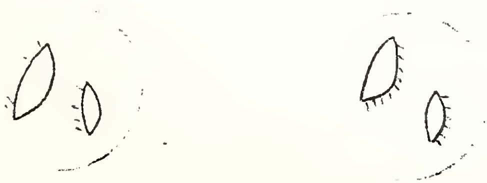
Now we shall briefly indicate the construction of some other Riemann surfaces. For example, suppose a and b are distinct complex numbers and consider the multiple-valued "function" $\sqrt{(z-a)(z-b)}$. The same procedure which works for $\sqrt{z}$ can be applied here if C or $\hat{C}$ is cut between a and b. In trying to define this function one finds that the sign changes when a circuit is made around either a or b, so two copies of $\hat{C}$ can be joined along the cut as before to provide a surface on which a function which is single-valued and has the properties of $\sqrt{(z-a)(z-b)}$ can be defined; the figure is exactly that which appears at the bottom of p. 8, where the two slits go from a to b on each sphere. Here it should be remarked that either branch of $\sqrt{(z-a)(z-b)}$ is meromorphic at ∞, since one branch is approximately z at ∞ and the other branch approximately -z. The branch points on the surface we have constructed are a and b, and the surface is again homeomorphic to $\hat{\mathbb{C}}$. However note that the function $\sqrt{(z-a)(z-b)}$ is not the homeomorphism in this case. Indeed, this function assumes every value in $\hat{\mathbb{C}}$ exactly twice. A natural homeomorphism in this case is the function on this surface corresponding to $\sqrt{\frac{z-a}{z-b}}$. Note in particular that if we begin with this function and two copies of $\hat{\mathbb{C}}$ cut from a to b, we obtain the same surface.

Using the same process, we shall now construct a Riemann surface which is not homeomorphic to a sphere. For this consider the expression $\sqrt{(z-a)(z-b)(z-c)}$, where a,b,c are distinct complex numbers. In order to attempt to define a single-valued function from this formula, consider two copies of the sphere each having two cuts, say from a to b and from c to ∞; these cuts should not intersect:

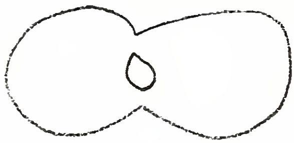
In defining continuously the square root in this case, a change of sign results in going around a, or b, or c, or ∞. The cuts we have provided prohibit this, and we also see just how to glue in order to obtain a continuous function: the shaded areas along the cuts from a to b are to be attached, and likewise along c to ∞. Now let $\tilde{S}$ denote the resulting surface with the four branch points a,b,c,∞ included, the topology being defined in the by now usual manner. This surface is <u>not</u> homeomorphic to a sphere. To see this we will exhibit a closed curve on $\tilde{S}$ which does not separate $\tilde{S}$ into two components. This is the curve shown on the left sphere

which encircles the cut from a to b. To see that this curve does not disconnect $\tilde{S}$ consider the typical example of the curve (shown by a dotted line) which connects two points which at first glance might be separated by the given closed curve.

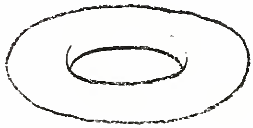
Probably the best way to see this topological property is to apply the method sketched on p. 8. After the first step we obtain the following spaces to be glued:



After the gluing, the resulting figure appears as shown:



This figure is clearly homeomorphic to a torus or a sphere with "one handle." The same topological type of surface arises from the function $\sqrt{(z-a)(z-b)(z-c)(z-d)}$, where a,b,c,d are distinct. The only difference is that the cuts on $\hat{\mathbb{C}}$ go from a to b and from c to d.



This same argument allows the treatment of the function $\sqrt{(z-a_1)(z-a_2)\ldots(z-a_m)}$, where $a_1,\ldots,a_m$ are distinct. Two copies of $\hat{\mathbb{C}}$ are used with cuts from a_1 to a_2, a_3 to a_4, etc. If m is even, the last cut is from a_{m-1} to a_m, and if m is odd, from a_m to ∞. The same gluing procedure gives a topological type as illustrated:

there are $\frac{m}{2}$ connecting tubes if m is even, $\frac{m+1}{2}$ if m is odd.

This is homeomorphic to a sphere with "handles":

there are $\frac{m-2}{2}$ or $\frac{m-1}{2}$ handles if m is even or odd, respectively. This is said to be a surface having genus equal to the number of handles.

Local coordinates.

In preparation for the definition of abstract Riemann surfaces to be given in the next chapter, we shall now examine a common property of all the surfaces we have constructed. Namely, each point on the surface has a neighborhood homeomorphic to an open subset of $\mathbb{C}$--the essential defining property for a surface. This assertion is of course completely trivial except where we have made cuts and where we have inserted branch points, for outside these exceptional points the neighborhoods can just be taken to be disks on the various copies of $\mathbb{C}$ and the homeomorphism essentially the identity mapping onto the same disk, now regarded as lying in some other fixed copy of $\mathbb{C}$. The situation for points on the cuts which are not branch points is not much more involved. Refer to the neighborhood defined and depicted on pp. 3,4; call this neighborhood $U(z)$ and let Δ be the disk $\{w \in \mathbb{C}: |w-z| < \varepsilon\}$. Then define

$$\varphi: U(z) \to \Delta$$

by the obvious relation

$$\varphi(w) = w.$$

The effect of φ is obviously to attach the two semi-disks used to make up $U(z)$. It is now trivial to check that each point which is not a branch point has a neighborhood homeomorphic to an open set (a disk) in $\mathbb{C}$, and this is true for all the surfaces we have constructed. If ∞ is not a branch point and does not lie on a cut, a neighborhood can be taken to be the complement of a large closed disk in the appropriate copy of $\hat{\mathbb{C}}$ and the mapping into $\mathbb{C}$ the function $\varphi(z) = z^{-1}$.

Now for the branch points. It should be no surprise that the branch points can be treated, for we have pointed out how the surface with branch points added is homeomorphic to a sphere or a sphere with handles (in the cases we have considered), making the neighborhoods of the branch points look not very special at all. Now we write down this homeomorphism explicitly in the case of the Riemann surface for $z^{1/m}$, since all the other branch points we have considered have the same behavior as is exhibited in this case (for $m = 2$). In fact, the homeomorphism is exactly the "function" $z^{1/m}$ (which has been made single-valued). In terms of the notation of p. 5, this is the function f. A similar construction works when the branch points at ∞ are considered.

Finally, consider how these various homeomorphisms are related. That is, suppose given two overlapping neighborhoods U_1 and U_2 on the surface with corresponding homeomorphisms φ_1 and φ_2. Then the function $\varphi_2 \circ \varphi_1^{-1}$ is defined on an open subset of $\mathbb{C}$ and has values in another open subset of $\mathbb{C}$, and is clearly a homeomorphism. The thing to be noted is that it is <u>holomorphic</u>. Except where U_1 or U_2 involves a branch point this is trivial, as the map $\varphi_2 \circ \varphi_1^{-1}$ is the identity where it is defined. If U_1 involves a branch point with m sheets, then φ_1^{-1} is essentially the m^{th} power, and $\varphi_2 \circ \varphi_1^{-1}(z) = z^m$, which is holomorphic. If U_2 involves a branch point, then $\varphi_2 \circ \varphi_1^{-1}$ is a holomorphic determination of the m^{th} root.

The observation made above will be used to give a definition of Riemann surface in the next chapter.

Chapter II

ABSTRACT RIEMANN SURFACES

In the introduction we have considered one method of constructing Riemann surfaces and have pointed out various properties. In the rest of the course several other methods will be given, especially the extremely important sheaf of germs of meromorphic functions in Chapter III and its generalization, the analytic configuration, in Chapter IV. Other examples will be considered in the present chapter. All of these Riemann surfaces have one feature that cries out for attention, so before coming to the concrete examples we shall define this characteristic feature and call any object which possesses it a Riemann surface.

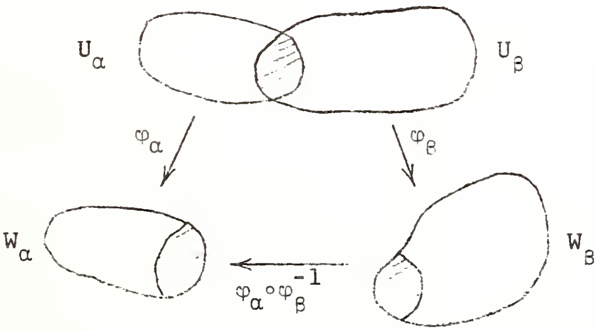
DEFINITION 1. A surface is a Hausdorff space S such that $\forall$ $p \in S$ $\exists$ an open neighborhood U of p and an open set $W \subset \mathbb{C}$ and a homeomorphism φ: $U \to W$. Such a mapping φ is called a chart or a coordinate mapping.

DEFINITION 2. Let S be a surface. An atlas for S is a collection of charts $\{\varphi_\alpha\}$, where α runs through some index set, such that every point of S belongs to the domain of some φ_α. If φ_α: $U_\alpha \to W_\alpha$, then we are saying that

$$S = \bigcup_{\alpha} U_\alpha .$$

Note that if U_α and U_β meet, then both φ_α and φ_β are defined on the intersection $U_\alpha \cap U_\beta$ and these mappings provide homeomorphisms between this intersection and the open sets $\varphi_\alpha(U_\alpha \cap U_\beta)$ and $\varphi_\beta(U_\alpha \cap U_\beta)$ in $\mathbb{C}$, respectively. Therefore, there is defined the function

$$\varphi_\alpha \circ \varphi_\beta^{-1}: \varphi_\beta(U_\alpha \cap U_\beta) \to \varphi_\alpha(U_\alpha \cap U_\beta) .$$



For brevity we shall frequently speak of $\varphi_\alpha \circ \varphi_\beta^{-1}$ without mentioning that it is defined only on $\varphi_\beta(U_\alpha \cap U_\beta)$. The functions $\varphi_\alpha \circ \varphi_\beta^{-1}$ are called <u>coordinate transition functions</u> of the atlas, because if φ_α and φ_β are thought of as defining coordinates on $U_\alpha \cap U_\beta$, the mapping $\varphi_\alpha \circ \varphi_\beta^{-1}$ determines how to change from one coordinate system to another.

<u>DEFINITION 3</u>. An atlas $\{\varphi_\alpha\}$ is <u>analytic</u> if each coordinate transition function $\varphi_\alpha \circ \varphi_\beta^{-1}$ is analytic.

Just note that this definition makes good sense, as $\varphi_\alpha \circ \varphi_\beta^{-1}$ is a complex-valued function on an open set in $\mathbb{C}$ and thus the usual meaning of analytic function is what is meant.

<u>DEFINITION 4</u>. Two charts φ_1 and φ_2 on a surface S are <u>compatible</u> if the functions $\varphi_1 \circ \varphi_2^{-1}$ and $\varphi_2 \circ \varphi_1^{-1}$ are analytic. A chart φ is <u>compatible with an analytic atlas</u> $\{\varphi_\alpha\}$ if φ and φ_α are compatible for all α.

<u>DEFINITION 5</u>. An analytic atlas is <u>complete</u> if it contains every chart compatible with it.

We are now almost ready to define a Riemann surface as a surface together with an analytic atlas. But there is a slight technical problem which must be overcome. Namely, there is almost never a convenient canonical atlas, and we therefore either need to define some sort of canonical atlas or need to define an equivalence relation between analytic atlases. Since these approaches are really the same, we arbitrarily pick the former possibility. This is the reason for Definition 5. Now we give a lemma which actually relates these concepts.

<u>LEMMA 1</u>. <u>For any analytic atlas</u> $\{\varphi_\alpha\}$ <u>on a surface</u> S, <u>there exists exactly one complete analytic atlas containing it</u>. <u>This complete analytic atlas is the collection of all charts compatible with</u> $\{\varphi_\alpha\}$.

<u>Proof</u>: Let $\mathcal{A}$ be the set of all charts compatible with $\{\varphi_\alpha\}$. We first prove that $\mathcal{A}$ is an atlas, then that it is complete. Suppose then that $\varphi,\varphi' \in \mathcal{A}$. Thus, $\varphi: U \to W$ and $\varphi': U' \to W'$ are homeomorphisms from open sets in S to open sets in $\mathbb{C}$. Suppose U and U′ meet and let $p_0 \in U \cap U'$. Since $\{\varphi_\alpha\}$ is an atlas, there exists $\varphi_\alpha: U_\alpha \to W_\alpha$ such that $p_0 \in U_\alpha$. Then

$$\varphi' \circ \varphi^{-1} = (\varphi' \circ \varphi_\alpha^{-1}) \circ (\varphi_\alpha \circ \varphi^{-1})$$

and $\varphi' \circ \varphi_\alpha^{-1}$ and $\varphi_\alpha \circ \varphi^{-1}$ are both holomorphic since φ,φ' are compatible with φ_α. Thus, $\varphi' \circ \varphi^{-1}$ is holomorphic. Thus, $\mathcal{A}$ is an analytic atlas. To prove that $\mathcal{A}$ is complete, suppose ψ is compatible with $\mathcal{A}$. Since $\mathcal{A}$ contains $\{\varphi_\alpha\}$ (since $\{\varphi_\alpha\}$ is itself an analytic atlas), ψ is compatible with $\{\varphi_\alpha\}$. That is, $\psi \in \mathcal{A}$. Thus, $\mathcal{A}$ is complete.

Finally, to prove that $\mathcal{A}$ is unique, suppose $\mathcal{B}$ is a complete analytic atlas containing $\{\varphi_\alpha\}$. If

$\varphi \in \mathcal{B}$, then φ is compatible with $\{\varphi_\alpha\}$, and thus $\varphi \in \mathcal{C}$. This proves $\mathcal{B} \subset \mathcal{C}$. Now suppose $\varphi \in \mathcal{C}$. Let $\psi \in \mathcal{B}$. Arguing as above, we find

$$\varphi \circ \psi^{-1} = (\varphi \circ \varphi_\alpha^{-1}) \circ (\varphi_\alpha \circ \psi^{-1}) ,$$

$$\psi \circ \varphi^{-1} = (\psi \circ \varphi_\alpha^{-1}) \circ (\varphi_\alpha \circ \varphi^{-1}) ,$$

and thus φ and ψ are compatible. Thus, φ is compatible with $\mathcal{B}$. As $\mathcal{B}$ is complete, $\varphi \in \mathcal{B}$. Thus, $\mathcal{C} \subset \mathcal{B}$. Hence, $\mathcal{C} = \mathcal{B}$.

QED

As a result of this lemma, we see that two analytic atlases are contained in the same complete analytic atlas if and only if each chart from one atlas is compatible with each chart from the other atlas, or if and only if the union of the two atlases is itself an analytic atlas.

DEFINITION 6. A Riemann surface is a surface together with a complete analytic atlas.

Thus, to specify an abstract Riemann surface, we must specify a surface and a complete analytic atlas.

The effective purpose of Lemma 1 is to enable us to forget about the rather cumbersome completeness assumption. So when we wish to construct a Riemann surface, we will be satisfied to exhibit <u>one</u> analytic atlas, keeping in the back of our minds that Lemma 1 implies the existence of a unique larger complete analytic atlas. This is quite helpful, as it will usually be more or less obvious what can be chosen to be an analytic atlas.

It is most important for beginners in this subject not to be beguiled by Definition 6. The crux of the theory of Riemann surfaces is <u>not</u> this definition. This definition just gives a convenient term in a bookkeeping sense to keep track of the structure implied in the definition of complete analytic atlas. Thus, this chapter has been called "<u>abstract</u> Riemann surfaces." It will be up to us to verify for the many concrete Riemann surfaces we find that the above definition obtains. Now we pass to some examples.

<u>Examples</u>.

1. This is by far the most trivial example. Let S be any open subset of $\mathbb{C}$; the atlas consists of the single chart φ which is the

identity mapping on S. In this case φ is obviously a homeomorphism and the only transition function is $\varphi\circ\varphi^{-1}$ = identity on S.

2. A most important example is the Riemann sphere. We take this to be the topological space $\hat{\mathbb{C}} = \mathbb{C}\cup\{\infty\}$, where points in $\mathbb{C}$ have their usual neighborhoods and a neighborhood basis of ∞ consists of the sets $\{z:|z|>a\} \cup\{\infty\}$ for $0<a<\infty$. This is clearly a topological space and stereographic projection is a homeomorphism of $\hat{\mathbb{C}}$ onto the unit Euclidean sphere in $\mathbb{R}^3$. The atlas we pick will consist of two charts. Let $U_1 = W_1 = \mathbb{C}$ and $\varphi_1: U_1 \to W_1$ be the identity. Let $U_2 = \hat{\mathbb{C}} - \{0\}$, $W_2 = \mathbb{C}$, and $\varphi_2: U_2 \to W_2$ be given by $\varphi_2(z) = z^{-1}$, $\varphi_2(\infty) = 0$. These are clearly charts, and $\varphi_2\circ\varphi_1^{-1}(z) = \varphi_2(z) = z^{-1}$, $\varphi_1\circ\varphi_2^{-1}(z) = z^{-1}$, which shows the coordinate transition functions are holomorphic.

3. As we mentioned above, $\hat{\mathbb{C}}$ is homeomorphic to the unit sphere in $\mathbb{R}^3$. It is a fact that any topological space homeomorphic to a Riemann surface can itself be made into a Riemann surface. To see this, suppose S is a Riemann

surface with analytic atlas $\{\varphi_\alpha\}$ and that T is a topological space and $\Phi: T \to S$ a homeomorphism. Then the maps $\{\varphi_\alpha \circ \Phi\}$ form an analytic atlas for T with transition functions

$$(\varphi_\alpha \circ \Phi) \circ (\varphi_\beta \circ \Phi)^{-1} = \varphi_\alpha \circ \varphi_\beta^{-1} \quad .$$

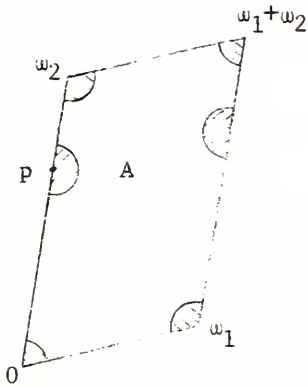
4. All the surfaces constructed in Chapter I are Riemann surfaces. The verification was briefly indicated on pp. 13-15.
5. Any open subset of a Riemann surface can be made into a Riemann surface in a natural way: If T is an open set in the Riemann surface S, then for a chart $\varphi_\alpha : U_\alpha \to W_\alpha$ on S let the mapping ψ_α be the restriction of φ_α to $U_\alpha \cap T$. Then an analytic atlas $\{\varphi_\alpha\}$ on S gives rise to an analytic atlas $\{\psi_\alpha\}$ on T.
6. The torus. Of course, the examples mentioned in 4 include a Riemann surface homeomorphic to a torus; cf. p. 12. Here is another way to make a torus into a Riemann surface.

Problem 1. Let ω_1 and ω_2 be nonzero complex numbers whose ratio is not real. Let $\Omega = \{n_1\omega_1 + n_2\omega_2 : n_1, n_2 \text{ integers}\}$, and for any $z \in \mathbb{C}$ let $[z] = z + \Omega$. Prove that $\exists\ \delta > 0$ such that $|n_1\omega_1 + n_2\omega_2| \geq \delta$ if n_1, n_2 are integers which are not both zero. Let $\mathbb{C}/\Omega$ be the set of all $[z]$ for $z \in \mathbb{C}$, noting that $[z] = [z'] \Leftrightarrow z - z' \in \Omega$. For any $[z]$ define a neighborhood basis of $[z]$ to consist of all sets

$$U_\varepsilon([z]) = \{[w] : |z - w| < \varepsilon\}$$

for $\varepsilon > 0$. Prove that $\mathbb{C}/\Omega$ becomes a Hausdorff space. For $\varepsilon \leq \delta/2$ let $\varphi : U_\varepsilon([z]) \to \Delta_\varepsilon = \{\zeta \in \mathbb{C} : |\zeta| < \varepsilon\}$ be defined by $\varphi([w]) = w - z$. Prove that these form charts in an analytic atlas for $\mathbb{C}/\Omega$.

The relation to a torus is that $\mathbb{C}/\Omega$ is homeomorphic to a torus in a natural way. This can perhaps best be seen by considering the set $A = \{t_1\omega_1 + t_2\omega_2 : 0 \leq t_1 < 1, 0 \leq t_2 < 1\} \subset \mathbb{C}$, which is obviously in one-to-one correspondence with $\mathbb{C}/\Omega$.



The topology in A is determined in a natural fashion: a neighborhood basis of a point $t_1\omega_1 + t_2\omega_2$ with $0 < t_1 < 1$, $0 < t_2 < 1$, can be taken to be sufficiently small disks centered at that point. For a point p as indicated in the figure, a neighborhood basis can be taken to be sets

$$\{z\in A: |z-p|<\varepsilon\}\cup\{z\in A: |z-p-\omega_1|<\varepsilon\}$$

for all sufficiently small ε. And a neighborhood basis of 0 can be described in a similar fashion, corresponding to the four smaller sectors in the figure. Of course, this topology just corresponds to a gluing in the sense of Chapter I and one easily sees that now A is homeomorphic to $\mathbb{C}/\Omega$, the homeomorphism being the mapping $A \to \mathbb{C}/\Omega$ which sends z to $[z]$. Finally, if one imagines this gluing carried out with a strip of paper the shape of A, it becomes clear that A is homeomorphic to a torus.

7. The _sheaf of germs of moromorphic functions_ to be discussed at length in Chapter III will be a Riemann surface in a natural way.

DEFINITION 7. A _path_ in a topological space S is a continuous function γ from $I = [0,1]$ into S. The _initial point_ of γ is $\gamma(0)$ and the _terminal point_ of γ is $\gamma(1)$. And γ is said to be a _path from_ $\gamma(0)$ _to_ $\gamma(1)$.

DEFINITION 8. A topological space S is _disconnected_ if $\exists$ open sets $A, B \subset S$ such that $S = A \cup B$, A and B are disjoint, and neither A nor B is empty. A topological space S is _connected_ if it is not disconnected.

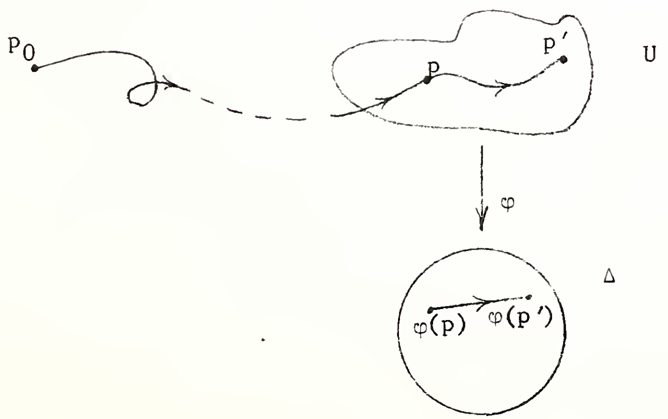
PROPOSITION 1. _A Riemann surface_ S _is connected if and only if for any points_ p_0 _and_ p_1 _in_ S _there exists a path in_ S _from_ p_0 _to_ p_1.

Proof: Suppose S is disconnected, and let A and B be the corresponding sets of Definition 8. Let $p_0 \in A$ and $p_1 \in B$. If there is a path γ in S

from p_0 to p_1, then $\gamma(I)$ is connected (it is a general result that a continuous image of a connected space is connected). However, the sets $A_1 = \gamma(I)\cap A$ and $B_1 = \gamma(I)\cap B$ show that in the sense of Definition 8 $\gamma(I)$ is disconnected.

Conversely, suppose S is connected and let p_0, $p_1 \in S$. Let $A = \{p\in S: \exists \text{ path in } S \text{ from } p_0 \text{ to } p\}$. Then A contains p_0 and is thus not empty. Also, A is open: if $p\in A$ then using an open neighborhood U of p and a chart $\varphi: U \to \Delta$ from U onto a disk Δ, then $U\subset A$. For if $p'\in U$ and if γ is the path from p_0 to p, then a path γ_1 from p_0 to p' is

$$\gamma_1(t) = \begin{cases} \gamma(2t), & 0\le t\le \frac{1}{2}, \\ \varphi^{-1}((2-2t)\varphi(p) + (2t-1)\varphi(p')), & \frac{1}{2}\le t\le 1. \end{cases}$$



Thus, A is open. A similar proof shows that A is closed: if p' is a limit point of A, then we can use the same picture as above, except that U is now picked to be a neighborhood of p' homeomorphic to a disk Δ. Since p' is a limit point of A, there is a point $p \in U \cap A$. Then the same construction as above shows that there is a path in S from p_0 to p'; i.e., $p' \in A$. Thus, A contains all its limit points and is therefore a closed set. Since A is open and closed and nonempty, and S is connected, we have $A = S$. Thus, $p_1 \in A$.

QED

Remark. Note that the above proof is entirely topological. In general topology this theorem states that a connected, locally arcwise connected space is arcwise connected.

Now we turn to the important concept of analytic functions.

DEFINITION 9. Let S_1 and S_2 be Riemann surfaces, U an open subset of S_1, and f a continuous function from U to S_2. Then f is analytic if for every chart $\varphi_1 : U_1 \to W_1$ on S_1 and every chart $\varphi_2 : U_2 \to W_2$ on S_2, the function $\varphi_2 \circ f \circ \varphi_1^{-1}$ is holomorphic. (Here and elsewhere when we use a phrase like "every chart φ_1" we mean every chart φ_1 in the complete analytic atlas for S_1.)

<u>Remark</u>. Since the coordinate transition functions are holomorphic, to check the analyticity of f in a neighborhood of a point $p_0 \in U$ it is sufficient to check the analyticity of $\varphi_2 \circ f \circ \varphi_1^{-1}$ for some chart φ_1 in a neighborhood of p_0 and some chart φ_2 in a neighborhood of $f(p_0)$. This remark also immediately leads to

<u>PROPOSITION 2</u>. <u>In the notation of Definition</u> 9 f <u>is analytic on</u> U <u>if and only if</u> f <u>is analytic in some neighborhood of each point of</u> U.

<u>Proof</u> is left to the reader.

<u>PROPOSITION 3</u>. <u>If</u> $f: S_1 \to S_2$ <u>is analytic and</u> $g: S_2 \to S_3$ <u>is analytic, then</u> $g \circ f: S_1 \to S_3$ <u>is analytic</u>.

<u>Proof</u>: Let $p_0 \in S_1$. Choose a chart $\varphi_3: U_3 \to W_3$ in a neighborhood of $g \circ f(p_0)$. Choose a chart $\varphi_2: U_2 \to W_2$ in a neighborhood of $f(p_0)$ such that $g(U_2) \subset U_3$. Choose a chart $\varphi_1: U_1 \to W_1$ in a neighborhood of p_0 such that $f(U_1) \subset U_2$. Then

$$\varphi_3 \circ g \circ f \circ \varphi_1^{-1} = (\varphi_3 \circ g \circ \varphi_2^{-1}) \circ (\varphi_2 \circ f \circ \varphi_1^{-1})$$

is a composition of holomorphic functions and is thus holomorphic. Thus, $g \circ f$ is analytic in a neighborhood of p_0 and Proposition 2 shows this suffices.

QED

Examples.

1. If S_1 is an open subset of the Riemann surface $\mathbb{C}$ and $S_2 = \mathbb{C}$, then $f: S_1 \to \mathbb{C}$ is analytic according to Definition 9 $\Leftrightarrow$ f is analytic in the usual sense (satisfies the Cauchy-Riemann equation).
2. If $S_1 = \hat{\mathbb{C}}$ and f is continuous from a neighborhood of ∞ into S_2, then f is analytic in a neighborhood of $\infty \Leftrightarrow$ the function $z \to f(z^{-1})$ is analytic in a neighborhood of 0. This follows because a chart near ∞ on $\hat{\mathbb{C}}$ is the mapping $\varphi(z) = z^{-1}$.
3. Likewise, if $S_2 = \hat{\mathbb{C}}$ and $f: S_1 \to \hat{\mathbb{C}}$ is continuous in a neighborhood of p_0 and $f(p_0) = \infty$, then f is analytic in a neighbor-hood of $p_0 \Leftrightarrow \frac{1}{f}$ is analytic from a neighbor-

of p_0 to C.

4. An analytic function from a Riemann surface to C is said to be <u>holomorphic</u>; an analytic function from a Riemann surface to $\hat{C}$ is said to be <u>meromorphic</u>.
5. Any chart in the complete analytic atlas of a Riemann surface is holomorphic.
6. Consider the torus C/Ω as discussed in 6 on p. 24. Let $\pi: C \to C/\Omega$ be the canonical mapping $\pi(z) = [z]$. Then π is analytic. To see this consider $\varphi: U_\varepsilon([z]) \to \Delta_\varepsilon$ as in Problem 1. Then in a neighborhood of the fixed point z we have $\varphi\circ\pi(w) = \varphi([w]) = w-z$, a holomorphic function of w.
7. Again for the torus C/Ω considered in 6, we show that if S is a Riemann surface and $f: C/\Omega \to S$, then f is analytic $\Leftrightarrow \exists\ F: C \to S$ analytic such that

$$F = f\circ\pi.$$

First, if f is analytic and F is defined this way, then F is a composition of analytic functions and is thus analytic.

Now suppose F is analytic and $F = f \circ \pi$. We shall then prove that f is analytic in a neighborhood of any point $[z] \in \mathbb{C}/\Omega$. Take $\varphi : U_\varepsilon([z]) \to \Delta_\varepsilon$ as in Problem 1. Then for $p \in U_\varepsilon([z])$ we can write $p = [w]$, where $|w-z| < \varepsilon$ and $\varphi(p) = w-z$. Thus

$$f(p) = f(\pi(w)) = F(w) = F(z + \varphi(p)),$$

and we have exhibited f as a composition of analytic functions, so that f is analytic near $p_0 = [z]$.

This example really indicates the importance of the notion of analytic functions, since we see that there is a natural identification of analytic functions on $\mathbb{C}/\Omega$ with analytic functions F on $\mathbb{C}$ which are doubly periodic, i.e., which satisfy

$$F(z+\omega_1) = F(z),$$
$$F(z+\omega_2) = F(z).$$

When $S = \hat{\mathbb{C}}$ these are the elliptic functions.

8. For the Riemann surfaces constructed in the

introduction there are corresponding analytic functions. For example, consider the Riemann surface S for $\log z$ and the function f on S corresponding to $\log z$ (pp. 3-4). Then f is holomorphic on S. Likewise, consider the Riemann surface $\tilde{T}$ for $z^{1/m}$ and the corresponding function f (pp. 5-6). Then f is meromorphic on $\tilde{T}$. This really follows from 5 above since near the branch point 0 the function f is a chart and likewise near the branch point ∞, and away from the branch points the verification is obvious.

9. The analytic functions from $\hat{\mathbb{C}}$ to $\hat{\mathbb{C}}$ are the rational functions.

10. The analytic functions from $\hat{\mathbb{C}}$ to $\mathbb{C}$ are the constant functions (Liouville's theorem).

Now we shall develop some general properties of analytic functions. The main thing to note is the fact that local properties of analytic functions of a complex variable usually go over to corresponding properties in the general case in an obvious and trivial fashion. For example, we have

PROPOSITION 4. An analytic function $f:S_1 \to S_2$ which is not constant on any neighborhood is an open mapping.

Proof: We must show that if $p_0 \in S_1$ and U_1 is a neighborhood of p_0, then $f(U_1)$ contains a neighborhood of $f(p_0)$. We can assume $\varphi_1: U_1 \to W_1$ is a chart for S_1 and $\varphi_2: U_2 \to W_2$ a chart for S_2 and $f(U_1) \subset U_2$. Then $\varphi_2 \circ f \circ \varphi_1^{-1}$ is a nonconstant holomorphic function on W_1 and by the known property that a holomorphic function of a complex variable is open if not constant we see that $\varphi_2 \circ f \circ \varphi_1^{-1}(W_1)$ contains a neighborhood G of $\varphi_2 \circ f(p_0)$. As φ_2 is a homeomorphism, this implies $f(U_1)$ contains a neighborhood $\varphi_2^{-1}(G)$ of $f(p_0)$.

QED

Also, global topological properties of Riemann surfaces can be combined with local properties of analytic functions in a decisive manner.

PROPOSITION 5. If S_1 is a connected Riemann surface and if

$$f:S_1 \to S_2, \quad g:S_1 \to S_2,$$

<u>are analytic functions such that</u> f <u>and</u> g <u>coincide on some set which has a limit point in</u> S_1, <u>then</u> $f = g$.

<u>Proof</u>: Let $A = \{p \in S_1$: f and g coincide in a neighborhood of p$\}$. Clearly, A is open by its very definition. Also, $A \neq \emptyset$, for if $f(p_n) = g(p_n)$ with $p_n \to p_0$ $(p_n \neq p_0)$, then $p_0 \in A$; to see this let $\varphi_i : U_i \to W_i$ be charts for S_i, $p_0 \in U_1$, $f(p_0) = g(p_0) \in U_2$. Then $\varphi_2 \circ f \circ \varphi_1^{-1}$ and $\varphi_2 \circ g \circ \varphi_1^{-1}$ are holomorphic in W_1 and agree on a sequence in W_1 tending to $\varphi_1(p_0) \in W_1$, and thus by the known property for holomorphic functions of a complex variable, $\varphi_2 \circ f \circ \varphi_1^{-1}$ and $\varphi_2 \circ g \circ \varphi_1^{-1}$ coincide in a neighborhood of $\varphi_1(p_0)$. Thus, f and g coincide in a neighborhood of p_0, and we see that $p_0 \in A$. A similar proof shows that A is closed; just use the previous argument with p_0 taking the role of a limit point of A. As S_1 is connected, $A = S_1$.

QED

<u>PROPOSITION 6</u>. <u>If</u> S <u>is a connected Riemann surface and if</u> $f: S \to \mathbb{C}$ <u>is holomorphic</u>, <u>then</u> $|f|$ <u>has no relative maximum in</u> S <u>unless</u> f <u>is constant</u>.

Proof: Suppose $|f|$ has a relative maximum at p_0: $|f(p)| \le |f(p_0)|$ for p near p_0. Then the maximum principle for holomorphic functions of a complex variable implies f is constant in a neighborhood of p_0. Proposition 5 implies f is constant on S.

QED

PROPOSITION 7. *If f is a holomorphic function on a Riemann surface minus a point, $S - \{p_0\}$, and if f is bounded in a neighborhood of p_0, then f has a unique extension to a holomorphic function on S.*

Proof: Apply the usual theorem on removable singularities to show that if $\varphi: U \to W$ is a chart in a neighborhood of p_0, then there is a holomorphic function g on W such that $f \circ \varphi^{-1} = g$ on $W - \{\varphi(p_0)\}$. The extension of f near p_0 is then $g \circ \varphi$.

QED

PROPOSITION 8. *If S is a compact connected Riemann surface, the only holomorphic functions on S are constants.*

Proof: Suppose $f: S \to \mathbb{C}$ is analytic. Since S is compact, the continuous function $|f|$ assumes its maximum at some point of S. Since S is connected, Proposition 6 implies f is constant.

QED

Now let us examine in some detail the local properties of meromorphic functions. Let f be meromorphic in a neighborhood of p_0 in a Riemann surface S. If $\varphi: U \to W$ is a chart in the complete analytic atlas for S and U is a neighborhood of p_0, then a translation of the set W in $\mathbb{C}$ allows us to assume $\varphi(p_0) = 0$. Thus, $f \circ \varphi^{-1}$ is meromorphic in a neighborhood of 0 in $\mathbb{C}$. Thus, $f \circ \varphi^{-1}$ has a Laurent expansion

$$f \circ \varphi^{-1}(z) = \sum_{k=N}^{\infty} a_k z^k \quad , \quad a_N \neq 0.$$

If $\psi: U_1 \to W_1$ is another chart in the complete analytic atlas for S, U_1 a neighborhood of p_0, $\psi(p_0) = 0$, then $\varphi \circ \psi^{-1}$ and its inverse are holomorphic and map 0 to 0, and thus near $w = 0$

$$\varphi \circ \psi^{-1}(w) = \sum_{k=1}^{\infty} c_k w^k, \qquad c_1 \neq 0.$$

Therefore,

$$f\circ\psi^{-1}(w) = f\circ\varphi^{-1}\circ\varphi\circ\psi^{-1}(w)$$
$$= a_N c_1^N w^N + \dots \quad ,$$

where the additional terms involve higher powers of w. Therefore,

$$f\circ\psi^{-1}(w) = \sum_{k=N}^{\infty} b_k w^k, \qquad b_N \neq 0.$$

Thus, the number N does not depend on the particular chart used, but depends only on the function f. It is called the <u>divisor of</u> f <u>at</u> p_0 and is written

$$N = \partial_f(p_0).$$

There is another integer associated with f which is perhaps more important. Suppose f is not constant near p_0. If the divisor N of f at p_0 is negative, then the <u>multiplicity</u> of f at p_0 is said to be $-N$. Now suppose $\partial_f(p_0) \geq 0$. Then the <u>multiplicity</u> of f at p_0 is the divisor of $f - f(p_0)$ at p_0. Thus, we have for $m_f(p_0)$, the <u>multiplicity of</u> f <u>at</u> p_0, the formula

$$m_f(p_0) = -\partial_f(p_0) \quad \text{if} \quad \partial_f(p_0) < 0,$$

$$m_f(p_0) = \partial_{f-f(p_0)}(p_0) \quad \text{if } \partial_f(p_0) \geq 0.$$

Thus, $m_f(p_0)$ is a positive integer which is completely determined by f.

In terms of $m_f(p_0)$ we can obtain a simple representation for f by choosing an appropriate chart near p_0. Thus, let $m = m_f(p_0)$ and consider two cases:

$\underline{\partial_f(p_0) \geq 0}$. In this case the Laurent expansion appears in the form

$$f\circ\varphi^{-1}(z) = f(p_0) + a_m z^m + \ldots, \quad a_m \neq 0.$$

Let α be one of the $m^{\underline{th}}$ roots of a_m and note that

$$a_m z^m + a_{m+1} z^{m+1} + \ldots = \alpha^m z^m \Big(1 + \sum_{k=1}^{\infty} \frac{a_k}{a_m} z^k\Big) .$$

Let $h(z)$ be the principal $m^{\underline{th}}$ root of $1 + \sum_{k=1}^{\infty} \frac{a_k}{a_m} z^k$ near $z = 0$, so that

$$f\circ\varphi^{-1}(z) = f(p_0) + (\alpha z h(z))^m \quad .$$

Now define a new chart near p_0 by the equation

$$\psi(p) = \alpha\varphi(p)h(\varphi(p)), \quad p \text{ near } p_0.$$

Then ψ is a chart in the complete analytic atlas for S since the mapping $z \to \alpha zh(z)$ is a conformal equivalence near 0; and

$$\begin{aligned} f\circ\psi^{-1}(w) &= f\circ\varphi^{-1}\circ\varphi\circ\psi^{-1}(w) \\ &= f(p_0) + (\alpha\varphi\circ\psi^{-1}(w)h(\varphi\circ\psi^{-1}(w)))^m \\ &= f(p_0) + (\psi(\psi^{-1}(w)))^m \\ &= f(p_0) + w^m \quad . \end{aligned}$$

<u>$\partial_f(p_0) < 0$</u>. Now the Laurent expansion is

$$\begin{aligned} f\circ\varphi^{-1}(z) &= a_{-m}z^{-m} + a_{1-m}z^{1-m} + \ldots, \quad a_{-m} \neq 0, \\ &= a_{-m}z^{-m}(1 + \frac{a_{1-m}}{a_{-m}} z + \ldots). \end{aligned}$$

In this case choose α such that $\alpha^{-m} = a_{-m}$ and h holomorphic near 0 with $h(0) = 1$, $h(z)^{-m} = 1 + \frac{a_{1-m}}{a_{-m}}z + .$ Then

$$f\circ\varphi^{-1}(z) = (\alpha zh(z))^{-m} \quad ,$$

so a similar argument shows that there is a chart ψ at p_0 such that

$$f\circ\psi^{-1}(w) = w^{-m} \quad .$$

Summarizing, if $m = m_f(p_0)$, then there is a chart ψ in a neighborhood of p_0 such that

$$f\circ\psi^{-1}(w) = f(p_0) + w^m \quad \text{if} \quad \partial_f(p_0) \geq 0 \;,$$

$$f\circ\psi^{-1}(w) = w^{-m} \quad \text{if} \quad \partial_f(p_0) < 0.$$

We note that it is easy to prove that

$$m_{g\circ f}(p_0) = m_g(f(p_0))m_f(p_0).$$

DEFINITION 10. Two Riemann surfaces S_1 and S_2 are equivalent if there are analytic functions $f: S_1 \to S_2$ and $g: S_2 \to S_1$ such that $f\circ g$ = the identity on S_2 and $g\circ f$ = the identity on S_1. Thus, each mapping f and g is bijective, analytic, and has analytic inverse.

It is routine to check that we have defined an equivalence relation. Note that equivalent Riemann surfaces are homeomorphic. The converse is not valid. We shall see that among the tori $\mathbb{C}/\Omega$ constructed in

6 on pp. 23-25 there are infinitely many nonequivalent Riemann surfaces. However, if a Riemann surface is homeomorphic to $\hat{\mathbb{C}}$, then it is equivalent to $\hat{\mathbb{C}}$ with its usual complete analytic atlas. This will be proved in Chapter VII.

Here is perhaps the simplest example of two homeomorphic nonequivalent Riemann surfaces. Let S_1 be $\mathbb{C}$ with the usual analytic atlas. Let $\Delta = \{z: |z| < 1\}$ and define a homeomorphism $\varphi: \mathbb{C} \to \Delta$. E.g.,

$$\varphi(z) = \frac{z}{\sqrt{1+|z|^2}} \quad .$$

Let Δ have the usual analytic atlas and define S_2 to be the Riemann surface induced on $\mathbb{C}$ by the homeomorphism φ, as in 3 on p. 22 . In other words, S_2 has an analytic atlas consisting of the single chart φ. Then S_1 and S_2 are not equivalent. For suppose $f: S_1 \to S_2$ is analytic. Then by definition $\varphi \circ f$ is a holomorphic function from $\mathbb{C}$ to Δ and is therefore constant by Liouville's theorem. Thus, f is constant.

We wish to consider a final feature of analytic functions. Suppose S and T are Riemann surfaces and that $f:S \to T$ is analytic. If $q \in T$ we say that f <u>takes the value</u> q n <u>times</u> if $f^{-1}(\{q\}) = \{p_1,\ldots p_\ell\}$ is finite and

$$\sum_{k=1}^{\ell} m_f(p_k) = n$$

(thus we are counting "according to multiplicity"). If $f^{-1}(\{q\}) = \emptyset$ we have $n = 0$. If this situation occurs, then there are charts $\varphi_k:U_k \to W_k$ in the complete analytic atlas for S with $p_k \in U_k$ and a chart $\varphi:U \to W$ in the complete analytic atlas for T such that the collection of sets $\{U_k\}$ is disjoint and $\varphi\circ f\circ\varphi_k^{-1}(z) = z^{m_f(p_k)}$ for $z \in W_k$. By diminishing the sizes of the U_k (if necessary) we can also assume that each U_k is contained in a compact set in S. Also, the explicit form for $\varphi\circ f\circ\varphi_k^{-1}$ given above shows that there exists a neighborhood V of q such that the restriction of f to U_k takes each value in V exactly $m_f(p_k)$ times. Therefore, the restriction of f to $\bigcup_{k=1}^{\ell} U_k$ takes each value in V exactly n times. Using this background information, we can prove

PROPOSITION 9. *Let S and T be Riemann surfaces and $f:S \to T$ an analytic function which is not constant on any neighborhood.*

1. *If S is compact and T is connected, then f takes every value in T the same number of times. Also, it follows that T is compact.*
2. *If f takes every value in T the same (finite) number of times, then f is proper. I.e., f^{-1} of any compact set is compact. In particular, if also T is compact, then S is compact.*

Proof: The rest of the proof is just topology. Assume the hypothesis of 1. Since $f(S)$ is a continuous image of a compact set, $f(S)$ is compact. Since T is Hausdorff, $f(S)$ is closed. Proposition 4 implies $f(S)$ is open. Since T is connected, $f(S) = T$. (Thus, T is itself compact.) If f takes any value infinitely often, then since S is compact there is a limit point in S of the set where f takes this value, and Proposition 5 implies f is constant in a neighborhood of this limit point. Thus,

f takes every value q in T a finite number $N(q)$ times. Now we use the argument just preceding this proposition with no change in notation. Since f does not take the value q on the compact set $S - \bigcup_{k=1}^{\ell} U_k$, there exists a neighborhood G of q such that the compact and thus closed set $f(S - \bigcup_{k=1}^{\ell} U_k)$ is disjoint from G. Thus, the restriction of f to $\bigcup_{k=1}^{\ell} U_k$ takes each value in $V \cap G$ exactly $N(q)$ times, and outside $\bigcup_{k=1}^{\ell} U_k$ f takes no value in $V \cap G$. Thus, $N(q') \equiv N(q)$ for $q' \in V \cap G$. Thus, the integer-valued function $q \to N(q)$ is continuous on T. Since T is connected, N is constant and part 1 is proved.

Now we prove 2. Let f take every value in T n times. If $q \in T$, the analysis preceding the proposition again shows that f takes each value in V exactly n times <u>in</u> $\bigcup_{k=1}^{\ell} U_k$. Therefore, by hypothesis $f^{-1}(V) \subset \bigcup_{k=1}^{\ell} U_k$. Therefore, $f^{-1}(V)$ is contained in a compact set in S since the U_k's are contained in compact sets in S. If $F \subset T$ is compact, then $F \subset \bigcup_{j=1}^{m} V_j$, where $f^{-1}(V_j)$ is contained in a compact set in S. Thus, $f^{-1}(F)$ is contained in a compact set S, and since $f^{-1}(F)$ is closed, it is compact.

<u>QED</u>

Chapter III

THE WEIERSTRASS CONCEPT OF A RIEMANN SURFACE

In this chapter we shall consider the process of analytic continuation and obtain the Weierstrass definition of analytic function. As we shall see, a concise language can be given to this process which brings us to construct a Riemann surface as a replacement for ("multiple-valued") analytic function.

The basic idea and the basic difficulty have been indicated on p. 1. Given a holomorphic or even a meromorphic function defined on an open set, we want to extend it as far as possible and somehow take care of the multiple-valuedness that arises.

What we shall basically consider is analytic continuation along paths. First, we shall describe the classical concept and then we shall fix everything up with lots of notation so that we end up with a Riemann surface and so that the problems of analytic continuation go over into statements about topological and other properties of Riemann surfaces. Analytic continuation is classically considered in the following way: suppose f is meromorphic in a disk Δ and has a Laurent expansion about the center a of Δ converging in Δ:

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z-a)^k, \quad z \in \Delta.$$

If $b \in \Delta$, it then makes sense to consider the Laurent expansion of f about b. This can be done in at least two ways; we can write

$$\begin{aligned} (z-a)^k &= (z-b + b-a)^k \\ &= (b-a)^k \left(1 + \frac{z-b}{b-a}\right)^k \\ &= (b-a)^k \sum_{n=0}^{\infty} \binom{k}{n} \frac{(z-b)^n}{(b-a)^n} \end{aligned}$$

by the binomial theorem, and then we insert this into the formula for f, rearrange terms (permissible if z is near b), and thus obtain for $b \neq a$ a <u>Taylor</u> series expansion for f in a neighborhood of b. The other procedure would be simply to write

$$f(z) = \sum_{n=0}^{\infty} \frac{f^{(n)}(b)}{n!} (z-b)^n, \quad z \text{ near } b.$$

If it so happens that the new series has radius of convergence larger than the distance from b to the boundary of Δ, we then have what is termed a <u>direct</u> or an <u>immediate</u> analytic continuation of f. Taking the new function in the new disk, we can again apply this process, etc. We can thus arrive at a sequence $\Delta_1, \Delta_2, \ldots$ of disks and corresponding meromorphic functions $f_1, f_2, \ldots$ such that

f_k is meromorphic in Δ_k,

the center of Δ_k belongs to Δ_{k-1},

f_k is a direct analytic continuation of f_{k-1}.

Our first adjustment of this process will be to ignore a definite procedure for direct analytic continuation. Thus, instead of considering f_k to be constructed from f_{k-1} by a definite process, we shall just require $f_{k-1} \equiv f_k$ in $\Delta_{k-1} \cap \Delta_k$. Also, there is then no reason to require the center of Δ_k to be in Δ_{k-1}.

DEFINITION 1. Let γ: $[0,1] \to \mathbb{C}$ be a path. An analytic continuation along γ is a collection of disks $\Delta_1, \Delta_2, \ldots, \Delta_n$ and meromorphic functions $f_1, f_2, \ldots, f_n$ such that

f_k is meromorphic in Δ_k,

$f_{k-1} \equiv f_k$ in $\Delta_{k-1} \cap \Delta_k$,

and such that there exist $0 = t_0 < t_1 < \ldots < t_n = 1$ with

$$\gamma([t_{k-1}, t_k]) \subset \Delta_k, \quad k = 1, 2, \ldots, n.$$

We clearly wish to consider all possible analytic continuations along paths, usually starting with a given meromorphic function f_1 in a given disk Δ_1. This will define a meromorphic function f_n in Δ_n, but the value $f_n(\gamma(1))$ is not in general independent of γ, so that we cannot in general define a meromorphic function in $\mathbb{C}$

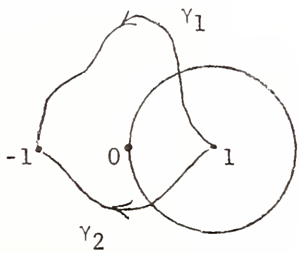
which extends f_1. For example, if

$$\Delta_1 = \{z: |z-1| < 1\}$$

and f_1 is the principal determination of $z^{\frac{1}{2}}$ in Δ_1:

$$f_1(z) = (1+(z-1))^{\frac{1}{2}} = \sum_{n=0}^{\infty} \binom{\frac{1}{2}}{n}(z-1)^n,$$

then analytic continuations along paths from 1 to -1 definitely depend on the path. Consider the figure:



analytic continuation along γ_1 yields a holomorphic function near -1 whose value at -1 is i; but analytic continuation along γ_2 yields the value -i.

These statements are trivial to justify since we can write $z = re^{i\theta}$ with $-\pi < \theta < \pi$ (except on the negative real axis) and then $z^{\frac{1}{2}} = r^{\frac{1}{2}}e^{i\theta/2}$. Along γ_1 θ increases to π and along γ_2 θ decreases to $-\pi$, and thus the two different values at -1 result.

Therefore, if we wish to define some meromorphic function which is a largest possible analytic continuation of f_1, or which is derived from f_1 by analytic continuation on all paths for which continuation is possible, we shall have to have something other than $\mathbb{C}$ on which to define the extended function. So we now begin to introduce the Riemann surface on which these continuations will be defined.

By the principle of the uniqueness of analytic continuation (p. 1), it suffices to know the original function in an arbitrarily small open neighborhood. Such a "germ" of a function uniquely will determine the function everywhere. So we make the following definitions.

DEFINITION 2. Let $a \in \mathbb{C}$ and suppose f and g are functions which are meromorphic in neighborhoods of a. Then f is equivalent to g, written $f \sim g$, if $f \equiv g$ in some neighborhood of a.

Clearly this is an equivalence relation, and we then make the following

DEFINITION 3. Let $a \in \mathbb{C}$. Then M_a is the collection of equivalence classes of functions meromorphic in a neighborhood of a. Any element of M_a is called a germ of a meromorphic function. If f is a meromorphic function in a neighborhood of a, then $[f]_a$ is the germ to which f belongs. We say $[f]_a$ is the germ of f at a.

By definition, $[f]_a = [g]_a \Leftrightarrow f \equiv g$ near a.

<u>DEFINITION 4</u>. $M = \bigcup_{a \in \mathbb{C}} M_a$. We also define the obvious mapping $\pi: M \to \mathbb{C}$ by $\pi([f]_a) = a$.

Below we shall make M into a topological space in a natural way and then M will be called the <u>sheaf of germs of meromorphic functions</u>, and M_a the <u>stalk</u> over the point a.

We shall define a topology on M by exhibiting a neighborhood basis for each point in M. Simple considerations then show that if we define a set in M to be <u>open</u> if it contains one of these special neighborhoods of each of its points, then the class of open sets in M forms a topology for which each point has the given neighborhood basis as a basis of open neighborhoods in this topology if the given neighborhood bases satisfy the following conditions:

1. any two neighborhoods of a point contain a third neighborhood of that point;
2. any neighborhood contains a neighborhood of each of its points.

Furthermore, the topology of M is <u>Hausdorff</u> if also

3. any two distinct points of M are contained in disjoint neighborhoods.

<u>The topology on M</u>. Suppose $[f]_a \in M$. Then there exists a disk Δ centered at a such that f is meromorphic

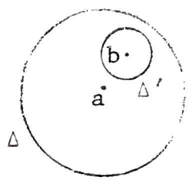
on Δ. Define

$$\boxed{U(a,f,\Delta) = \{[f]_b : b \in \Delta\}.}$$

A neighborhood basis of $[f]_a$ is defined to be all sets $U(a,f,\Delta)$ such that f is meromorphic on Δ. Although the definition is quite simple, we have already incorporated into it the notion of direct analytic continuation, for the definition states that the germs "close" to $[f]_a$ are just the germs of the function f itself at points close to a. Now we check the various requirements for neighborhood bases:

1. $U(a,f,\Delta_1) \cap U(a,f,\Delta_2) \supset U(a,f,\Delta_3)$
   if $\Delta_3 \subset \Delta_1 \cap \Delta_2$;
2. suppose $[f]_b \in U(a,f,\Delta)$. Then let Δ' be a disk centered at b such that $\Delta' \subset \Delta$, and note that

$$U(b,f,\Delta') \subset U(a,f,\Delta) \ ;$$



3. now we check that M is Hausdorff. Suppose that $[f]_a$ and $[g]_{a'}$ are distinct points in M. If $a \neq a'$, then take Δ and Δ' to be

disjoint disks centered at a and a′, respectively, and note that $U(a,f,\Delta)$ and $U(a',g,\Delta')$ are obviously disjoint. If $a = a'$, then choose a disk Δ such that f and g are meromorphic in Δ. Then $U(a,f,\Delta)$ and $U(a,g,\Delta)$ are disjoint. For otherwise there would exist a point $[f]_b = [g]_b$ for some $b \in \Delta$, and thus $f \equiv g$ in a neighborhood of b. By the uniqueness of analytic continuation, $f \equiv g$ in Δ, contradicting $[f]_a \neq [g]_a$.

Now M is a topological space. Note how much information is contained in the statement of the validity of the Hausdorff separation axiom--namely, this property reflects the uniqueness of analytic continuation.

M is a surface. The charts are almost obvious. Just use the mapping π restricted to the various neighborhoods $U(a,f,\Delta)$. Suppose we call φ the restriction of π to $U(a,f,\Delta)$. Then

$$\varphi\colon U(a,f,\Delta) \to \Delta$$

is given by $\varphi([f]_b) = b$, and $\varphi^{-1}(b) = [f]_b$. Thus, φ is a bijection. Also, if $U(b,f,\Delta') \subset U(a,f,\Delta)$, then clearly

$$\varphi(U(b,f,\Delta')) = \Delta'.$$

Thus, φ induces a one-to-one correspondence between a neighborhood basis of $[f]_b$ and a neighborhood basis of $\varphi([f]_b) = b$. Thus, φ is a homeomorphism. This proves that M is a surface.

Moreover, we now have a nice atlas on M and we claim it is an analytic atlas and thus

<u>M is a Riemann surface</u>. Suppose φ is the restriction of π to $U(a,f,\Delta)$ and ψ is the restriction of π to $U(b,g,\Delta')$. If $z \in \Delta$ and $\varphi^{-1}(z) \in U(a,f,\Delta) \cap U(b,g,\Delta')$, then $\varphi^{-1}(z) = [f]_z = [g]_z$, and thus $\psi(\varphi^{-1}(z)) = z$. Thus, where it is defined we have

$$\psi \circ \varphi^{-1} = \text{identity!}$$

The coordinate transition functions are thus trivially holomorphic and M is a Riemann surface.

The mapping $\pi: M \to \mathbb{C}$ is holomorphic. This really needs no checking at all, since π restricted to any neighborhood $U(a,f,\Delta)$ is a chart in the analytic atlas we have constructed, and such charts are always holomorphic (p. 31, no. 5).

<u>Problem 2</u>. Define $V: M \to \hat{\mathbb{C}}$ by the formula

$$V([f]_a) = f(a).$$

Prove that V is meromorphic.

Thus, we have two meromorphic functions π and V on M which are quite natural and simple functions to consider. We shall in the next chapter define an extension of M which is quite a bit more complicated, and again will be able to single out two natural meromorphic functions, which we shall again designate π and V. In that context these functions will appear very much alike, although on M the function π seems to be somewhat simpler than V.

In terms of M we can give a characterization of analytic continuation along paths (see Definition 1).

PROPOSITION 1. *Let $[f]_a \in M$ be given, and let γ be a path in $\mathbb{C}$ starting at a. A necessary and sufficient condition that there exists an analytic continuation along γ with $f_1 = f$ in a disk Δ_1 containing a* (using the notation of Definition 1) *is that there exists a path $\tilde{\gamma}$ in M such that*

$$\pi \circ \tilde{\gamma} = \gamma,$$
$$\tilde{\gamma}(0) = [f]_a.$$

Proof: The necessity is quite clear. Using the notation of Definition 1 we define

$$\tilde{\gamma}(t) = [f_k]_{\gamma(t)}, \quad t_{k-1} \le t \le t_k .$$

Since $\gamma(t_{k-1}) \in \Delta_{k-1} \cap \Delta_k$ and $f_{k-1} \equiv f_k$ in $\Delta_{k-1} \cap \Delta_k$, we have $[f_{k-1}]_{\gamma(t_{k-1})} = [f_k]_{\gamma(t_{k-1})}$. Thus, $\tilde{\gamma}$ is unambiguously defined, and clearly $\pi(\tilde{\gamma}(t)) = \gamma(t)$, $\tilde{\gamma}(0) = [f_1]_{\gamma(0)} = [f]_a$. The continuity of $\tilde{\gamma}$ is immediate from the definition of the topology of M and the continuity of γ.

The proof of sufficiency relies on a compactness argument. The continuity of $\tilde{\gamma}$ and the definition of the topology of M show that for each $t \in [0,1]$ there exists an open interval I_t (open relative to $[0,1]$) containing t and a meromorphic function f_t defined on a disk Δ_t centered at $\gamma(t)$ such that

$$\tilde{\gamma}(I_t) \subset U(\gamma(t), f_t, \Delta_t) \equiv U_t \subset M.$$

The compactness result we need is that there exists $\varepsilon > 0$ such that any interval in $[0,1]$ of length not exceeding ε is contained in some one of the intervals I_t. The proof proceeds in the following manner. For each $s \in [0,1]$ there exists $r(s) > 0$ such that

$$[0,1] \cap (s - r(s), s + r(s)) \subset I_s.$$

As $[0,1]$ is compact, there exist points $s_1, \ldots, s_k$ such that

$$[0,1] \subset \bigcup_{j=1}^{k} (s_j - \tfrac{1}{2} r(s_j), s_j + \tfrac{1}{2} r(s_j)).$$

Let $\varepsilon = \min\{r(s_j) : 1 \le j \le k\}$. Then if $x \in [0,1]$ choose j such that $|x - s_j| < \tfrac{1}{2} r(s_j)$. If $|y - x| \le \tfrac{1}{2}\varepsilon$, then

$$|y-s_j| \le |y-x|+|x-s_j| < \tfrac{1}{2}\varepsilon+\tfrac{1}{2}r(s_j) \le r(s_j),$$

so

$$y \in (s_j-r(s_j), s_j+r(s_j)) \subset I_{s_j}.$$

Thus,

$$[0,1] \cap [x-\tfrac{1}{2}\varepsilon, x+\tfrac{1}{2}\varepsilon] \subset I_{s_j},$$

as required.

Now choose points $t_0,\ldots,t_n$ such that $0 = t_0<t_1<\ldots<t_n = 1$, $t_k-t_{k-1} \le \varepsilon$, $1 \le k \le n$. By choice of ε, the interval $[t_{k-1},t_k]$ is contained in some set I_{τ_k} constructed above. Thus, we are given a collection of disks Δ_{τ_k} and meromorphic functions f_{τ_k} on Δ_{τ_k}, and we have to check that we have thereby obtained an analytic continuation along γ. Since $\tilde{\gamma}([t_{k-1},t_k]) \subset \tilde{\gamma}(I_{\tau_k}) \subset U_{\tau_k}$, we obtain

$$\gamma([t_{k-1},t_k]) = \pi(\tilde{\gamma}([t_{k-1},t_k])) \subset \pi(U_{\tau_k}) = \Delta_{\tau_k}.$$

If $z \in \Delta_{\tau_{k-1}} \cap \Delta_{\tau_k}$, then the corresponding points in $U_{\tau_{k-1}}$ and U_{τ_k} are $[f_{\tau_{k-1}}]_z$ and $[f_{\tau_k}]_z$, respectively. In particular for $z = \gamma(t_{k-1})$ we have

$$\tilde{\gamma}(t_{k-1}) = [f_{\tau_{k-1}}]_z = [f_{\tau_k}]_z .$$

so that $f_{\tau_{k-1}} \equiv f_{\tau_k}$ in a neighborhood of $\gamma(t_{k-1})$, and by analytic continuation $f_{\tau_{k-1}} \equiv f_{\tau_k}$ in $\Delta_{\tau_{k-1}} \cap \Delta_{\tau_k}$. Finally, it is clear that since $\tilde{\gamma}(0) = [f]_a$, we have $f_{\tau_1} \equiv f$ in

$\Delta_1 \cap \Delta_{\tau_0}$, so the analytic continuation along γ which we have constructed begins with the given meromorphic function f in a neighborhood of a.

QED

DEFINITION 5. If γ and $\tilde{\gamma}$ are paths into C and M, respectively, such that $\gamma = \pi\circ\tilde{\gamma}$, then $\tilde{\gamma}$ is said to be a lifting of γ.

PROPOSITION 2. "The uniqueness of analytic continuation" (Topologically speaking, "The unique lifting theorem"). If $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are paths in M such that $\pi\circ\tilde{\gamma}_1 = \pi\circ\tilde{\gamma}_2$, then either

$$\tilde{\gamma}_1(t) = \tilde{\gamma}_2(t) \text{ for every } t \in [0,1] ,$$

or

$$\tilde{\gamma}_1(t) = \tilde{\gamma}_2(t) \text{ for no } t \in [0,1] .$$

Proof: Let $A = \{t\in[0,1]: \tilde{\gamma}_1(t) = \tilde{\gamma}_2(t)\}$. By continuity of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$, A is a closed set. It is also an open set, for consider any $t_0 \in A$. Let $a = \pi\circ\tilde{\gamma}_i(t_0)$ and $\tilde{\gamma}_i(t_0) = [f]_a$. Then f is meromorphic on a disk Δ centered at a and a neighborhood $U(a,f,\Delta)$ of $[f]_a$ is defined. By continuity of $\tilde{\gamma}_i$, $\tilde{\gamma}_i(t)$ is contained in $U(a,f,\Delta)$ for t sufficiently near t_0 and thus for those values of t

$$\tilde{\gamma}_i(t) = [f]_{\pi\circ\tilde{\gamma}_i(t)}$$

and since $\pi\circ\tilde{\gamma}_1 = \pi\circ\tilde{\gamma}_2$ we obtain $\tilde{\gamma}_1(t) = \tilde{\gamma}_2(t)$, or, $t\in A$.

Since A is both open and closed and [0,1] is connected, we have either A = [0,1] or A is empty.

QED

Thus, although in the sense of Definition 1 analytic continuation is not a uniquely defined construction (since different choices could be made for the t_k's and the disks Δ_k), yet viewed as a lifting problem we do have a strong uniqueness statement. Moreover, we see now that the natural choice we have used in the proof of necessity of Proposition 1 was really forced upon us. There was no other way to choose $\tilde{\gamma}(t)$.

This discussion definitely does not imply that the unique continuation property along paths leads to a germ at the end point of the path which is uniquely determined by the end point. The discussion on p. 49 makes this clear. In terms of the example of the square root mentioned there, observe that if f_1 is the principal determination of $z^{\frac{1}{2}}$ near 1 and if γ is the path $\gamma(t) = e^{2\pi it}$, $0\le t\le 1$, then γ can be lifted to a path $\tilde{\gamma}$ such that $\tilde{\gamma}(0) = [f_1]_1$, but $\tilde{\gamma}(1) = [-f_1]_1$. Thus, $\tilde{\gamma}(0) \ne \tilde{\gamma}(1)$, although $\gamma(0) = \gamma(1)$. Another way of stating this is that $[f_1]_1$ and $[-f_1]_1$ are "far apart" in the topology of M, and yet both lie in M_1 and can be connected by a path in M.

Now we begin to prove the famous "monodromy theorem," which essentially states that the phenomenon just discussed cannot occur on simply connected regions. A consequence will be the fact that on any simply connected region in $\mathbb{C}$ which does not contain the origin one can define a (single-valued) analytic determination of $z^{1/m}$, $\log z$, etc. First we introduce the notation

$$I = [0,1] ,$$
$$I^2 = [0,1]\times[0,1] .$$

<u>LEMMA 1</u>. <u>Let</u> $\Gamma: I^2 \to \mathbb{C}$ <u>be continuous</u>, <u>and let</u> $\tilde{\Gamma}: I^2 \to M$ <u>satisfy</u> $\pi\circ\tilde{\Gamma} = \Gamma$. <u>Assume that for each fixed</u> $u \in I$, $\tilde{\Gamma}(t,u)$ <u>is a continuous function of</u> t; <u>and also that</u> $\tilde{\Gamma}(0,u)$ <u>is a continuous function of</u> u. <u>Then</u> $\tilde{\Gamma}$ <u>is continuous</u>.

<u>Proof</u>: This follows in a purely topological manner from the unique lifting theorem (Proposition 2) and the description of lifting in terms of analytic continuation given in Proposition 1. Let $\beta \in I$. We shall then prove that there exists $\varepsilon > 0$ such that $\tilde{\Gamma}$ is continuous on $I \times (I\cap(\beta-\varepsilon,\beta+\varepsilon))$, and the lemma will then be proved. For each fixed u define for $t \in I$

$$\gamma_u(t) = \Gamma(t,u) ,$$
$$\tilde{\gamma}_u(t) = \tilde{\Gamma}(t,u) .$$

Then γ_u and $\tilde{\gamma}_u$ are continuous and $\pi\circ\tilde{\gamma}_u = \gamma_u$. Now we apply Proposition 1, which guarantees the existence of a collection of disks $\Delta_1,\ldots,\Delta_n$, meromorphic f_k defined in Δ_k, $1\le k\le n$, and points t_k such that $0 = t_0<t_1<\ldots<t_n = 1$ and

$$\gamma_\beta([t_{k-1},t_k]) \subset \Delta_k \ ,$$

$$f_{k-1} \equiv f_k \text{ on } \Delta_{k-1} \cap \Delta_k \ ,$$

and

$$\tilde{\gamma}_\beta(t) = [f_k]_{\gamma_\beta(t)} \ , \quad t_{k-1}\le t\le t_k \ ,$$

the latter choice being forced as follows from the proof of Proposition 1 and the unique lifting theorem. Note that we have <u>fixed</u> β and applied Proposition 1 to the paths γ_β and $\tilde{\gamma}_\beta$.

Since Γ is <u>uniformly</u> continuous, there exists $\varepsilon_1 > 0$ such that

$$\gamma_u([t_{k-1},t_k]) \subset \Delta_k \text{ if } |u-\beta| < \varepsilon_1 \ , \ u\in I$$

(recall that Δ_k is an <u>open</u> disk). Also, since $\tilde{\Gamma}(0,u)$ is a continuous function of u, there exists $\varepsilon_2 > 0$ such that

$$\tilde{\Gamma}(0,u) \in U(a,f_1,\Delta_1) \text{ if } |u-\beta| < \varepsilon_2 \ , \ u\in I$$

(a = center of Δ_1). Thus, if $\varepsilon = \min(\varepsilon_1,\varepsilon_2)$,

$$\tilde{\gamma}_u(0) = [f_1]_{\gamma_u(0)} \text{ if } |u-\beta| < \varepsilon \ , \ u\in I.$$

The unique lifting theorem now implies that

$$\tilde{\gamma}_u(t) = [f_k]_{\gamma_u(t)}\ , \quad t_{k-1} \le t \le t_k\ , \quad |u-\beta|<\varepsilon\ , \quad u \in I.$$

That is,

$$\tilde{\Gamma}(t,u) = [f_k]_{\Gamma(t,u)}\ , \quad t_{k-1} \le t \le t_k\ , \quad |u-\beta|<\varepsilon\ , \quad u \in I.$$

But the continuity of $\Gamma(t,u)$ in the indicated range implies that of $\tilde{\Gamma}(t,u)$ in the same range.

QED

DEFINITION 6. If T is a topological space and $\gamma_0 \colon I \to T$ and $\gamma_1 \colon I \to T$ are paths in T having the same end points, then γ_0 and γ_1 are homotopic with fixed end points if there exists a continuous

$$\Gamma \colon I^2 \to T$$

such that

$$\Gamma(t,0) = \gamma_0(t)\ ,$$
$$\Gamma(t,1) = \gamma_1(t)\ ,$$
$$\Gamma(0,u) = \gamma_0(0) = \gamma_1(0)\ ,$$
$$\Gamma(1,u) = \gamma_0(1) = \gamma_1(1)\ .$$

The function Γ is called a homotopy between γ_0 and γ_1. If there is possibility of confusion we will say γ_0 and γ_1 are T-homotopic with fixed end points.

<u>DEFINITION 7</u>. A connected topological space T is <u>simply connected</u> if each pair of paths γ_0 and γ_1 in T having the same end points are homotopic with fixed end points.

Now we can state various trivial consequences of Lemma 1.

<u>Covering Homotopy Theorem</u>. <u>Let</u> Γ: $I^2 \to \mathcal{C}$ <u>be a homotopy in</u> $\mathcal{C}$ <u>and let</u> $p \in M$ <u>such that</u> $\pi(p) = \Gamma(0,u)$ $(0 \le u \le 1)$, <u>and suppose for each</u> $u \in I$ <u>the path</u> $t \to \Gamma(t,u)$ <u>can be lifted to a path in</u> M <u>starting at</u> p, <u>say</u> $\tilde{\Gamma}(t,u)$, <u>so that</u>

$$\pi \circ \tilde{\Gamma} = \Gamma \ .$$

<u>Then</u> $\tilde{\Gamma}$ <u>is a homotopy in</u> M.

<u>Proof</u>: For each $u \in I$ the function $t \to \tilde{\Gamma}(t,u)$ is continuous, by hypothesis. And $\tilde{\Gamma}(0,u) = p$ is constant and thus a continuous function of u. Therefore, Lemma 1 implies $\tilde{\Gamma}$ is continuous. Finally, consider the two paths

$$\tilde{\gamma}(u) = \tilde{\Gamma}(1,u) \ ,$$
$$\tilde{\gamma}'(u) = \tilde{\Gamma}(1,0) \ .$$

We then have

$$\tilde{\gamma}(0) = \tilde{\gamma}'(0)$$

and

$$\pi \circ \tilde{\gamma}(u) = \Gamma(1,u) = \Gamma(1,0) = \pi \circ \tilde{\gamma}'(u).$$

Thus Proposition 2 implies $\tilde{\gamma} = \tilde{\gamma}'$. That is, $\tilde{\Gamma}(1,u) \equiv \tilde{\Gamma}(1,0)$, and thus $\tilde{\Gamma}$ is a homotopy.

QED

Monodromy Theorem. Let D be a simply connected region in $\mathbb{C}$, $a \in D$, and f a meromorphic function in a neighborhood of a. Assume that f has an analytic continuation along every path in D which starts at a. Then there exists a meromorphic function F on D such that $f \equiv F$ in a neighborhood of a.

Proof: The hypothesis means that for every path γ in D such that $\gamma(0) = a$, there exists a path $\tilde{\gamma}$ in M such that $\pi \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(0) = [f]_a$. If γ_0 and γ_1 are paths in D from a to z, then γ_0 and γ_1 are D-homotopic with fixed end points, and by the covering homotopy theorem the paths $\tilde{\gamma}_0$ and $\tilde{\gamma}_1$ are homotopic with fixed end points; in particular, $\tilde{\gamma}_0(1) = \tilde{\gamma}_1(1)$. Thus, we can define unambiguously

$$F(z) = V(\tilde{\gamma}(1)) ,$$

where $\tilde{\gamma}$ is a path in M such that $\tilde{\gamma}(0) = [f]_a$ and $\pi \circ \tilde{\gamma}$ is a path in D from a to z. Now we must check the properties of F. First, suppose $\tilde{\gamma}(1) = [g]_z$, where g is holomorphic in a disk Δ centered at z. For $w \in \Delta$ we use the path which goes from a to z along γ and then from z to w along a line segment. The lifting from a to z is $\tilde{\gamma}$ and the lifting from z to w is just the germ of g at points on the segment from z to w. Since F is unambiguously defined, $F(w) = V([g]_w) = g(w)$. Thus, F is meromorphic in Δ and this proves F is meromorphic in D. In particular, if $z = a$ we can take $g = f$ and we obtain $F(w) = f(w)$ for w near a.

QED

One application of the monodromy theorem has already been mentioned. Namely, on a simply connected region $D \subset \mathbb{C} - \{0\}$, there exists a holomorphic determination of log z. The only hypothesis which needs to be checked is that log z can be analytically continued along all paths in D. This can be verified in a simple manner, but we omit the proof now since a slightly different version of the same result will be given in the discussion of algebraic functions in Chapter V.

We next want to give an example pertinent to the monodromy theorem, but we shall first give a rather simple but important theorem on analytic continuation, the so-called "permanence of functional relations." This is a generalization of a familiar result on single-valued functions, an example of which is the fact that the identity $\sin 2z = 2 \sin z \cos z$ follows from its validity for <u>real</u> z and the analyticity of all the functions involved. The theorem we shall give is really a generalization of usual theorems on unique analytic continuation because we are not here dealing with single-valued functions. Also, a more general theorem could be stated.

<u>Permanence of Functional Relations</u>. <u>Let</u> $A(z,w)$ <u>be a holomorphic function for</u> z <u>in a region</u> $D \subset \mathbb{C}$ <u>and all</u> $w \in \mathbb{C}$. <u>Let</u> $\tilde{\gamma}$ <u>be a path in</u> M <u>such that</u> $\gamma = \pi \circ \tilde{\gamma}$ <u>is a path in</u> D <u>and each</u> $t \in I$ <u>yields</u> $\tilde{\gamma}(t) = [f_t]_{\gamma(t)}$, <u>where</u> f_t <u>is a holomorphic function in a neighborhood of</u> $\gamma(t)$. <u>If</u> $A(z,f_0(z)) \equiv 0$ <u>in a neighborhood of</u> $\gamma(0)$, <u>then for each</u> $t \in I$, $A(z,f_t(z)) \equiv 0$ <u>in a neighborhood of</u> $\gamma(t)$.

<u>Remark</u>. We have not given a definition for a function A to be holomorphic in two complex variables. One definition states that for any (z_o,w_o) in the domain of definition of A, A has a power series expansion

$$A(z,w) = \sum_{j,k=0}^{\infty} a_{jk}(z-z_o)^j(w-w_o)^k$$

converging absolutely for z near z_o and w near w_o. The important property we need is that if f is a holomorphic function of one variable near z_o, then $A(z,f(z))$ is holomorphic for z near z_o. For the most important case we shall consider this is quite obvious; namely, the case in which the function $A(z,w)$ is a polynomial in w with coefficients holomorphic functions of z:

$$A(z,w) = a_0(z)w^n+\ldots+a_{n-1}(z)w+a_n(z).$$

Proof: Since f_t is holomorphic near $\gamma(t)$, the function $A(z,f_t(z))$ is holomorphic near $\gamma(t)$ and thus defines a germ at $\gamma(t)$ which we denote

$$\tilde{\gamma}_1(t) = [A(z,f_t(z))]_{\gamma(t)} \; .$$

Since $\tilde{\gamma}(t) = [f_t]_{\gamma(t)}$, it follows that $\tilde{\gamma}_1$ is a path in M (i.e., $\tilde{\gamma}_1$ is continuous) and obviously $\pi\circ\tilde{\gamma}_1 = \gamma$. Also define

$$\tilde{\gamma}_2(t) = [0]_{\gamma(t)} \; .$$

Then $\tilde{\gamma}_2$ is a path in M with $\pi\circ\tilde{\gamma}_2 = \gamma$. By hypothesis, $\tilde{\gamma}_1(0) = \tilde{\gamma}_2(0)$. Thus, Proposition 2 implies $\tilde{\gamma}_1 = \tilde{\gamma}_2$ and this implies the result.

QED

Before giving the example, let us make one important observation about analytic continuation. This is the fact that if two germs at a point a are different, then they remain different under analytic continuation along any fixed path. This is another consequence of Proposition 2, which in this case would read that if $\tilde{\gamma}_1(1) = \tilde{\gamma}_2(1)$, then $\tilde{\gamma}_1(0) = \tilde{\gamma}_2(0)$. Also, if $[f]_a$ is a germ at a and if f can be continued analytically along every path in a region D and if the continuation of f depends only on the terminal point of the path and not on the path itself, then there is a meromorphic F defined in D such that F = f near a. The proof of this is exactly like the proof of the monodromy theorem was once we knew that analytic continuation did not depend on the path (see pp. 64-65).

The monodromy theorem has of course two critical hypotheses. We have already indicated the reason for assuming D is simply connected, and now we shall examine the other main hypothesis, that f has an analytic continuation along <u>every</u> path in D. Note especially that the hypothesis does not state that f can be continued analytically to each point of D along <u>some</u> path in D. We shall now give an example to refute such a possibility for a weakening of the hypothesis of the theorem.

This example will be the Riemann surface for the "inverse" of the function $G(w) = w^3 - 3w$, and the analytic continuation process will reduce to finding paths on the surface. As $G'(w) = 3w^2 - 3$, the inverse function theorem of complex analysis will apply if $w \neq 1$ and $w \neq -1$. Since $G(1) = -2$ and $G(-1) = 2$, we conclude that if $G(w_o) = z_o \neq \pm 2$, then there exists a unique holomorphic function f in a neighborhood of z_o such that $G(f(z)) = z$ near z_o and $f(z_o) = w_o$. But for each $z_o \neq \pm 2$ there are three distinct corresponding values of w_o and thus three distinct solutions f of $G(f(z)) = z$ defined near z_o. We shall make this multiple-valued correspondence $z \to w$ into a single-valued function on an appropriate Riemann surface by the technique of the introduction, even though we no longer possess an explicit formula for w in terms of z. Thus, we select three copies of the z-plane cut along the real axis from 2 to ∞ and from -2 to ∞:

Each of these slit planes is simply connected, so the monodromy theorem applies to show that in each plane we can define a global solution f to the equation $G(f(z)) = z$ and f is holomorphic in the slit plane.

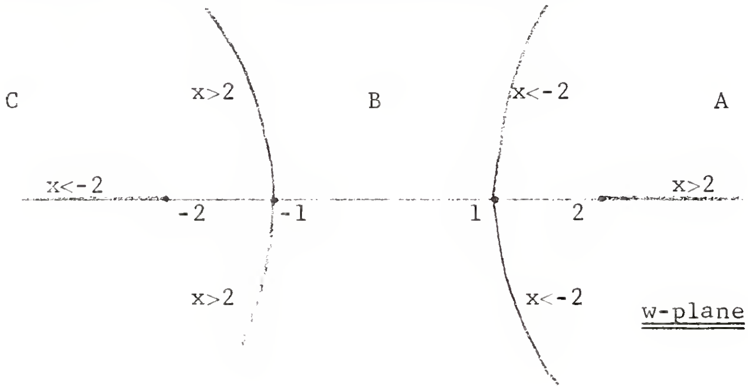
In order to accomplish the corresponding gluing we must see what happens to these functions at the slits. So we wish to examine carefully the values of w corresponding to real z such that $2<|z|<\infty$. To do this we introduce coordinates $z = x+iy$, $w = u+iv$ and compute from

$$(u+iv)^3 - 3(u+iv) = x+iy .$$

We find

$$u^3-3uv^2-3u = x \ ,$$
$$3u^2v-v^3-3v = y \ .$$

Along the slits we have $y = 0$, or $3v(u^2 - \frac{v^2}{3} - 1) = 0$. Thus, $v = 0$ or $u^2 - \frac{v^2}{3} = 1$. This locus in the w-plane looks like the real axis and a hyperbola:



For $v = 0$ we have $x = u^3-3u$. Thus, $x>2 \Leftrightarrow u>2$ and $x<-2 \Leftrightarrow u<-2$, as one easily sees by considering the graph of u^3-3u. For $u^2 - \frac{v^2}{3} = 1$ we have $x = u^3-3u(3u^2-3)-3u = -8u^3+6u$. Again, it is easily seen that $x>2 \Leftrightarrow u<-1$ and $x<-2 \Leftrightarrow u>1$.

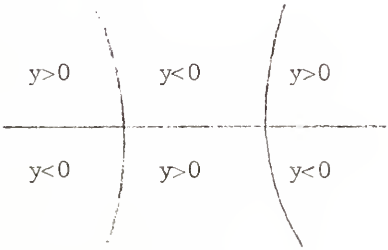
Now we distinguish three regions in the w-plane:

$$A = \{(u,v): u^2-\frac{v^2}{3}>1,\ u>0\} - \{(u,0): 2\leq u<\infty\} \ ,$$

$$B = \{(u,v): u^2-\frac{v^2}{3}<1\} \ ,$$

$$C = \{(u,v): u^2-\frac{v^2}{3}>1,\ u<0\} - \{(u,0): -\infty<u\leq-2\}.$$

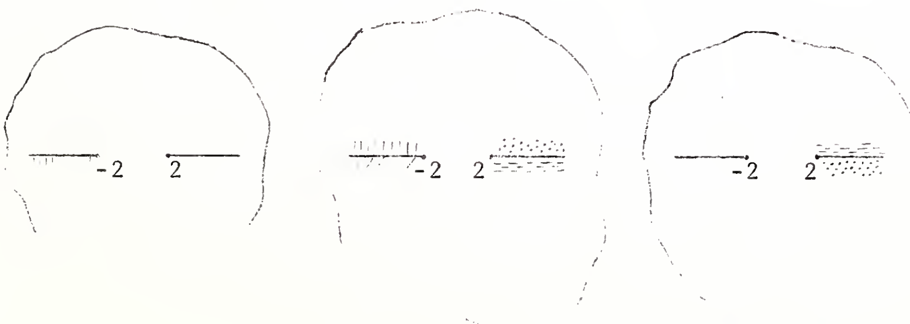
Then one easily sees that the function G maps A,B, and C each onto a copy of the z-plane, cut as described. Suppose we use three copies of the z-plane, labeled $\mathbb{C}_A$, $\mathbb{C}_B$, and $\mathbb{C}_C$. In order to see how these should be glued along the cuts, we just need to check the sign of y near the boundaries of A,B, and C in the w-plane. This is indicated in the figure.



$$y = 3v\left(u^2 - \frac{v^2}{3} - 1\right)$$

w-plane

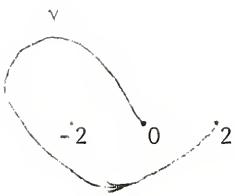
Now we can easily indicate the method of gluing the planes $\mathbb{C}_A, \mathbb{C}_B, \mathbb{C}_C$:



$\mathbb{C}_A$ $\mathbb{C}_B$ $\mathbb{C}_C$

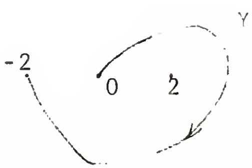
Note in particular that the cuts from 2 to ∞ in $\mathbb{C}_A$ and from -2 to ∞ in $\mathbb{C}_C$ can now be erased. This is the basic reason this example has been introduced. "Over" the point $z = 2$ lie two points of our Riemann surface, one a branch point, the other not. Likewise for $z = -2$.

Now we have a function f defined on this Riemann surface which represents all the solutions of $G(w) = z$ for any z. Now suppose we start at $z = 0$ with the solution f_o of the equation $G(f_o(z)) = z$ near $z = 0$, $f_o(0) = 0$, f_o holomorphic. Given any complex number a, there is some path γ from 0 to a along which f_o has an analytic continuation. If $a \neq \pm 2$, one can indeed go along any path from 0 to a which does not pass through ± 2. If $a = 2$, use the path:



Here is the reason. The starting point corresponds to $z = 0$, $w = 0$ and thus to the origin in $\mathbb{C}_B$. In order to get to the point 2 in $\mathbb{C}_A$ (where this is not a branch point), we pass through the cut joining $\mathbb{C}_B$ to $\mathbb{C}_A$.

Likewise, if $a = -2$, use the path



But the conclusion of the monodromy theorem fails. Otherwise, by the permanence of functional relations there would exist a function F holomorphic in all of $\mathbb{C}$ such that $G(F(z)) \equiv z$, $z \in \mathbb{C}$. It is rather clear that this cannot happen since by its very nature the relation $z \to w$ must be multiple valued. A direct proof would be this. Since $F^3 - 3F \equiv z$, we have $F(z) \to \infty$ as $z \to \infty$. Thus, F has a pole at ∞ and so the Laurent expansion of F at ∞ shows that $F(z) = \alpha z^n + \ldots$ (smaller powers of z), where $\alpha \neq 0$ and n is a positive integer. But then $F^3 - 3F = \alpha^3 z^{3n} + \ldots$ and there is no way this can behave like z near ∞.

We shall return to this example in Chapter V, where algebraic functions in general are treated. But it should even be noted here that the branch point 2 lying in $\mathbb{C}_B$ and $\mathbb{C}_C$ and the branch point -2 lying in $\mathbb{C}_A$ and $\mathbb{C}_B$ can be added to the surface in the way described in Chapter 1, and likewise ∞ in $\mathbb{C}_A, \mathbb{C}_B$, and $\mathbb{C}_C$ can be added, all three sheets being joined there. The resulting surface is a Riemann surface and the function f on it corresponding to the mapping $z \to w$ is meromorphic. Also, f is easily seen to be one-to-one since the inverse mapping $w \to z$ is single-valued. Therefore, since f is also onto, f is an analytic equivalence with $\hat{\mathbb{C}}$, so this Riemann surface is equivalent to $\hat{\mathbb{C}}$.

<u>Theorem of Poincaré and Volterra</u>. <u>Let</u> S <u>be a connected open subset of</u> M. <u>Then for any</u> $a \in \mathbb{C}$ <u>the set</u>

$$\{[f]_a : [f]_a \in S\} = S \cap \pi^{-1}(a)$$

<u>is countable or finite</u>.

<u>Proof</u>: Since S is connected we can consider some fixed $[g]_b \in S$ and then note that each element of $S \cap \pi^{-1}(a)$ can be connected to $[g]_b$ by a path $\tilde{\gamma}$ in S, by Proposition 1 of Chapter II. That is, if $\gamma = \pi \circ \tilde{\gamma}$, then analytic continuation of g along γ results in f, if $[f]_a$ is the point we are considering. By Proposition 1 it follows that if γ' is another path such that for a sufficiently small $\epsilon > 0$

$$|\gamma'(t) - \gamma(t)| < \epsilon \ , \quad 0 \le t \le 1 \ ,$$

then $[g]_{\gamma'(0)}$ can be analytically continued along γ' and the resulting germ is $[f]_{\gamma'(1)}$. We have again appealed to the unique lifting theorem and the argument used in the proof of Lemma 1. Now there is such a γ' with initial point b and terminal point a, such that γ' is a <u>polygon</u> with vertices (except for a and b) at <u>rational</u> complex numbers (i.e., complex numbers whose real and imaginary parts are both rational). Thus, $S \cap \pi^{-1}(a)$ consists of germs $[f]_a$ which come from analytic continuation from $[g]_b$ along paths which are polygons with rational vertices. There are only countably many such paths so the theorem is proved.

QED

Of course, the example which is immediately suggested by this theorem is the Riemann surface for log z, which has countably many sheets. In the language of germs, we have over a point $a \neq 0, \infty$, the germs $[\log z + 2n\pi i]_a$, where log z represents an arbitrary determination of the logarithm near a, and n is any integer.

DEFINITION 8. Let f be a meromorphic function in a neighborhood of a point $a \in \mathbb{C}$. The Riemann surface (in M) of f is the component of $[f]_a$ in M.

Here we have used a topological word "component," which by definition is a maximal connected set--a connected set contained in no strictly larger connected set. Since M is a surface, in this case the component containing $[f]_a$ (the component of $[f]_a$) is the collection of germs which can be joined to $[f]_a$ by a path (in M).

For example, the Riemann surface of any determination of $z^{1/m}$ near a point $a \neq 0$ consists of all germs $[f]_b$ such that $f(z)^m \equiv z$ near b, $b \neq 0$. By the permanence of functional relations all the germs in this Riemann surface must satisfy this identity, and we thus need only verify that any germ satisfying the identity can be joined to any other such germ. This can of course be easily checked directly, but an argument will be given in Chapter V for a general theorem along these lines.

There is an obvious deficiency in the Riemann surface for $z^{1/m}$. Namely, the branch points 0 and ∞ are missing. This situation is true in general for M--it has been constructed without branch points (a phrase which we haven't even yet defined), and also it does not contain germs of functions meromorphic at ∞. The latter is not a serious omission and indeed we could have considered from the start germs of meromorphic functions on any fixed Riemann surface. But in the next chapter we shall construct a Riemann surface which contains M in a very precise sense and has all the branch points and also the germs at ∞. Then we shall give a satisfactory definition of the Riemann surface of a meromorphic function, replacing Definition 8.

Chapter IV

BRANCH POINTS AND ANALYTIC CONFIGURATIONS

Before going to the definitions we give some motivating thoughts. The basic thing we want to do is give up the special role played by the independent variable. So consider $[f]_a$. This germ of course is determined by a meromorphic function f defined near a, the correspondence being written $z \to f(z)$. We could also consider z as depending on a complex parameter t and write for example $a + t \to f(a + t)$ as the correspondence, where now t is near zero. But also we could write $a + \sin t \to f(a + \sin t)$, or $a + e^{3t} - 1 \to f(a + e^{3t} - 1)$, etc. All these would be legitimate representations of f because the correspondence $t \to z$ indicated in each case is a conformal equivalence of a neighborhood of $t = 0$ onto a neighborhood of $z = a$. Thus, in general we could consider a <u>pair</u> of meromorphic functions

$$P(t) = a + \rho(t),$$

$$Q(t) = f(a + \rho(t)),$$

where ρ is a conformal equivalence of a neighborhood of 0 onto a neighborhood of 0. Thus, each small parameter value t corresponds uniquely to a value of z $(=P(t))$ near a and the corresponding value $Q(t)$ of f. We would not like to allow a representation of the form

$$P(t) = a + t^2,$$
$$Q(t) = f(a + t^2),$$

however. The reason is basically because two different values of t can give the same value of P. However, the thing that is <u>really</u> wrong here is that two different values of t can give the same value both of P and of Q. This will be an important observation in our preparation for the definition.

Now consider the Riemann surface in M for the function $z^{1/m}$. This consists of germs $[f]_a$, $a\neq 0$, such that f is some determination of $z^{1/m}$ near a. So we have a representation

$$P(t) = a + t,$$
$$Q(t) = (a + t)^{1/m} \quad \text{(some determination)},$$

for t near 0. Suppose $Q(0) = \alpha$ so that α is one of the $m^{\underline{th}}$ roots of a. Then we can introduce a new parameter τ by the equation

$$a + t = (\alpha + \tau)^m .$$

In fact, $\frac{dt}{d\tau}(\tau=0) = m\alpha^{m-1} \neq 0$. Thus, we can also represent $[f]_a$ by the pair of functions

$$P_1(\tau) = (\alpha + \tau)^m ,$$

$$Q_1(\tau) = \alpha + \tau .$$

In our desire to obtain a representation near the branch point, we would like to use a pair $P(t) = t$, $Q(t) = t^{1/m}$. Of course, this is not allowed, but the answer to the dilemma is obtained by just formally setting $\alpha = 0$ in the above formulas to obtain the pair

$$P_1(\tau) = \tau^m ,$$

$$Q_1(\tau) = \tau .$$

Note how useful such a pair is. We obtain all the values of $z^{1/m}$ just by using the m different solutions of $\tau^m = z$. These yield the same value of P_1 (regarded as the independent variable) for the m different corresponding values of Q_1. Thus in a very real sense we have introduced a point corresponding to the branch point 0, and it fits in very well with the regular points near 0. Of course, we again would allow parameter changes as before, so that the pair

$$P(t) = \rho(t)^m ,$$

$$Q(t) = \rho(t) ,$$

is regarded as equivalent to the pair P_1, Q_1 if ρ is a conformal equivalence, with $\rho(0) = 0$. And as before we do not allow a pair such as

$$P(t) = t^{2m} ,$$

$$Q(t) = t^2 ,$$

because different values of t can yield the same values for both P and Q.

Finally, we exhibit pairs which we want to imagine as germs at ∞. If f is meromorphic in a neighborhood of ∞, then we use the parameter t near 0 and let the independent variable be $z = \frac{1}{t}$. Thus we have

$$P(t) = t^{-1} ,$$

$$Q(t) = f(t^{-1}) ,$$

defined for t near 0. More generally, we can also consider ∞ as a branch point, yielding for the Riemann surface for $z^{1/m}$ the pair of functions

$$P(t) = t^{-m} ,$$

$$Q(t) = t^{-1} .$$

Now we are ready for the formal development.

DEFINITION 1. A parameter change is a function ρ holomorphic in a neighborhood of 0 such that

$$\rho(0) = 0 ,$$

$$\rho'(0) \neq 0 .$$

Equivalently, we could say that $\rho(0) = 0$ and ρ is one-to-one in a neighborhood of 0.

DEFINITION 2. A pair is an ordered couple of functions P, Q meromorphic in a neighborhood of 0 such that in a sufficiently small neighborhood of 0

1. P is not constant,
2. the mapping $t \to (P(t),Q(t))$ is one-to-one.

Examples of pairs have already been given. Here are some other examples. First, $(\sin t, \sin t)$ is a pair, although the points $t = 0$ and $t = \pi$ give the same value to both P and Q. Second, (t^m, t^n) is a pair if and only if $m \neq 0$, and either $n = 0$ and $m = \pm 1$ or $n \neq 0$ and m and n are relatively prime. The only thing which really needs checking here is that if m and n are relatively prime, then (t^m, t^n) is a pair. This follows because the Euclidean algorithm (see Chapter V) shows there exist integers m' and n' such that $mm' + nn' = 1$. Now suppose

$$(t_1^m,\ t_1^n) = (t_2^m,\ t_2^n).$$

Then $t_1^{mm'} = t_2^{mm'}$ and $t_1^{nn'} = t_2^{nn'}$. Multiplying, we obtain

$$t_1^{mm'+nn'} = t_2^{mm'+nn'} ;$$

$$t_1 = t_2 .$$

<u>DEFINITION 3</u>. Let (P,Q) and (P_1,Q_1) be pairs. Then (P,Q) is <u>equivalent</u> to (P_1,Q_1) if there exists a parameter change ρ such that the equations

$$P_1 = P\circ\rho \quad ,$$

$$Q_1 = Q\circ\rho \quad ,$$

are valid in a neighborhood of 0. If (P,Q) is equivalent to (P_1,Q_1), this will be written $(P,Q) \sim (P_1,Q_1)$.

PROBLEM 3. Suppose (P,Q) and (P_1,Q_1) are pairs. Prove that if there exists a function ρ holomorphic in a neighborhood of 0 such that $\rho(0) = 0$ and $P_1 = P \circ \rho$, $Q_1 = Q \circ \rho$ near 0, then ρ must be a parameter change. Also, ρ is uniquely determined (near 0) by these equations.

LEMMA 1. $\sim$ *is an equivalence relation.*

Proof: *Reflexive*: $(P,Q) \sim (P,Q)$ since $\rho(t) \equiv t$ works.

Symmetric: If $(P,Q) \sim (P_1,Q_1)$ and ρ satisfies

$$P_1 = P \circ \rho \quad ,$$
$$Q_1 = Q \circ \rho \quad ,$$

then also

$$P = P_1 \circ \rho^{-1} \quad ,$$
$$Q = Q_1 \circ \rho^{-1} \quad ,$$

near 0 and ρ^{-1} is holomorphic,

proving $(P_1,Q_1) \sim (P,Q)$.

Transitive: If $(P,Q) \sim (P_1,Q_1)$ and $(P_1,Q_1) \sim (P_2,Q_2)$ and we have parameter changes ρ and ρ_1 satisfying $P_1 = P\circ\rho$, $P_2 = P_1\circ\rho_1$, likewise for the Q's, then

$$P_2 = P_1\circ\rho_1 = P\circ\rho\circ\rho_1 \quad ,$$

$$Q_2 = Q_1\circ\rho_1 = Q\circ\rho\circ\rho_1 \quad ,$$

and $\rho\circ\rho_1$ is also a parameter change, showing that

$$(P,Q) \sim (P_2,Q_2).$$

QED

DEFINITION 4. An equivalence class of pairs is a *meromorphic element*. The *meromorphic element* containing a pair (P,Q) is designated $e(P,Q)$. Thus,

$$e(P,Q) = \{(P_1,Q_1) : (P_1,Q_1) \text{ is a pair and } (P,Q) \sim (P_1,Q_1)\}.$$

Define $\overline{M}$ to be the collection of all meromorphic elements.

DEFINITION 5. The two functions $\pi : \overline{M} \to \hat{C}$,
$V : \overline{M} \to \hat{C}$,

are given by the formulas

$$\pi(e(P,Q)) = P(0),$$
$$V(e(P,Q)) = Q(0).$$

We simply remark that π and V are well defined since if $(P,Q) \sim (P_1,Q_1)$, then clearly $P(0) = P_1(0)$ and $Q(0) = Q_1(0)$. The number $\pi(e(P,Q))$ is sometimes called the <u>center</u> of $e(P,Q)$, and $V(e(P,Q))$ is called the <u>value</u> of $e(P,Q)$.

Another remark which is simple but useful is that if (P,Q) is a pair and ρ is a parameter change, then $(P\circ\rho, Q\circ\rho)$ is also a pair and is therefore equivalent to (P,Q).

As has been indicated in the motivation for $\overline{M}$, we definitely wish to consider $M \subset \overline{M}$ in a natural manner. Of course, the way we do this is to define a function on M with values in $\overline{M}$ and prove this function is one-to-one. This means that each element of M is identified with an element of $\overline{M}$ in a one-to-one fashion, and the identification is this: to a germ $[f]_a$ we associate the meromorphic element $e(a + t, f(a + t))$. Now we prove this is a one-to-one function. Suppose $[g]_b$ is another germ and that $e(a + t, f(a + t)) = e(b + t, g(b + t))$. This means that there exists a parameter change ρ such that for t near 0

$$a + t = b + \rho(t) \quad ,$$

$$f(a + t) = g(b + \rho(t)) \quad .$$

The first equation implies $a = b$ and $\rho(t) = t$, and then the second equation implies $f(a + t) = g(a + t)$ for t near 0. Thus, $[f]_a = [g]_b$.

We now begin to topologize $\overline{M}$, then make $\overline{M}$ a surface, then a Riemann surface. We remark that as <u>sets</u> the inclusion $M \subset \overline{M}$ is an isomorphism of M onto its image in $\overline{M}$ (this we have just proved), and we will eventually see that as <u>Riemann surfaces</u> this is still true: the mapping of M onto its image in $\overline{M}$ will be seen to be an analytic equivalence.

Before beginning this program we wish to spell out a notational convenience. Frequently we shall write

$$e(P,Q) = e(P(t),Q(t))$$

to designate a meromorphic element. We have already used this type of notation in the discussion of $M \subset \overline{M}$, where we wrote $e(a + t, f(a + t))$. Of course, this means $e(P,Q)$, where $P(t) = a + t$, $Q(t) = f(a + t)$, but it would seem pedantic to be so strict with the notation and certainly would be confusing. We couldn't even use notation such as $e(t^2,t)$. In order to attempt to be consistent we shall try to use t for the dummy variable in an expression such as the above. Thus, for example,

$$e(P(t_o + t),\ Q(t_o + t))$$

stands for the meromorphic element $e(P_1,Q_1)$, where for small t

$$P_1(t) = P(t_o + t) \quad ,$$
$$Q_1(t) = Q(t_o + t) \quad .$$

DEFINITION 6. Let (P,Q) be a pair and assume P and Q are both meromorphic on a disk Δ. Then let $U(P,Q,\Delta)$ be the collection of meromorphic elements according to the formula

$$U(P,Q,\Delta) = \{e(P(t_o + t),Q(t_o + t)):t_o \in \Delta\}.$$

We have assumed Δ sufficiently small that the mapping $t \to (P(t),Q(t))$ is one-to-one on Δ (cf. Definition 2).

Note that by this latter assumption each couple $(P(t_o + t),\ Q(t_o + t))$ for $t_o \in \Delta$ is indeed a pair, and thus $U(P,Q,\Delta)$ makes sense.

The sets $U(P,Q,\Delta)$ will form a neighborhood basis of $e(P,Q)$ when Δ is allowed to vary over all sufficiently small disks centered at 0. Since $U(P,Q,\Delta)$ is not defined in terms of the equivalence class $e(P,Q)$ but rather in terms of the particular pair $(P,Q)\in e(P,Q)$, we shall need a lemma comparing two neighborhoods constructed with different but equivalent pairs.

LEMMA 2. *Suppose* $(P,Q) \sim (P_1,Q_1)$. *If* $U(P,Q,\Delta)$ *is defined, then there exists a disk* Δ_1 *centered at* 0 *such that*

$$U(P_1,Q_1,\Delta_1) \subset U(P,Q,\Delta).$$

Proof: By definition there exists a parameter change ρ such that $P_1 = P\circ\rho$, $Q_1 = Q\circ\rho$ in a disk Δ_1 centered at 0. We choose Δ_1 sufficiently small that ρ' never vanishes in Δ_1 and $\rho(\Delta_1)\subset\Delta$, and also that $U(P_1,Q_1,\Delta_1)$ is defined. Now let $e\in U(P_1,Q_1,\Delta_1)$. Then $e = e(P_1(t_o + t), Q_1(t_o + t))$ for some $t_o\in\Delta_1$. Now

$$P_1(t_o + t) = P(\rho(t_o + t)) = P\ (\rho(t_o) + \rho_1(t)),$$

where

$$\rho_1(t) = \rho(t_o + t) - \rho(t_o) \quad .$$

Note that $\rho_1(0) = 0$ and $\rho_1'(0) = \rho'(t_o) \neq 0$ and thus ρ_1 is a parameter change. Since also

$$Q_1(t_o + t) = Q(\rho(t_o) + \rho_1(t)) \quad ,$$

we conclude that $(P_1(t_o + t), Q_1(t_o + t)) \sim (P(\rho(t_o) + t), Q(\rho(t_o) + t))$. Thus,

$$e = e(P(\rho(t_o) + t), Q(\rho(t_o) + t)) \in U(P,Q,\Delta)$$

and this proves the lemma.

<u>QED</u>

<u>PROPOSITION 1</u>. <u>The collection of sets</u> $U(P,Q,\Delta)$ <u>is a system of basic neighborhoods for a topology on</u> $\overline{M}$.

<u>Proof</u>: Clearly any point $e(P,Q)$ in $\overline{M}$ belongs to $U(P,Q,\Delta)$, and just as on p. 52 we have two things to check:

1. Suppose there are given $U(P,Q,\Delta)$ and $U(P_1,Q_1,\Delta_1)$, basic sets defined in terms of pairs $(P,Q) \sim (P_1,Q_1)$. By Lemma 2 there exists a disk Δ_2 centered at 0 such that $\Delta_2 \subset \Delta_1$ and $U(P_1,Q_1,\Delta_2) \subset U(P,Q,\Delta)$. Thus,

$$U(P,Q,\Delta) \cap U(P_1,Q_1,\Delta_1) \supset U(P_1,Q_1,\Delta_2).$$

2. Suppose $e(\tilde{P},\tilde{Q}) \in U(P,Q,\Delta)$. Then for a point $t_o \in \Delta$, $(\tilde{P},\tilde{Q}) \sim (P(t_o + t),Q(t_o + t))$. If Δ' is the disk centered at 0 whose radius is the radius of Δ minus $|t_o|$, then it is clear that

$$U(P(t_o + t),Q(t_o + t),\Delta') \subset U(P,Q,\Delta).$$

By Lemma 2 there exists a disk $\tilde{\Delta}$ centered at 0 such that $U(\tilde{P},\tilde{Q},\tilde{\Delta}) \subset U(P(t_o+t),Q(t_o+t),\Delta')$. Thus,

$$U(\tilde{P},\tilde{Q},\tilde{\Delta}) \subset U(P,Q,\Delta).$$

<u>QED</u>

Before proving that $\overline{M}$ is a Hausdorff space, we introduce some <u>normal</u> representations for meromorphic elements. Suppose we consider an element $e(P,Q)$. The discussion of pp. 37-41 defines the multiplicity m of P at 0 and shows a particularly simple form P has in terms of a judiciously chosen chart for the Riemann surface (a neighborhood of 0 in this case). Thus, in the present framework we conclude that there exists a parameter change ρ such that near $t = 0$

$$P(\rho(t)) = P(0) + t^m \quad \text{if} \quad P(0) \neq \infty,$$

$$P(\rho(t)) = t^{-m} \quad \text{if} \quad P(0) = \infty.$$

Thus, if $Q_1 = Q\circ\rho$, we see that

$$e(P,Q) = e(P(0) + t^m, Q_1) \quad \text{if} \quad P(0) \neq \infty,$$

$$e(P,Q) = e(t^{-m}, Q_1) \quad \text{if} \quad P(0) = \infty.$$

Note that the integer m is well defined, being the multiplicity of P. For if P_1 is derived from P by means of any parameter change, then P_1 has the same multiplicity m.

A point of $\overline{M}$ of the form $e(a + t^m, Q)$ or $e(t^{-m}, Q)$ is called a <u>branch point of order</u> $m-1$.

It should be remarked that the normal form is not unique if $m>1$. In fact, if ω is any root of $\omega^m = 1$, then for example

$$(a + t^m, Q(t)) \sim (a + t^m, Q(\omega t))$$

as the parameter change $t \to \omega t$ shows. Thus, e.g. $e(t^2,t) = e(t^2,-t)$. This is the only possible type of ambiguity.

<u>PROPOSITION 2</u>. $\overline{M}$ <u>is a Hausdorff space</u>.

<u>Proof</u>: Compare pp. 52-53. The fact that M is Hausdorff is an obvious and immediate consequence of the uniqueness of analytic continuation. The present proof is surprisingly more involved. Suppose that $e(P,Q)$ and $e(P_1,Q_1)$ are not contained in disjoint

neighborhoods. We can assume both these elements to be in normal representation, so that

$$P(t) = a + t^m \quad \text{or} \quad t^{-m}$$

and

$$P_1(t) = b + t^n \quad \text{or} \quad t^{-n} \quad .$$

Let Δ_k be the disk centered at 0 with radius k^{-1}. Then for any sufficiently large k the neighborhoods $U(P,Q,\Delta_k)$ and $U(P_1,Q_1,\Delta_k)$ have a common point, say

$$e(P(s_k + t),Q(s_k + t)) = e(P_1(t_k + t),Q_1(t_k + t)),$$

where $s_k, t_k \in \Delta_k$. In particular π and V have the same value at these two points, so

$$P(s_k) = P_1(t_k), \qquad Q(s_k) = Q_1(t_k).$$

If ever $s_k = t_k = 0$, then $e(P,Q) = e(P_1,Q_1)$, which is what we're trying to prove. Thus, we can assume s_k or $t_k \neq 0$. Now letting $k \to \infty$ implies first $P(0) = P_1(0)$, so we have either

$$P(t) = a + t^m \text{ and } P_1(t) = a + t^n,$$

or

$$P(t) = t^{-m} \quad \text{and} \quad P_1(t) = t^{-n} .$$

We shall eventually prove that $m = n$, so then $P = P_1$. Also, we see immediately that in either case $s_k^m = t_k^n$.

Choose arbitrary $n^{\underline{th}}$ roots of s_k, say

$$\sigma_k^n = s_k \quad .$$

Then

$$\left(\frac{\sigma_k^m}{t_k}\right)^n = \frac{s_k^m}{t_k^n} = 1.$$

Since there are only n choices for each number $\dfrac{\sigma_k^m}{t_k}$, we can choose a subsequence of k's such that these numbers are all equal to a common $n^{\underline{th}}$ root of 1, say ω^{-1}. Renaming this subsequence, it follows that we can assume

$$\omega\sigma_k^m = t_k \quad .$$

Then

$$Q(\sigma_k^n) = Q_1(\omega\sigma_k^m) \quad .$$

Since this equation is valid for $\sigma_k \to 0$, $\sigma_k \neq 0$, we can now apply the uniqueness of analytic continuation to conclude

$$Q(s^n) \equiv Q_1(\omega s^m), \quad s \text{ small} .$$

Thus, note that

$$(P(s^n),Q(s^n)) \equiv (P_1(\omega s^m),Q_1(\omega s^m)), \quad s \text{ small}.$$

Since the mapping $t \to (P(t),Q(t))$ is one-to-one (small t), then the mapping $s \to (P(s^n),Q(s^n))$ is <u>exactly</u> n-to-one (small s). Likewise, the mapping $s \to (P_1(\omega s^m), Q_1(\omega s^m))$ is <u>exactly</u> m-to-one. Since these mappings are identical, we must have $\underline{\underline{m = n}}$!

Thus, $Q(s^m) \equiv Q_1(\omega s^m)$ and we conclude

$$Q(t) \equiv Q_1(\omega t).$$

Since $P(t) \equiv P_1(\omega t)$ (as $\omega^m = 1$), we have

$$(P,Q) \sim (P_1,Q_1),$$

the parameter change being just $\rho(t) = \omega t$. Thus,

$$e(P,Q) = e(P_1,Q_1).$$

<u>QED</u>

Now we have certain obvious charts for $\overline{M}$. Namely, we define the mapping

$$\varphi : U(P,Q,\Delta) \to \Delta$$

by the formula

$$\varphi(e(P(t_o + t),Q(t_o + t))) = t_o \quad ,$$

or

$$\varphi^{-1}(t_o) = e(P(t_o + t),Q(t_o + t)).$$

The definition of φ^{-1} is of course clear enough, but for φ to be well defined something must be checked. Namely, if two points in $U(P,Q,\Delta)$ are the same, then they correspond to the same t_o. Another way of saying this is that φ^{-1} is one-to-one. But if $\varphi^{-1}(t_o) = \varphi^{-1}(t_o')$, then $\pi(\varphi^{-1}(t_o)) = \pi(\varphi^{-1}(t_o'))$ and $V(\varphi^{-1}(t_o)) = V(\varphi^{-1}(t_o'))$, so that $P(t_o) = P(t_o')$ and $Q(t_o) = Q(t_o')$. Since the mapping $t \to (P(t),Q(t))$ is one-to-one for $t \in \Delta$, this implies $t_o = t_o'$. This shows that at least φ maps $U(P,Q,\Delta)$ to Δ in a bijective fashion (one-to-one and onto).

It is now easy to see that φ is a homeomorphism. In fact, if $e_o = e(P(t_o + t),Q(t_o + t))$ is any point in $U(P,Q,\Delta)$, then a neighborhood basis for e_o consists of the sets $U(P(t_o + t), Q(t_o + t), \Delta') = \{e(P(t_o + t_1 + t), Q(t_o + t_1 + t)) : t_1 \in \Delta'\}$, where Δ' is a sufficiently small disk centered at 0. The image of a set like this under the mapping φ is precisely $\{t_o + t_1 : t_1 \in \Delta'\}$, and these sets form a neighborhood basis for the point $t_o \in \Delta$. Thus, φ induces a one-to-one correspondence between a neighborhood basis for e_o and a neighborhood basis for $\varphi(e_o)$. Thus, φ is a homeomorphism. Thus, $\overline{M}$ is a surface.

PROPOSITION 3. *The given charts form an analytic atlas for* $\overline{M}$. *Thus,* $\overline{M}$ *is a Riemann surface*.

Proof: Suppose two neighborhoods $U(P, \dot{Q}, \Delta)$ and $U(P_1, Q_1, \Delta_1)$ meet. Let φ and φ_1 be the respective charts. Let e_o be a common point in these neighborhoods and $\varphi(e_o) = t_o$, $\varphi_1(e_o) = t_1$. We need to check the analyticity of $\varphi_1 \circ \varphi^{-1}$ in a neighborhood of t_o. Now by definition

$$e_o = e(P(t_o + t), Q(t_o + t)) = e(P_1(t_1 + t), Q_1(t_1 + t)).$$

Therefore there exists a parameter change ρ such that for t near 0

$$P(t_o + t) = P_1(t_1 + \rho(t)),$$

$$Q(t_o + t) = Q_1(t_1 + \rho(t)).$$

For z near t_o we have

$$\varphi^{-1}(z) = e(P(z + t),Q(z + t)).$$

We need to express this in terms of P_1 and Q_1, rather than P and Q. So we compute as follows:

$$P(z + t) = P(t_o + (z-t_o+t)) = P_1(t_1 + \rho(z-t_o+t))$$

$$= P_1(t_1 + \rho(z-t_o) + [\rho(z-t_o+t) - \rho(z-t_o)]).$$

The function

$$\rho_1(t) = \rho(z-t_o+t) - \rho(z-t_o)$$

satisfies $\rho_1(0) = 0$ and $\rho_1'(0) = \rho'(z - t_o)$. Thus, $\rho_1'(0) \neq 0$ if $z - t_o$ is sufficiently small, so ρ_1 is also a parameter change. Since the same computation is valid for Q and Q_1, we obtain

$$\varphi^{-1}(z) = e(P_1(t_1 + \rho(z - t_o) + t),Q_1(t_1 + \rho(z - t_o) + t)).$$

Therefore,

$$\varphi_1 \circ \varphi^{-1}(z) = t_1 + \rho(z - t_o),$$

which holds for all z sufficiently near t_o. Since

ρ is holomorphic, we have now proved that $\varphi_1 \circ \varphi^{-1}$ is holomorphic near t_o.

QED

Now we list various properties of $\overline{M}$.

PROPOSITION 4. $\overline{M} - M$ *is a discrete set. That is, each point of* $\overline{M}$ *has a neighborhood consisting only of itself and points of* M.

Proof: Suppose $e(P,Q) \in \overline{M}$ and that Δ is a disk centered at 0 such that $U(P,Q,\Delta)$ is defined and for $t \in \Delta - \{0\}$, $P'(t) \neq 0$, $P(t) \neq \infty$. Then $U(P,Q,\Delta)$ is a neighborhood of $e(P,Q)$ having the required properties. Indeed, if $e_o \in U(P,Q,\Delta)$ but $e_o \neq e(P,Q)$, then for some $t_o \in \Delta - \{0\}$

$$e_o = e(P(t_o + t), Q(t_o + t)).$$

Now $t \to P(t_o + t)$ is holomorphic and one-to-one near $t = 0$, so

$$\rho(t) = P(t_o + t) - P(t_o)$$

is a coordinate change. Thus, if $a = P(t_o)$

$$e_o = e(P(t_o) + t, Q(t_o + \rho^{-1}(t)))$$

$$= [Q(t_o + \rho^{-1}(z - a)]_a,$$

that is, e_o is the germ of the meromorphic function $z \to Q(t_o + \rho^{-1}(z - a))$ at a. Thus, $e_o \in M$.

QED

PROBLEM 4. Prove that π and V are meromorphic functions on $\overline{M}$. Prove that the mapping which identifies M as a subset of $\overline{M}$ is an analytic equivalence of M onto its image in $\overline{M}$.

The second half of this problem completely justifies regarding M as a subset of $\overline{M}$. Of course, we previously could only consider $M \subset \overline{M}$ as sets, but now also as Riemann surfaces. Also, M is open in $\overline{M}$, as is implied by Proposition 4.

PROPOSITION 5. *If* $e \in \overline{M}$, *then* e *is a branch point of order* $m - 1$ *if and only if* $m_\pi(e) = m$ (definition on p. 38).

Proof: Suppose $e = e(a + t^m, Q)$, and $\varphi: U(a+t^m, Q, \Delta) \to \Delta$ is a related chart. Then

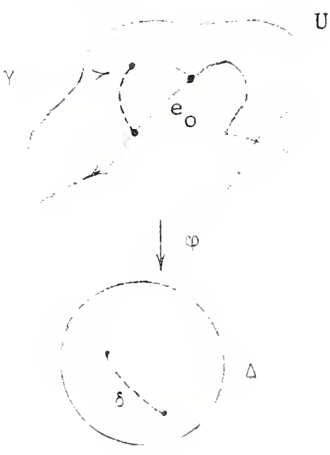
$$\pi \circ \varphi^{-1}(t_o) = a + t_o^m \quad ,$$

so that the multiplicity of $\pi \circ \varphi^{-1}$ at 0 is m. A similar computation applies if $e = e(t^{-m}, Q)$.

QED

PROPOSITION 6. *Any two points in the same component of $\overline{M}$ can be joined by a path in $\overline{M}$ every point of which except the initial and terminal points lies in* M.

Proof: This is a topological consequence of Proposition 4. Suppose γ is a path. Since $\gamma(I)$ is compact and $\overline{M} - M$ is discrete and closed, $\gamma(I) \cap (\overline{M} - M)$ is finite. Let e_0 be a point in this set which is not an initial or terminal point of γ. Let t_0 be the smallest and t_1 the largest numbers t in $(0,1)$ such that $\gamma(t) = e_0$. Choose a neighborhood U of e_0 and a chart $\varphi: U \to \Delta$, where $\Delta \subset \mathbb{C}$ is a disk.



Choose $0 < t_0' < t_0 \le t_1 < t_1' < 1$ such that $\gamma(t_0') \in U$, $\gamma(t_1') \in U$. Choose a path δ in Δ joining $\varphi \circ \gamma(t_0')$ and $\varphi \circ \gamma(t_1')$ and missing $\varphi(e_0)$. Then let

$$\gamma_1(t) = \begin{cases} \gamma(t), & 0 \le t \le t_0', \\ \varphi^{-1} \circ \delta\left(\dfrac{t - t_0'}{t_1' - t_0'}\right), & t_0' \le t \le t_1', \\ \gamma(t), & t_1' \le t \le 1. \end{cases}$$

Then γ_1 is a path in $\overline{M}$ having the same end points as γ, but γ_1 does not pass through e_o. Since we need to remove only finitely many points like e_o, the theorem follows.

QED

PROPOSITION 7. There is a natural one-to-one correspondence between components of M and components of $\overline{M}$. Namely, if S is a component of M, S is contained in a unique component of $\overline{M}$, which is the closure of S in $\overline{M}$; conversely, if $\overline{S}$ is a component of $\overline{M}$, then $\overline{S}$ contains a unique component of M, which is $\overline{S}\cap M$.

Proof: Let S be a component of M. Certainly S is contained in a unique component $\overline{S}$ of $\overline{M}$. Since components are closed, $\overline{S}$ contains the closure of S. But also any $e\in\overline{S}$ can be joined to a fixed $e_o\in S$ by a path γ such that $\gamma(0) = e_o$, $\gamma(1) = e$, $\gamma([0,1))\subset M$, by Proposition 6. Since $\gamma([0,1))$ is connected and $\gamma(0)\in S$, it follows that $\gamma([0,1))\subset S$ (since S is a component). Thus, e is a limit point of S and thus belongs to the closure of S, showing $\overline{S}$ is contained in the closure of S.

Conversely, let $\overline{S}$ be a component of $\overline{M}$. If a component of M is contained in $\overline{S}$, this component is also contained in $\overline{S}\cap M$. Thus, it suffices to show

$\overline{S}\cap M$ is a component of M. Proposition 6 shows that $\overline{S}\cap M$ is connected, for if e and e are in $\overline{S}\cap M$, they can be joined by a path in M. Since the end points are in $\overline{S}$, the entire path is in $\overline{S}$ ($\overline{S}$ is a *component*) and thus is in $\overline{S}\cap M$. And if a point of M can be joined by a path in M to a point in $\overline{S}\cap M$, that point must be in $\overline{S}$ and thus in $\overline{S}\cap M$. Thus, $\overline{S}\cap M$ is a component.

QED

DEFINITION 7. A component of $\overline{M}$ is called an *analytic configuration*. This is a translation of the term "analytische Gebilde" used by Weyl. Another term is *analytic entity*.

DEFINITION 8. Let f be a meromorphic function in a neighborhood of a point $a\in\mathbb{C}$. The *Riemann surface of* f is the analytic configuration containing $[f]_a$.

This definition is finally the complete idea which was begun in Definition 8 of Chapter III. We have now included the branch points in the surface and nothing else needs to be added.

It is important to observe that the nice analytic continuation or lifting properties of M do not hold in $\overline{M}$. For example, Proposition 2, the unique lifting theorem, of Chapter III would be false if phrased for $\overline{M}$. Just consider a neighborhood of a branch point to see this. For example, let

$$\tilde{\gamma}_1(s) = e((s+t)^2,\ s+t), \quad -1 \le s \le 1 \quad ,$$

$$\tilde{\gamma}_2(s) = e(|s|+t)^2,\ |s|+t), \quad -1 \le s \le 1 \quad .$$

Then $\pi \circ \tilde{\gamma}_1(s) = \pi \circ \tilde{\gamma}_2(s) = s^2$, but

$$\tilde{\gamma}_1(s) = \tilde{\gamma}_2(s) \text{ for } 0 \le s \le 1 \quad ,$$
$$\tilde{\gamma}_1(s) \ne \tilde{\gamma}_2(s) \text{ for } -1 \le s < 0 \quad .$$

We thus are led to a pictoral idea of branch point: two liftings of a given path in $\mathbb{C}$ which begin at a common point in $\overline{M}$ must be the same until a branch point is reached. But then the liftings can branch into several different paths in $\overline{M}$.

If f is a meromorphic function which is not one-to-one, it of course has no inverse. But we can easily consider the Riemann surface inverse to its Riemann surface, as follows.

PROPOSITION 8. *Let* S *be an open connected subset of* $\overline{M}$ *such that* V *is not constant on* S. *Then the mapping*

$$i : S \to \overline{M} \qquad \text{(i for "inverse")}$$

<u>defined by</u>

$$i(e(P,Q)) = e(Q,P)$$

<u>is an analytic equivalence of</u> S <u>with</u> $i(S)$.

<u>Proof</u>: First, note that i is well defined. This depends on the obvious fact that if $(P,Q) \sim (P_1,Q_1)$, then $(Q,P) \sim (Q_1,P_1)$. Now we prove i is analytic. To do this we introduce charts in the canonical way:

$$\varphi : U(P,Q,\Delta) \to \Delta \quad ,$$

$$\psi : U(Q,P,\Delta) \to \Delta \quad ,$$

where

$$\varphi^{-1}(t_o) = e(P(t_o+t),Q(t_o+t)),$$

$$\psi^{-1}(t_o) = e(Q(t_o+t),P(t_o+t)).$$

Then

$$\psi \circ i \circ \varphi^{-1}(t_o) = \psi(e(Q(t_o+t),P(t_o+t))) = t_o,$$

so that trivially $\psi \circ i \circ \varphi^{-1}$ is analytic. Thus, i is analytic.

Since i is analytic, $i(S)$ is an open connected subset of $\overline{M}$, and π is not constant on $i(S)$ since V is not constant on S. Furthermore, $i:i(S) \to S$ is analytic by what we have already proved and $i \circ i$ = identity. Thus, the inverse of i is analytic.

<u>QED</u>

We now give an interesting and rather surprising application of some of these ideas.

<u>DEFINITION 9</u>. Let f be meromorphic on an open set $D \subset \mathbb{C}$ and let $w \in \hat{\mathbb{C}}$. Then w is an <u>asymptotic value</u> of f if there exists a path of the form

$$\gamma:[0,\sigma) \to D \quad \text{(where } 0<\sigma\leq\infty)$$

such that

$$\lim_{s\to\sigma} f(\gamma(s)) = w$$

and $\gamma \to \partial D$, meaning that for every compact set $K \subset D$, there exists s_0 such that $\gamma((s_0,\sigma)) \subset D-K$.

<u>THEOREM 1</u>. <u>If f is holomorphic on an open set</u> $D \subset \mathbb{C}$, <u>then there exists an asymptotic value of</u> f.

Proof: Clearly it suffices to treat the case in which D is connected and f is not constant on D. Now define

$$S = \{[f]_a : a \in D\} \quad .$$

Clearly, S is an open subset of the sheaf of germs M, and the mapping $\pi : S \to D$ is an analytic equivalence. Now we complicate the situation by regarding $S \subset \overline{M}$ and letting $T = i(S)$ in the sense of Proposition 8. Then $V : T \to D$ is an analytic equivalence since $V = \pi \circ i^{-1}$. According to the definition of $M \subset \overline{M}$, we have

$$T = \{e(f(a+t), a + t) : a \in D\} \quad .$$

Thus, $f \circ V = \pi$ on T. That is, we have a commutative diagram

$$\begin{tikzcd} & T \arrow[dl, "V"'] \arrow[d, "\pi"] \\ D \arrow[r, "f"] & C \end{tikzcd}$$

and V represents the multiple-valued inverse of f. Note that a point $e(f(a+t), a+t)$ is a branch point (order at least 1) if and only if $f'(a) = 0$, and this holds only for a discrete and thus countable set of points $a \in D$. Let $E = \{a \in D : f'(a) = 0\}$ and note that $f(E)$ is a countable subset of $\mathbb{C}$. Choose arbitrarily $a_o \in D - E$. Since $f(E)$ is countable and there are uncountably many rays from $f(a_o)$ to ∞, it follows that there exists a ray from $f(a_o)$ to ∞ which contains no point of $f(E)$. Let this ray be represented by a path $\alpha : [0,\infty) \to \mathbb{C}$, so that $\alpha(0) = f(a_o)$, $\alpha(s) \notin f(E)$, $\lim_{s\to\infty} \alpha(s) = \infty$.

Now we consider the process of lifting α to T by the mapping π. Note that if $e \in T$ and $\pi(e) \notin f(E)$, then $V(e) \notin E$ and thus $e \in M$. Thus, lifting α is a problem of lifting to M, not merely $\overline{M}$, and the unique lifting theorem obtains. Let s_o be the supremum of all numbers s_1 such that there exists a lifting $\tilde{\alpha}$ on $[0,s_1)$ with $\tilde{\alpha}(0) = e(f(a_o+t), a_o + t)$. Then there is a unique path $\tilde{\alpha}$ corresponding to the maximal s_o,

$$\tilde{\alpha} : [0, s_o) \to T$$

such that $\pi \circ \tilde{\alpha} = \alpha$ and $\tilde{\alpha}(0) = e(f(a_o+t), a_o + t)$. Then define $\gamma = V \circ \tilde{\alpha}$, so that γ is a path in D

such that

$$\lim_{s\to s_o} f(\gamma(s)) = \lim_{s\to s_o} f\circ V\circ\tilde{\alpha}(s) = \lim_{s\to s_o} \pi\circ\tilde{\alpha}(s)$$

$$= \lim_{s\to s_o} \alpha(s) = \alpha(s_o) \quad .$$

Here $\alpha(s_o) = \infty$ if $s_o = \infty$. Thus, the theorem follows if we know that γ leaves every compact set in D. If $s_o = \infty$ this is perfectly clear since $\lim_{s\to\infty} f(\gamma(s)) = \infty$. Suppose $s_o < \infty$, and suppose that for some compact $K\subset D$, γ does not eventually leave K. By the Bolzano- Weierstrass theorem, there exists a point $z_o\in K$ and a sequence $s_1 < s_2 < \ldots$, $s_n \to s_o$, such that $\gamma(s_n) \to z_o$. Since $V:T \to D$ is a homeomorphism, $\tilde{\alpha}(s_n) \to V^{-1}(z_o) = e(f(z_o+t), z_o+t)$. Since π is continuous, $f(z_o) = \lim \pi\circ\tilde{\alpha}(s_n) = \lim \alpha(s_n)$ $= \alpha(s_o)\notin f(E)$, so $e_o = e(f(z_o + t), z_o + t)\in T\cap M$. This contradicts the maximality of s_o, for the topology of M implies that a neighborhood U of e_o is homeomorphic by a homeomorphism φ to a disk Δ centered at $f(z_o)$ and φ is just the restriction of π to U. Therefore, if n is chosen such that $\alpha(s_n)\in\Delta$, . then for sufficiently small $\varepsilon > 0$ we can define

$$\tilde{\alpha}(s) = \varphi^{-1}\circ\alpha(s), \quad s_n \leq s < s_o + \varepsilon$$

and we obtain a lifting of the required sort past the supposedly maximal s_o. This contradiction shows that γ leaves every compact set in D.

QED

A comprehensive reference to questions of this sort can be found in MacLane, G. R., "Asymptotic values of holomorphic functions," Rice University Studies 49 (No. 1) 1963, pp. 1-83. The example we have just treated can be found on page 7 of MacLane's monograph.

Chapter V

ALGEBRAIC FUNCTIONS

What we are going to study in this section is solutions of an algebraic equation in two complex variables; i.e., equations of the form

$$A(z,w) = 0,$$

where A is a polynomial in z and w. The viewpoint is that we want to regard w as a function of z satisfying $A(z,w(z)) = 0$. Of course, we expect w to be multiple-valued and then we construct a Riemann surface on which a function like w can be defined. Examples of this procedure were given in the introduction. There we treated the following examples of A:

$$w^m - z,$$
$$w^2 - (z-a)(z-b),$$
$$(z-b)w^2 - (z-a),$$
$$w^2 - (z-a_1)(z-a_2)\dots(z-a_m);$$

also on pp. 68 ff. we discussed the polynomial

$$w^3 - 3w - z.$$

All the Riemann surfaces associated with these examples can be easily visualized as subsets of $\overline{M}$ and as such enjoy the topological property of compactness. The main fact to come out of this section is that algebraic equations always lead to compact surfaces and that, conversely, every compact analytic configuration has a unique algebraic equation associated with it.

It follows from general topological considerations that every compact orientable surface (as Riemann surfaces are) is homeomorphic to a sphere with a certain number g of handles and g is called the genus of the surface; cf. p. 13. Before analyzing algebraic equations, we shall discuss heuristically a remarkable formula involving the genus, the number of sheets, and the branching of a compact Riemann surface.

The Riemann-Hurwitz formula. Consider a compact analytic configuration S. We first discuss its Euler characteristic. This can be defined in terms of a "triangulation" of S. We do not wish to pause to define triangulation, but if f is the number of triangles (faces), e the number of edges, and v the number of vertices, then the Euler charactersitic is $v-e+f$. A theorem of topology is that this number is a topological invariant of the surface and equals $2-2g$:

$$v-e+f = 2-2g.$$

Now S has certain branch points $e_1,\dots,e_\ell$ of orders $b_1,\dots,b_\ell$, respectively, $b_j \geq 1$. Define

$$V = \sum_{j=1}^{\ell} b_j .$$

The number V is called the ramification index or total branching order of S. Also S has a certain number n of sheets when viewed as spread over $\hat{C}$; this is the number such that π takes every value in $\hat{C}$ n times ... see pp. 43-44. The Riemann-Hurwitz formula is

$$\boxed{\frac{V}{2} = n + g - 1.}$$

To prove this formula consider a triangulation of the sphere $\hat{C}$ such that every point $\pi(e_j)$ is a vertex. Let f,e, and v be the number of faces, edges, and vertices. Since $\hat{C}$ has genus 0, we have the Euler formula

$$v-e+f = 2.$$

Now consider the preimage by π of these triangles. By lifting the triangulation of $\hat{C}$ to S we obtain nf faces and ne edges in the triangulation of S, since S has n sheets. And each vertex which is not a $\pi(e_j)$ is lifted to n new vertices. But each vertex $\pi(e_j)$ does not get lifted to n new vertices. Rather, if z_o is one of these values, then $\pi^{-1}(\{z_o\})$ consists of exactly

$$n - \sum_{\pi(e_j)=z_0} b_j$$

distinct points. Thus, the number of vertices in the triangulation of S is

$$nv - V.$$

Therefore,

$$(nv-V) - ne + nf = 2-2g.$$

Since $v-e+f = 2$ we can write this relation as

$$2n - V = 2-2g,$$

and the assertion is proved.

Let us test this formula on some of the cases we have considered. For example, on p. 12 we treated

$$w^2 - (z-a_1)\ldots(z-a_m),$$

$a_1,\ldots,a_m$ distinct. If m is even there are branch points of order 1 at each a_j and nowhere else, so $V = m$ and thus

$$\frac{m}{2} = n + g - 1 = 2 + g - 1,$$

$$g = \frac{m-2}{2}.$$

If m is odd then also ∞ is a branch point of order 1 and so $V = m+1$ and $g = \frac{m-1}{2}$. These results agree with p. 13.

Next, consider the example w^3-3w-z discussed on pp. 68 - 73. There the points 2 and -2 are branch points of order 1 and ∞ is a branch point of order 2, so $V = 4$. Since $n = 3$, we find $g = 0$ and the Riemann surface is homeomorphic to a sphere. This again agrees with our earlier findings, for on p. 73 we discussed an analytic equivalence of the surface with $\hat{C}$.

Of course, we have not rigorously derived the formula, but we _have_ given a sketch of a rigorous proof. But the formula should prove useful as a check in working out other examples. Every time one sees a compact Riemann surface, he should try out this formula on it. Two things in the formula deserve special attention. One is that V is always an _even_ integer. The other is that a purely topological number g is equal to the number $\frac{V}{2} - n + 1$ which depends very much on features of the surface which are not purely topological.

Now we proceed to the analytical discussion of algebraic equations.

Problem 5. In the spirit of pp. 68-73, discuss the algebraic equation $(w^2-1)^2 - z = 0$.

Lemma 1. _Let_ $a_1,\ldots,a_n$ _be holomorphic on an open set_ $D \subset \hat{C}$ _and_ $A(z,w) = w^n + a_1(z)w^{n-1}+\ldots+a_{n-1}(z)w + a_n(z)$.

Suppose $z_0 \in D$ *and*

$$A(z_0, w_0) = 0,$$

$$\frac{\partial A}{\partial w}(z_0, w_0) \neq 0.$$

Then there exists a function f *holomorphic in a neighborhood of* z_0 *such that*

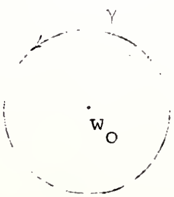
$$A(z, f(z)) \equiv 0, \quad z \text{ near } z_0,$$

$$f(z_0) = w_0,$$

$$A(z,w) = 0, \; z \text{ near } z_0, \; w \text{ near } w_0 \;\Rightarrow\; w = f(z).$$

Proof: This is merely an implicit function theorem and could be derived from the general implicit function theorem of differential calculus - we would just have to check the validity of the Cauchy-Riemann equation. However, the proof is much simpler in the present case than the proof of the general theorem and is even almost elegant, so we present it.

Since A is not constant in w, the zeros of $A(z_0, w)$ are isolated. Thus, there exists $\varepsilon > 0$ such that $A(z_0, w) \neq 0$ for $0 < |w - w_0| \leq \varepsilon$. Let γ be the path $\gamma(t) = w_0 + \varepsilon e^{2\pi i t}$, $0 \leq t \leq 1$. Since the image of γ



is compact and $A(z_0, w) \neq 0$ there, there exists $\delta > 0$ such that

$$A(z,w) \neq 0 \text{ for } |z - z_0| < \delta,$$
$$|w - w_0| = \varepsilon.$$

Therefore, the residue theorem implies that for each fixed z, $|z-z_0| < \delta$,

$$\frac{1}{2\pi i}\int_\gamma \frac{\frac{\partial A}{\partial w}(z,w)}{A(z,w)}\,dw$$

is equal to the number of zeros of A(z,w) (minus the number of poles of A(z,w)) for $|w-w_0| < \varepsilon$. And it is clear that this number is a continuous function of z for $|z-z_0| < \delta$, and is therefore constant. For $z = z_0$ we are counting the number of zeros of $A(z_0,w)$ in $|w-w_0| < \varepsilon$. Since $A(z_0,w) = 0$ only at $w = w_0$ and since w_0 is a <u>first order</u> zero ($\frac{\partial A}{\partial w}(z_0,w_0) \neq 0$), we have proved that the above integral is equal to 1 for $|z-z_0|<\delta$. Thus, $|z-z_0| < \delta$ implies there exists a unique f(z) such that $|f(z)-w_0| < \varepsilon$ and $A(z,f(z)) = 0$. Again, the residue theorem implies

$$f(z) = \frac{1}{2\pi i}\int_\gamma w\,\frac{\frac{\partial A}{\partial w}(z,w)}{A(z,w)}\,dw, \quad |z-z_0| < \delta.$$

From this formula it follows immediately that f is holomorphic. Of course, $f(z_0) = w_0$.

To prove uniqueness, suppose g is holomorphic near z_0 and $A(z,g(z)) = 0$, $g(z_0) = w_0$. Then by continuity of g it follows that there exists $0 < \delta_1 \leq \delta$ such that for $|z-z_0| < \delta_1$, $|g(z)-w_0| < \varepsilon$. Therefore, $g(z) = f(z)$ for

$|z-z_o| < \delta_1$.

QED

COROLLARY. *Suppose* $z_o \in D$ *and that there exists no* w *satisfying*

$$A(z_o,w) = 0,$$

$$\frac{\partial A}{\partial w}(z_o,w) = 0.$$

Then there exist unique holomorphic functions $f_1,\ldots,f_n$ *in a neighborhood of* z_o *such that*

$A(z,f_k(z)) \equiv 0$ *near* z_o, $1 \leq k \leq n$,

for each z *near* z_o, *the numbers* $f_1(z),\ldots,$ $f_n(z)$ *are distinct*.

Proof: Since $A(z_o,w)$ is a polynomial in w of degree n, it has n zeros. By hypothesis these zeros are distinct, say $A(z_o,w_k) = 0$, $1 \leq k \leq n$, $w_1,\ldots,w_n$ distinct. Apply Lemma 1 to $w_o = w_k$ to obtain the holomorphic solutions f_k. Since $f_1(z_o),\ldots,f_n(z_o)$ are distinct, it follows by continuity that for z sufficiently near z_o, $f_1(z),\ldots,f_n(z)$ are distinct.

QED

LEMMA 2. *Let* A *be defined as in Lemma* 1. *Assume that for every* $z \in D$ *there exists no* w *satisfying*

$$A(z,w) = 0,$$

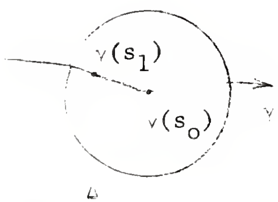
$$\frac{\partial A}{\partial w}(z,w) = 0.$$

Let f be a holomorphic function in a neighborhood of $z_o \in D$ satisfying

$$A(z,f(z)) \equiv 0 \text{ near } z_o.$$

Then f can be analytically continued along any path in D starting at z_o.

Proof: Let γ: $[0,1] \to D$ be a path with $\gamma(0) = z_o$. We are trying to prove the existence of a path $\tilde{\gamma}$: $[0,1] \to M$ such that $\tilde{\gamma}(0) = [f]_{z_o}$ and $\pi \circ \tilde{\gamma} = \gamma$. By the general discussion of analytic continuation we know that $\tilde{\gamma}$ exists on some interval $[0,t_o]$, $t_o > 0$, and that $\tilde{\gamma}$ is uniquely determined (Proposition 2 of Chapter III). Let s_o be the supremem of such t_o. Then $\tilde{\gamma}$ exists on the interval $[0,s_o)$. Now we apply the above corollary to the point $\gamma(s_o)$. obtaining holomorphic functions $f_1,\ldots,f_n$ in a neighborhood of $\gamma(s_o)$ satisfying the conclusion of the corollary on a disk Δ centered at $\gamma(s_o)$. Choose any $s_1 < s_o$ such that $\gamma(s_1) \in \Delta$. Then $\tilde{\gamma}(s_1) = [g]_{\gamma(s_1)}$, where g is holomorphic in a neighborhood of $\gamma(s_1)$ and by the permanence of functional relations (p. 66)



$$A(z,g(z)) \equiv 0 \text{ near } \gamma(s_1).$$

Thus, $g(z)$ is one of the n zeros of the polynomial $A(z,w)$ and must therefore be equal to one of the $f_k(z)$. Thus, by Lemma 1 and its corollary we find that for a unique k, $g(z) \equiv f_k(z)$, z near $\gamma(s_1)$. By the uniqueness of analytic continuation,

$$\tilde{\gamma}(s) = [f_k]_{\gamma(s)}, \quad s_1 \leq s < s_0.$$

This formula serves to define $\tilde{\gamma}$ for $s = s_0$ as well and even for $s > s_0$, $s-s_0$ sufficiently small, if $s_0 < 1$. Thus we conclude that $s_0 = 1$ and that $\tilde{\gamma}$ exists on $[0,1]$.

<u>QED</u>

<u>COROLLARY</u>. <u>In addition to the hypothesis of Lemma 2, assume that D is a simply connected region. Then there exist holomorphic functions</u> $f_1,\ldots,f_n$ <u>on</u> D <u>such that</u>

$A(z,f_k(z)) \equiv 0$ <u>for</u> $z \in D$, $1 \leq k \leq n$,

<u>for each</u> $z \in D$, <u>the numbers</u> $f_1(z),\ldots,$ $f_n(z)$ <u>are distinct</u>.

<u>Proof</u>: Use the corollary of Lemma 1 to obtain $f_1,\ldots,f_n$ near some point in D, say z . Use Lemma 2 and the <u>monodromy theorem</u> (p. 64) to obtain holomorphic extensions on all of D, noting that $A(z,f_k(z)) \equiv 0$ on D follows from the permanence of functional relations. If for some $z \in D$, $f_j(z) = f_k(z)$, Lemma 1 implies $f_j \equiv f_k$

near z and then $f_j \equiv f_k$ in D by analytic continuation, contradicting $f_j(z_o) \neq f_k(z_o)$ if $j \neq k$. Thus, $j = k$.

QED

The above corollary is about as far as we can go without really analyzing what happens near points z such that $A(z,w)$ has a multiple zero. To carry out such an analysis will require a little algebraic background, which we now begin.

First of all, what we shall be considering is functions A which are polynomials in z and w. It is always possible and frequently useful to arrange A according to powers of w or according to powers of z. Thus, we write

$$A(z,w) = a_0(z)w^n + a_1(z)w^{n-1}+\ldots+a_{n-1}(z)w + a_n(z),$$

where $a_0, a_1, \ldots, a_n$ are polynomials in z, and we assume $a_0 \neq 0$. We then say that A has degree n with respect to w. We say that a polynomial B is a factor of A if there exists another polynomial C such that $A = BC$. If A has no factors other than constants or constant multiples of A, we say that A is irreducible. It will also frequently be useful to factor a_0 from A, writing

$$A(z,w) = a_0(z)[w^n+\alpha_1(z)w^{n-1}+\ldots+\alpha_{n-1}(z)w+\alpha_n(z)],$$

where $\alpha_k = \frac{a_k}{a_0}$ is a rational function of z. Conversely, given rational functions of z, $\alpha_1,\ldots,\alpha_n$, we can let a_0 be the least common multiple of the denominators of $\alpha_1,\ldots,\alpha_n$, and use the above formula to define a polynomial A. This innocent statement will prove to be extremely useful in constructing polynomials. We shall frequently be able to construct holomorphic functions α_k on $\hat{C}$ minus a finite set, and by some argument show that α_k has no essential singularities in $\hat{C}$. Then we use the fact that a meromorphic function α_k on $\hat{C}$ must be rational; cf. p. 33, no. 9.

LEMMA 3. *Let A and B be polynomials in z and w which have no common nontrivial factor, and assume A, B $\neq$ 0. Then there are at most finitely many z such that there exists w such that*

$$A(z,w) = 0,$$

$$B(z,w) = 0.$$

Proof: We shall use the Euclidean algorithm. To do this it is most convenient to regard A and B as polynomials in w. Then we employ the factorization mentioned above to write

$$A = a_0(z)A'$$

$$B = b_0(z)B'$$

where

$$A'(z,w) = w^n + \alpha_1(z)w^{n-1} + \dots + \alpha_n(z),$$

$$B'(z,w) = w^m + \beta_1(z)w^{m-1} + \dots + \beta_m(z),$$

and $\alpha_1, \dots, \alpha_n$, $\beta_1, \dots, \beta_m$ are rational functions of z. We rely heavily on the fact that the rational functions of z form a field. Also, we write for short $\deg A' = n$ and $\deg B' = m$. By long division we have uniquely

$$A' = B'Q_1 + R_1, \quad \deg R_1 < \deg B'.$$

Here Q_1 and R_1 are polynomials in w with coefficients in the field of rational functions of z, and if $R_1 \equiv 0$ we set $\deg R_1 = -\infty$. If $R_1 \not\equiv 0$, we apply this again to obtain

$$B' = R_1Q_2 + R_2, \quad \deg R_2 < \deg R_1.$$

Continue this division process:

$$\begin{aligned} R_1 &= R_2Q_3 + R_3, \quad \deg R_3 < \deg R_2, \\ &\vdots \\ R_{k-2} &= R_{k-1}Q_k + R_k, \quad \deg R_k < \deg R_{k-1}, \\ R_{k-1} &= R_kQ_{k+1} . \end{aligned}$$

As indicated in this scheme, the process eventually terminates ($R_{k+1} \equiv 0$) since the degrees of the R_j's keep decreasing. We assume of course that $R_k \not\equiv 0$. Note that if $R_1 \equiv 0$, then B' is a factor of both A' and B', thus B is a polynomial in z alone and the conclusion of the lemma is trivial. Thus, we can assume

$R_1 \neq 0$. Working up through the above scheme, we see successively that R_k is a factor of R_{k-1}, thus $R_{k-2}, \ldots$, and finally R_k is a factor of B', and thus of A'. By hypothesis, R_k must have degree 0 in w, so R_k is just a rational function of z. Now we eliminate finitely many z by requiring that $a_0(z) \neq 0$, $b_0(z) \neq 0$, and z is not a pole of any of the coefficients of any of the polynomials $Q_1, \ldots, Q_k$, and $R_k(z) \neq 0$. Then we claim that there does not exist w such that $A(z,w) = 0$, $B(z,w) = 0$. For suppose such w exists. Then also $A'(z,w) = 0$, $B'(z,w) = 0$, since $a_0(z) \neq 0$, $b_0(z) \neq 0$. Since $Q_1(z,w) \neq \infty$, the first equation in our division scheme implies $R_1(z,w) = 0$. Likewise, $R_2(z,w) = 0$, and on down the line until we reach the contradiction $R_k(z) = 0$.

QED

Remark. Perhaps a cleaner way of giving this argument is to work up through the above equations to write

$$R_k = CA' + DB',$$

where C and D are polynomials in w with coefficients which are rational function of z. By clearing all the fractions out of this expression, we obtain

$$R = EA + FB,$$

where R is a not identically vanishing polynomial in z alone, and E and F are polynomials in z and w. Then if $A(z,w) = 0$ and $B(z,w) = 0$, it follows that $R(z) = 0$. Since R has only finitely many zeros, this proves the lemma.

For our purposes the most important applications of this lemma occur when the polynomial A is _irreducible_. Then A and B have no common nontrivial factor except perhaps A itself. Thus, if A is not a factor of B, Lemma 3 is in force. The most important example is the case in which the degree of B with respect to w is lower than that of A.

DEFINITION 1. Let A be a polynomial,

$$A(z,w) = a_0(z)w^n + a_1(z)w^{n-1}+\ldots+a_n(z),\ a_0 \not\equiv 0.$$

Then a point $z \in \hat{C}$ is a _critical point_ for A if one of the following conditions holds:

1. $z = \infty$;
2. $a_0(z) = 0$;
3. there exists $w \in C$ such that

$$A(z,w) = 0,$$
$$\frac{\partial A}{\partial w}(z,w) = 0.$$

If z is not critical, then z is a _regular point_ for A.

PROPOSITION 1. _If A is irreducible. then there are only finitely many critical points for A._

Proof: Since a_0 has only finitely many zeros, there are only finitely many z satisfying 1 or 2. Since the degree of $\frac{\partial A}{\partial w}$ with respect to w is less than n, A and $\frac{\partial A}{\partial w}$ have no nontrivial factor in common, and

Lemma 3 implies that at most finitely many z satisfy condition 3.

QED

Of course, what we are aiming for is an analytic description of the solutions of $A(z,w) = 0$. If we wish to do this in a neighborhood of a regular point z_o, the corollary to Lemma 1 contains all the information we need, namely that there are n distinct holomorphic solutions $f_1,\ldots,f_n$ near z_o: $A(z,f_k(z)) \equiv 0$. Viewed as points in $\overline{M}$, we have found

$$e_k = e(z_o+t,\ f_k(z_o+t))$$

such that

$$A(z_o+t,\ f_k(z_o+t)) \equiv 0,\ t \text{ near } 0.$$

Another way of expressing this relation is that

$$A(\pi(e),V(e)) \equiv 0 \text{ for } e \text{ near } e_k.$$

Likewise, for the simplest example of critical point we have

$$A(z,w) = w^n-z$$

and the element

$$e_o = e(t^n,t)$$

satisfies $A(t^n,t) \equiv 0$ near 0, or

$$A(\pi(e),\ V(e)) \equiv 0 \text{ for } e \text{ near } e_o.$$

Thus, we make the following definition.

DEFINITION 2. The Riemann surface of the polynomial $A(z,w)$ is the largest open subset of $\overline{M}$ on which $A(\pi, V) = 0$. Thus, a meromorphic element $e(P,Q)$ belongs to the Riemann surface of A if and only if $A(P(t), Q(t)) \equiv 0$ for t near 0.

This latter assertion follows from the fact that if φ is a chart defined on $U(P,Q,\Delta)$ in the canonical way indicated on p. 95, then

$$P(t_o) = \pi \circ \varphi^{-1}(t_o),$$
$$Q(t_o) = V \circ \varphi^{-1}(t_o),$$

so that

$$e(P,Q) = e(\pi \circ \varphi^{-1}, V \circ \varphi^{-1}).$$

Notation. S_A is the Riemann surface of A.

The first main result we shall obtain is that if A is irreducible and has degree n in w, then S_A is compact, connected, and π restricted to S_A takes every value in $\hat{\mathbb{C}}$ n times. First, we need a lemma on polynomials and their zeros.

LEMMA 4. If $w, \alpha_1, \ldots, \alpha_n$ are complex numbers such that

$$w^n + \alpha_1 w^{n-1} + \ldots + \alpha_{n-1} w + \alpha_n = 0,$$

then

$$|w| < |\alpha_1| + \ldots + |\alpha_n| + 1.$$

Proof: If $|w| < 1$, the result holds. If $|w| \geq 1$, then

$$|w|^n \leq |\alpha_1||w|^{n-1} + \ldots + |\alpha_{n-1}||w| + |\alpha_n|$$

$$\leq (|\alpha_1| + \ldots + |\alpha_n|)|w|^{n-1},$$

so that

$$|w| \leq |\alpha_1| + \ldots + |\alpha_n|.$$

QED

THEOREM 1. *If A is irreducible, then S_A is an analytic configuration.*

Proof: By Proposition 1, if D is the set of regular points for A, then $\hat{\mathbb{C}}-D$ is finite. We shall first prove that $S_A \cap \pi^{-1}(D)$ is connected; this assertion forms the main point of the proof. Note that

$$S_A \cap \pi^{-1}(D) \subset M.$$

For suppose $e(P,Q) \in S_A \cap \pi^{-1}(D)$, and let $z_o = P(0)$, $w_o = Q(0)$. Then $z_o \in D$ and $A(z_o,w_o) = 0$. Since z_o is a regular point, $\frac{\partial A}{\partial w}(z_o,w_o) \neq 0$. Thus, Lemma 1 implies there is a unique holomorphic f near z_o such that

$A(z,f(z)) \equiv 0$, $f(z_o) = w_o$. Since $A(P(t), Q(t)) \equiv 0$ and $P(t)$ is near z_o, $Q(t)$ near w_o for small t, we then have $Q(t) = f(P(t))$. If the mapping $t \to (P(t), Q(t))$ is to be one-to-one (as it must), then the mapping $t \to P(t)$ must be one-to-one, showing that P has multiplicity 1 at 0. Thus, $e(P,Q) \in M$.

So we must now prove that if z_0 and $z_1 \in D$ and $[f]_{z_0}$ and $[g]_{z_1} \in S_A$, then there is a path in $S_A \cap \pi^{-1}(D)$ connecting these two germs. Since $\hat{C}-D$ is finite, D is connected, and thus there is a path γ in D with initial point z_1 and terminal point z_0. By Lemma 2 there exists a (unique) path $\tilde{\gamma}$ in M such that $\pi \circ \tilde{\gamma} = \gamma$ and $\tilde{\gamma}(0) = [g]_{z_1}$. By the permanence of functional relations, $\tilde{\gamma}(t)$ is in S_A for every t. In particular, $\tilde{\gamma}(1) \in S_A$ and thus is represented by a holomorphic function near z_0 which forms zeros of A. By the corollary to Lemma 1, there are unique holomorphic functions $f_1,\ldots,f_n$ in a neighborhood of z_0 such that $[f_k]_{z_0} \in S_A$ and $f_1(z),\ldots,f_n(z)$ are the distinct zeros of the function $A(z,w)$, if z is near z_0. Thus, $[f]_{z_0} = [f_j]_{z_0}$ and $\tilde{\gamma}(1) = [f_k]_{z_0}$ for some j and k. To finish the proof that $S_A \cap \pi^{-1}(D)$ is connected, it suffices to prove that for any j and k there exists a path γ in D from z_0 to z_0 such that analytic continuation of f_j along γ leads to f_k. Let us suppose that in all such analytic continuations f_1 can be analytically continued to $f_1, f_2, \ldots, f_m$, but not to $f_{m+1}, \ldots, f_n$ (where we have renumbered the f_j's). Here $1 \leq m \leq n$, and we want to prove $m = n$.

Now consider the function

$$B(z,w) = \prod_{k=1}^{m} (w - f_k(z))$$

defined for all $w \in \mathbb{C}$ and for z in a neighborhood of z_0. For each fixed w the function $B(z,w)$ can be analytically continued along all paths in D with initial point z_0 (Lemma 2), and analytic continuation along a closed path of this nature must simply lead to a permutation of $f_1, \ldots, f_m$: such a continuation could not lead to any of the $f_{m+1}, \ldots, f_n$, and two different f_j's could not be continued to the same f_k, by the unique lifting theorem. Therefore, $B(z,w)$ is analytically continued into itself along any closed path in D from z_0 to z_0, since B is a symmetric function of $f_1, \ldots, f_m$. Another way of looking at this is to perform the indicated multiplication in B and write near z_0

$$B(z,w) = w^m + \alpha_1(z) w^{m-1} + \ldots + \alpha_m(z),$$

where

$$\alpha_k(z) = (-1)^k \sum_{i_1 < i_2 < \ldots < i_k} f_{i_1} f_{i_2} \ldots f_{i_k}.$$

By the same reasoning, each α_k is symmetric in $f_1, \ldots, f_m$, so α_k has the property that it can be analytically continued along all paths in D and analytic continuation along closed paths leads back to α_k. Thus each α_k can be extended to a <u>single-valued</u> holomorphic function in D.

Now for a trick that will be used over and over. The function α_k is holomorphic in $\hat{\mathbb{C}}$ except at finitely many points. We shall now estimate the growth of α_k at these points to conclude α_k does not possess any essential singularity. Suppose now that a is one of the critical points(one of the points in $\hat{\mathbb{C}}$-D). Then for some sufficiently large integer N we have near a

$$|a_0(z)| \geq |z-a|^N, \quad |a_k(z)| \leq C \quad (1 \leq k \leq n)$$

(C is some constant) if $a \neq \infty$; if $a = \infty$ we have near a

$$|a_0(z)| \geq c, \quad |a_k(z)| \leq |z|^N \quad (1 \leq k \leq n)$$

(c is some positive constant). Thus, for z near a and $A(z,w) = 0$, Lemma 4 implies:

$$\text{if } a \neq \infty, \ |w| < nC|z-a|^{-N} + 1,$$

$$\text{if } a = \infty, \ |w| < \frac{n}{c}|z|^N + 1.$$

Since $A(z,f_k(z)) = 0$, we thus obtain for z near a,

$$|f_k(z)| \leq \text{const}|z-a|^{-N} \text{ or const } |z|^N$$

if $a \neq \infty$ or $a = \infty$, respectively. Thus, the formula for α_k shows that for z near a,

$$|\alpha_k(z)| \le \text{const } |z-a|^{-Nk} \quad \text{or const } |z|^{Nk}$$

in the two cases. Thus, α_k has either a pole or a removable singularity at a. Since this is true at each critical point, α_k is meromorphic in $\hat{C}$ and is thus a rational function.

Let b_0 be the least common multiple of all the denominators of the α_k's expressed as fractions without common factors, and let

$$\begin{aligned}\tilde{B}(z,w) &= b_0(z)B(z,w) \\ &= b_0(z)w^m + b_1(z)w^{m-1}+\ldots+ b_m(z),\end{aligned}$$

a polynomial in z,w of degree m in w. Since for z near z_0

$$\tilde{B}(z,f_1(z)) = A(z,f_1(z)) = 0,$$

the conclusion of Lemma 3 does not hold for the polynomials A and $\tilde{B}$. Thus, A and $\tilde{B}$ possess a common non-trivial factor. Since A is irreducible, this factor must be A itself. Thus, the degree of $\tilde{B}$ must be at least the degree of A, so m = n.

We have now completed the proof that $S_A \cap \pi^{-1}(D)$ is connected. The rest is easy. Suppose $e(P,Q) \in S_A$. Then for a sufficiently small disk Δ centered at 0, $U(P,Q,\Delta)$ consists only of points in $S_A \cap \pi^{-1}(D)$ with the possible exception of e(P,Q), since $\hat{C}$-D is finite. Thus, e(P,Q) can be joined to a point in $S_A \cap \pi^{-1}(D)$ by a path in $\overline{M}$. Thus, S_A is connected.

To prove S_A is a component we show it is both open and closed in $\overline{M}$. It is trivially open by Definition 2. Suppose $e(P,Q)$ is in the closure of S_A and let $\varphi\colon U(P,Q,\Delta) \to \Delta$ be a canonical chart. Then there exists $t_o \in \Delta$ such that $\varphi^{-1}(t_o) \in S_A$. Thus, since $\varphi^{-1}(t_o) = e(P(t_o+t), Q(t_o+t))$, we have

$$A(P(t_o+t),Q(t_o+t)) \equiv 0 \text{ for } t \text{ small.}$$

Thus, since $A(P(t),Q(t))$ is a meromorphic function for $t \in \Delta$ which vanishes for t near t_o, $A(P(t),Q(t)) \equiv 0$ in Δ. That is, $e(P,Q) \in S_A$, proving S_A is closed.

<u>QED</u>

<u>WARNING</u> It is tempting to think that if z_o is a critical point of the type 3, that is, if the equation $A(z_o,w) = 0$ has a double root; and if $e(P,Q) \in S_A$, $P(0) = z_o$, and $Q(0)$ is a double zero, then $e(P,Q)$ is a branch point of order at least 1. This is not true in general. For example, let

$$A(z,w) = w^2 - z^2 - z^3.$$

Then A is irreducible and $z = 0$ is a critical point, the zeros of $A(0,w) = w^2$ both vanishing. There are two points in S_A lying near $z = 0$, and these are given by

$$e(t, t\sqrt{1+t}), \qquad e(t, -t\sqrt{1+t}),$$

where $\sqrt{1+t}$ is the principal determination of the square root for t small. Clearly, neither of those meromorphic elements is a branch point of order ≥ 1.

THEOREM 2. S_A is compact.

Proof: We again write

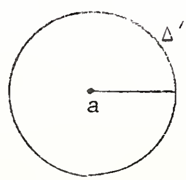
$$A(z,w) = a_0(z)w^n + \ldots + a_n(z).$$

Consider the function $\pi\colon S_A \to \hat{C}$. By Proposition 9.2 of Chapter II, it suffices to prove that the restriction of π to S_A takes every value in $\hat{C}$ n times. Of course, it suffices to consider the case in which A is irreducible. Let D be the set of regular points for A; by Proposition 1 the set $\hat{C}$-D is finite. Let $a \in \hat{C}$ and choose $\epsilon > 0$ sufficiently small that

$$\Delta = \{z\colon |z-a| < \epsilon\} \quad (\Delta = \{z\colon |z| > \epsilon^{-1}\} \text{ if } a = \infty)$$

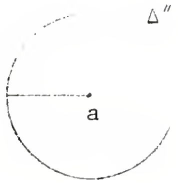
contains only points of D except possibly for a itself. Let Δ' be the set Δ with a line from a to the circumference removed; for definiteness, let

$$\Delta' = \{z\colon z \in \Delta,\ z-a \text{ not a nonnegative real number}\}.$$



Since Δ' is simply connected and contains no critical points for A, the corollary to

Lemma 2 implies that there are functions $f_1, \ldots, f_n$ holomorphic in Δ' such that for each $z \in \Delta'$, $f_1(z), \ldots, f_n(z)$ are the distinct solutions of $A(z,w) = 0$. Likewise, there are functions $g_1, \ldots, g_n$ holomorphic in the region Δ'' as illustrated:



Now just <u>below</u> the slit in Δ' the function f_k must coincide with a unique g_j. In turn, g_j must coincide with a unique f_ℓ just <u>above</u> the slit in Δ'. Let us denote $\ell = \sigma(k)$. Thus, $f_{\sigma(k)}$ is the result of analytically continuing f_k in a counterclockwise manner around Δ'. By the unique lifting theorem, the function σ is a permutation of the integers $1, 2, \ldots, n$. This permutation has a unique decomposition into cycles. Let us consider a cycle of length m and let us renumber the functions f_k so that this cycle is represented by $\sigma(1) = 2$, $\sigma(2) = 3, \ldots, \sigma(m-1) = m$, $\sigma(m) = 1$. Define for small t

$$Q(t) = \begin{cases} f_1(a+t^m), & 0 < \arg t < \frac{2\pi}{m}, \\ f_2(a+t^m), & \frac{2\pi}{m} < \arg t < 2\,\frac{2\pi}{m}, \\ \quad\vdots & \\ f_m(a+t^m), & (m-1)\frac{2\pi}{m} < \arg t < 2\pi. \end{cases}$$

(If $a = \infty$ replace $a + t^m$ by t^{-m} throughout.) By the definition of σ and the particluar enumeration of this cycle, Q has an obvious extension to a holomorphic function defined for $0 < |t| < \varepsilon^{1/m}$. Also, since each $f_k(a + t^m)$ is a solution of $A(a + t^m, w) = 0$, Lemma 4 can be applied exactly as on p. 130 to show that $|f_k(a + t^m)| \le \text{const } |t|^{-N}$ as $t \to 0$, for some positive integer N. Therefore, Q cannot have an essential singularity at 0, and thus Q is meromorphic for $|t| < \varepsilon^{1/m}$.

Now we prove that $(a+t^m, Q(t))$ is a pair. Suppose that for small s and t, $a + t^m = a + s^m$, $Q(t) = Q(s)$. Then $t^m = s^m$. If $(k-1)\frac{2\pi}{m} \le \arg t < k\frac{2\pi}{m}$ and $(j-1)\frac{2\pi}{m} \le \arg s < j\frac{2\pi}{m}$, then

$$Q(t) = f_k(a+t^m) \quad (\equiv \lim_{\delta \to 0+} f_k(a + t^m e^{i\delta})),$$

$$Q(s) = f_j(a+s^m).$$

Since $t^m = s^m$, $f_k(a+t^m) = f_j(a+t^m)$. Since the functions $f_1, \dots, f_n$ represent distinct solutions (and likewise $g_1, \dots, g_n$), we must have $j = k$. By the inequalities for $\arg t$ and $\arg s$, the equation $t^m = s^m$ now implies $t = s$. Thus, the mapping $t \to (a+t^m, Q(t))$ is one-to-one.

Since $A(a+t^m, Q(t)) \equiv 0$, this argument finds an element

$$e(a+t^m, Q(t))$$

belonging to S_A.

If the permutation σ is decomposed into cycles of lengths $m_1,\ldots,m_\ell$, then by the same argument we produce elements in S_A of the forms

$$e(a + t^{m_1}, Q_1(t)),$$
$$\vdots$$
$$e(a + t^{m_\ell}, Q_\ell(t)).$$

Since the multiplicity of π at each point $e(a+t^{m_i}, Q_i(t))$ is m_i (Proposition 5 of Chapter IV), it follows that π takes the value a <u>at least</u> $m_1+\ldots+m_\ell = n$ times. The same is true if $a = \infty$, though we of course need to use slightly different notation.

An obvious remark shows that π takes each value <u>at most</u> n times... of course, we speak of the restriction of π to S_A. In fact, if a is a regular point for A, then the points in $S_A \cap \pi^{-1}(\{a\})$ are in M (cf. p. 128) and these elements are exactly the germs of the n holomorphic solutions near a by the corollary to Lemma 1. Thus, π takes on the value a exactly n times in S_A. By the argument on p. 43, if π takes some value (a critical value) more than n times in S_A, then π takes every neighborhing value more than n times in S_A, which implies π takes some regular value more than n times in S_A, a contradiction.

<u>QED</u>

Remarks. 1. One sees finally the reason for discussing $\overline{M}$ - it contains precisely enough points to discuss branch points in general, and in particular to discuss all the solutions of algebraic equations. If $e(P,Q) \in S_A$ and $\varphi: U(P,Q,\Delta) \to \Delta$ is the canonical chart, the function φ^{-1} is called a uniformizer for A near the point P(0). It replaces the multiple-valued solutions of A = 0 by two single-valued meromorphic functions. It is of course only defined locally.

2. The elements $e(a + t^{m_i}, Q_i(t))$ produced in the above proof are obviously different. The only possibility for two of them to coincide is for two of the multiplicities m_i and m_j to coincide and for $Q_i(t) \equiv Q_j(\omega t)$ for some root of unity ω. But this would force the corresponding cycles to overlap, as can be easily checked.

3. The function Q on p. 134 is meromorphic and thus has a Laurent expansion:

$$Q(t) = \sum_{k=N}^{\infty} a_k t^k.$$

Substituting formally $z = a + t^m$, or $t = (z-a)^{1/m}$, gives a series

$$\sum_{k=N}^{\infty} a_k (z-a)^{k/m},$$

with a similar series

$$\sum_{k=N}^{\infty} a_k z^{-k/m}$$

in case $a = \infty$. These are called Puiseaux series, and have the property that for any determination of $z^{1/m}$ the sum of the series gives a solution of $A(z,w) = 0$, and differeint determinations of $z^{1/m}$ yield different solutions. Of course, all this information is contained in the idea of the corresponding meromorphic element.

4. It is almost amazing how easy it was to find the elements $e(a+t^{m_i}, Q_i(t))$ in S_A. However, when one observes what had to be known, it is quite obvious that it should be easy. Namely, we had to have completely solved the equation $A(z,w) = 0$ away from the critical points, and then it was a simple matter of checking what S_A looks like above these finitely many critical points. But this sort of procedure can almost never be carried out in practice for rather obvious reasons. We can't even usually hope to solve the equation near a critical point and observe how the zeros behave under analytic continuation around the critical point.

5. Even without knowing Proposition 9.2 of Chapter II, it is almost obvious why S_A is compact. For S_A consists essentially of n copies of the (compact) sphere $\hat{\mathbb{C}}$ branched above certain finitely many points. The only way S_A could fail to be compact would be for certain of these branch points not to be included in S_A. Essentially the

proof shows they are indeed all included and this statement is phrased in the perhaps deceptive statement that the restriction of π to S_A takes every value n times.

Problem 6. Let $A(z,w) = w^3-3zw+z^3$. Prove that A is irreducible. Find its critical points and discover the types of meromorphic elements which belong to S_A. Compute the genus of S_A by the Riemann-Hurwitz formula (p. 112).

Problem 7. Same for $A(z,w) = zw^3-3w+2z^a$, where a is any integer (positive or negative). Of course, if $a < 0$ then this is interpreted to be the problem for the polynomial

$$z^{1-a}w^3 - 3z^{-a}w + 2.$$

Now we pass to the converse of Theorem 2. This states that every compact analytic configuration is the Riemann surface of a unique (to within a constant factor) irreducible algebraic function. In Chapter VI this statement will be improved considerably and will state that any compact connected Riemann surface is analytically equivalent to a compact analytic configuration (and thus has an associated irreducible polynomial).

Before stating this converse of Theorem 2, we make a useful observation about S_A. First, divide out the leading coefficient $a_0(z)$ to write $A(z,w) = a_0(z)A'(z,w)$, where

$$A'(z,w) = w^n + \alpha_1(z)w^{n-1}+\ldots+\alpha_n(z)$$

and $\alpha_1,\ldots,\alpha_n$ are rational functions of z. We assume A (and thus A′) to be irreducible. If a is a regular point for A, then the corollary to Lemma 2 shows the existence of the holomorphic zeros $f_1,\ldots,f_n$ as usual. Thus,

$$e_k = e(a+t,\ f_k(a+t))$$

is a point in S_A, $1 \leq k \leq n$, and the elements e_k are the only ones in $S_A \cap \pi^{-1}(\{a\})$. Also, $V(e_k) = f_k(a)$, so the numbers $V(e_k)$ are the n solutions of $A'(a,w)=0$. Thus, we obtain a factorization

$$\begin{aligned} A'(a,w) &= \prod_{k=1}^{n} (w-V(e_k)) \\ &= \prod_{\substack{e \in S_A \\ \pi(e)=a}} (w-V(e)). \end{aligned}$$

<u>THEOREM 3</u>. <u>Let</u> S <u>be a compact analytic configuration</u>. <u>Then there exists a unique (up to constant factor) irreducible polynomial</u> A <u>such that</u> $S = S_A$.

<u>Proof</u>: Since S is compact and $\pi\colon S \to \hat{\mathbb{C}}$ is analytic, Proposition 9.1 of Chapter II shows that the restriction of π to S takes every value the same number n of times. Let D be the subset of $\hat{\mathbb{C}}$ defined by

$$\hat{\mathbb{C}}-D = \{\pi(e)\colon e \in S,\ m_\pi(e) > 1 \text{ or } V(e) = \infty\}.$$

Thus, if $e \in S$ and $\pi(e) \in D$, then $m_\pi(e) = 1$ and $V(e) \neq \infty$. Since S is compact, the set of elements e such that $e \in S$ and $m_\pi(e) > 1$ or $V(e) = \infty$ is <u>finite</u>. <u>A fortiori</u>, $\hat{C}-D$ is finite. We now take our clue from the discussion on p. 140 and define for $z \in D$, $w \in C$

$$A'(z,w) = \prod_{\substack{e \in S \\ \pi(e)=z}} (w-V(e)).$$

That discussion implies that <u>if</u> $S = S_A$ for some A, then this must be the formula for $A'(z,w)$ for regular points z, since all the regular points must be contained in D. Thus, the uniqueness assertion of the theorem is established. Moreover, we have an <u>explicit</u> formula for A' and we now just have to check various details.

First, if $z_o \in D$ then there are exactly n elements $e_1,\dots,e_n \in S$ with $\pi(e_k) = z_o$, since π takes the value z_o n times and π must have multiplicity 1 at each e_k. Suppose $e_k = e(P(t),Q_k(t))$, where $P(t) = z_o+t$ if $z_o \neq \infty$ and $P(t) = t^{-1}$ if $z_o = \infty$. Let $\varphi_k\colon U(P,Q_k,\Delta) \to \Delta$ be a canonical chart. For small t_o,

$$S \cap \pi^{-1}(\{P(t_o)\}) = \{\varphi_1^{-1}(t_o),\dots,\varphi_n^{-1}(t_o)\},$$

so that

$$A'(P(t_o),w) = \prod_{k=1}^{n} (w-Q_k(t_o))$$

since $V(\varphi_k^{-1}(t_o)) = Q_k(t_o)$. This equation shows that if we expand

$$A'(z,w) = w^n + \alpha_1(z)w^{n-1} + \ldots + \alpha_n(z),$$

then $\alpha_1, \ldots, \alpha_n$ are holomorphic on D.

Now we examine the behavior of α_k at the (isolated) points of $\hat{C}$-D. Suppose $a \in \hat{C}$-D. Let $e(a+t^m, \tilde{Q}(t))$ be one of the points in $S \cap \pi^{-1}(\{a\})$; in case $a = \infty$ this must be replaced by $e(t^{-m}, \tilde{Q}(t))$. Then for z near a but not equal to a, there are m points in $S \cap \pi^{-1}(\{z\})$ determined by this one element, namely,

$$e(a+(t_k + t)^m, \tilde{Q}(t_k + t)), \text{ where } t_k^m = z-a.$$

(We now discuss the case $a \neq \infty$; the case $a = \infty$ is handled entirely similarly.) The corresponding values of V(e) are $\tilde{Q}(t_k)$, $1 \le k \le m$. Thus, for some N we have

$$|V(e)| \le |t_k|^{-N} = |z-a|^{-N/m}$$

for these m points $e \in S \cap \pi^{-1}(z)$. Treating the other points in $S \cap \pi^{-1}(z)$ in a similar fashion, we obtain for some integer M

$$|V(e)| \le |z-a|^{-M} \text{ if } e \in S \cap \pi^{-1}(z), \text{ z near a}.$$

Thus

$$|\alpha_k(z)| \le \text{const } |z-a|^{-Mk} \text{ if z is near a;}$$

if $a = \infty$ this estimate should read

$$|\alpha_k(z)| \le \text{const } |z|^{Mk}.$$

Therefore, α_k is meromorphic on $\hat{\mathbb{C}}$ and thus α_k is <u>rational</u>.

Now that we have produced a polynomial A we must show that A is irreducible and that its Riemann surface is S. This will essentially be done all at once. Suppose that there exists a factorization of A in the form $A = BC$, where B and C are polynomials and B is irreducible...in fact, there is always such a factorization with a polynomial B of degree at least one in w (perhaps C is constant). Then B has a Riemann surface S_B which is a compact analytic configuration. Let e be an element in S_B such that $m_\pi(e)=1$ and $V(e) \neq \infty$; this includes all but finitely many points in S_B. We also assume $\pi(e) \in D$, eliminating again at most finitely many points. Then if $\pi(e) = z_o$ we let $P(t) = z_o + t$ if $z_o \neq \infty$ and $P(t) = t^{-1}$ if $z_o = \infty$. Then $e = e(P,Q)$, and $B(P(t), Q(t)) \equiv 0$ for t near 0. Thus, since $A = BC$ we have

$$A'(P(t),Q(t)) \equiv 0 \text{ for } t \text{ near } 0.$$

The formula for A' at the bottom of p. 141 implies

$$\prod_{k=1}^{n}(Q(t)-Q_k(t)) \equiv 0 \quad \text{for } t \text{ near } 0.$$

Since each factor $Q-Q_k$ which is not identically zero can have only isolated zeros, it follows that for some k,

$$Q = Q_k.$$

Thus, $e = e(P,Q_k) = e_k \in S$. Thus, except for finitely many points $S_B \subset S$. Since S is compact in the Hausdorff space $\overline{M}$, S is closed and thus

$$S_B \subset S.$$

Since S_B is a component of $\overline{M}$ and since S is connected, it follows that $\underline{S_B = S.}$

Now it is all done. For, π assumes (when restricted to S_B) every value the same number of times, this number being the degree of B as a polynomial in w. But π assumes (when restricted to S) every value n times. Thus, B has degree n in w. Thus, C has degree 0 in w and thus is just a polynomial in z. Therefore, if we discard all the common polynomial factors in z from the polynomial A(z,w), we must have $C \equiv$ const. This shows that A is irreducible, its only possible nontrivial factor turning out to be itself. And $S_A = S_B = S$.

<u>QED</u>

We thus see that on any compact analytic configuration S the two meromorphic functions π and V are related by an algebraic equation. These two functions of course allow us to construct other meromorphic functions on S; namely any rational function of π and V is meromorphic on S. The amazing fact is that there are no other meromorphic functions on S. In fact, we have

THEOREM 4. Let S be a compact analytic configuration on which π assumes every value n times. Let f be any meromorphic function on S. Then there exist unique rational functions $\alpha_0, \ldots, \alpha_{n-1}$ such that

$$f = \sum_{j=0}^{n-1} \alpha_j \circ \pi \cdot V^j .$$

Proof: Suppose z is a regular point for A, the polynomial is such that A is irreducible, and $S = S_A$. If the the formula for f is to hold, then we must have

$$f(e) = \sum_{j=0}^{n-1} \alpha_j(z) V(e)^j \quad \text{if } e \in S,\ \pi(e) = z .$$

Now $S \cap \pi^{-1}(\{z\}) = \{e_1, \ldots, e_n\}$ has exactly n points and the numbers $V(e_k)$ are distinct. The above equations read

$$f(e_k) = \sum_{j=0}^{n-1} \alpha_j(z)\ V(e_k)^j, \qquad 1 \leq k \leq n .$$

These are n equations in n "unknowns", $\alpha_o(z), \ldots, \alpha_{n-1}(z)$, and the determinant of the system is

$$\det \begin{pmatrix} 1 & V(e_1) & V(e_1)^2 & \ldots & V(e_1)^{n-1} \\ 1 & V(e_2) & V(e_2)^2 & \ldots & V(e_2)^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & V(e_n) & V(e_n)^2 & \ldots & V(e_n)^{n-1} \end{pmatrix} .$$

This is a so-called Vandermonde determinant and its value is well known and easily seen to be

$$\prod_{1 \le \ell < k \le n} (V(e_k)-V(e_\ell)),$$

which is not zero. Thus, $\alpha_0(z),\ldots,\alpha_{n-1}(z)$ are uniquely determined. It is also clear that these numbers $\alpha_j(z)$ really depend only on z and not on a particular ordering $e_1,\ldots,e_n$ of the points in $\pi^{-1}(\{z\})$. Thus, $\alpha_0,\ldots,\alpha_{n-1}$ are uniquely determined at the regular points for A, and thus are unique since they are to be rational functions.

Knowing what α_j must be, we now prove that they exist. By Cramer's rule, we can write down a formula for $\alpha_j(z)$ in terms of a determinant involving $f(e_k)$ and $V(e_k)$, divided by the Vandermonde determinant. Near a fixed regular point we can choose the e_k in terms of charts to be analtyic functions and thus $f(e_k)$ and $V(e_k)$ become analytic functions, proving α_j is holomorphic on the set of regular points. As usual, we now prove that α_j cannot have any essential singularities. Since we obtain upper bounds for $f(e_k)$ and $V(e_k)$ in the standard manner we are used to by now, it remains to obtain a lower bound for the Vandermonde.

Suppose then that a is a critical point. We assume in the following that $a \ne \infty$; the case $a = \infty$ is treated by mere formal changes in the analysis. The points in $S \cap \pi^{-1}(\{a\})$ have the forms

$$e(a+t^{m_j},\ Q_j(t)),\ 1 \le j \le J,\ \sum_{j=1}^{J} m_j = n,$$

where Q_j is meromorphic near 0. Let the positive integer m be the least common multiple of the integers m_j. If z is a number sufficiently near a but not equal to a, choose an arbitrary $s \in \mathbb{C}$ such that

$$z-a = s^m.$$

Let

$$\omega_j = e^{2\pi i/m_j}.$$

Then for $0 \le \ell \le m_j-1$,

$$z-a = (\omega_j^{\ell}\, s^{m/m_j})^{m_j},$$

and the numbers $\omega_j^{\ell} s^{m/m_j}$ are different for $0 \le \ell \le m_j-1$. Thus, $S \cap \pi^{-1}(\{z\})$ consists of the points

$$e_{j\ell} = e(a+(\omega_j^{\ell}\, s^{m/m_j} + t)^{m_j},\ \ Q_j(\omega_j^{\ell}\, s^{m/m_j} + t))$$

for $0 \le \ell \le m_j-1$, $1 \le j \le J$. Thus, $V(e_{j\ell}) = Q_j(\omega_j^{\ell} s^{m/m_j})$. The Vandermonde contains terms of the form

$$V(e_{j\ell})-V(e_{j'\ell'}) = Q_j(\omega_j^{\ell} s^{m/m_j}) - Q_{j'}(\omega_{j'}^{\ell'} s^{m/m_{j'}}),$$

which is a <u>meromorphic function</u> of s, not vanishing for small $s \ne 0$ since z is regular for a. Thus, there exists an integer N such that

$$|V(e_{j\ell}) - V(e_{j'\ell'})| \geq |s|^N = |z-a|^{N/m},$$

so the Vandermonde has modulus bounded below by

$$|z-a|^{\frac{N}{m}\frac{n(n-1)}{2}} .$$

Thus, we have proved that each α_j is rational and by definition

$$f(e) = \sum_{j=0}^{n-1} \alpha_j(\pi(e))V(e)^j$$

for all but finitely many $e \in S$ (those such that $\pi(e)$ is a critical point for A). Thus, these two meromorphic functions on S coincide.

<u>QED</u>

Chapter VI

EXISTENCE OF MEROMORPHIC FUNCTIONS

The thrust of this chapter is the proof that there exist nonconstant meromorphic functions on <u>any</u> Riemann surface. It will take a tremendous amount of machinery to achieve this result; in particular, we will need to give a careful and fairly complete discussion of harmonic functions on Riemann surfaces. But before beginning this topic, we shall exhibit one problem which can be solved using the existence of meromorphic functions.

First, we introduce a lemma which really logically belongs in Chapter IV, but has not been needed before now.

<u>LEMMA 1</u>. *Let* m *be a positive integer and* Q *a meromorphic function near* 0 *having Laurent expansion*

$$Q(s) = \sum_{j=-\infty}^{\infty} \alpha_j s^j,$$

and assume that no positive integer except 1 *is a common factor of all* j *such that* $\alpha_j \neq 0$. *Let* n *be an integer relatively prime to* m. *Then* $(t^m, Q(t^n))$ *is a pair*.

<u>Proof</u>: We have to prove that the mapping $t \to (t^m, Q(t^n))$ is one-to-one near 0. If this is not the case, then there exist $s_k \to 0$ and $t_k \to 0$ such that $s_k \neq t_k$ and $s_k^m = t_k^m$,

$Q(s_k^n) = Q(t_k^n)$. Thus, $(s_k/t_k)^m = 1$, and by taking a subsequence we can assume that there exists a fixed ω such that $\omega \neq 1$, $\omega^m = 1$, $s_k = \omega t_k$ (cf. p. 93). Therefore, $Q(\omega^n t_k^n) = Q(t_k^n)$, so that the two functions $Q(\omega^n s)$ and $Q(s)$ agree on a sequence $s = t_k^n \to 0$. Since they are both meromorphic, they must be identical:

$$\sum_{j=-\infty}^{\infty} \alpha_j \omega^{nj} s^j \equiv \sum_{j=-\infty}^{\infty} \alpha_j s^j, \quad s \text{ near } 0.$$

Therefore the coefficients must agree: $\alpha_j \omega^{nj} = \alpha_j$ for all j. This means that $\alpha_j \neq 0$ implies $\omega^{nj} = 1$. It follows easily that $\omega^n = 1$. For the set $\{j: \omega^{nj} = 1\}$ is an additive subgroup of the integers and the Euclidean algorithm implies that any subgroup equals the integer multiples of a fixed positive integer j_o. Thus, $\alpha_j \neq 0$ implies j contains j_o as a factor. By hypothesis, $j_o = 1$ and therefore $\omega^n = 1$. Since n and m are relatively prime, the Euclidean algorithm again implies there exist integers p and q such that $pm + qn = 1$. Thus,

$$\omega = \omega^{pm+qn} = (\omega^m)^p(\omega^n)^q = 1,$$

a contradiction.

<u>QED</u>

<u>THEOREM 1</u>. <u>Let</u> S <u>be any connected Riemann surface</u>. <u>Let</u> f <u>and</u> g <u>be meromorphic functions on</u> S <u>such that</u> f $\not\equiv$ <u>constant</u>. <u>Then there exists a unique analytic</u> $\Phi: S \to \overline{M}$

such that

$$f = \pi \circ \Phi,$$

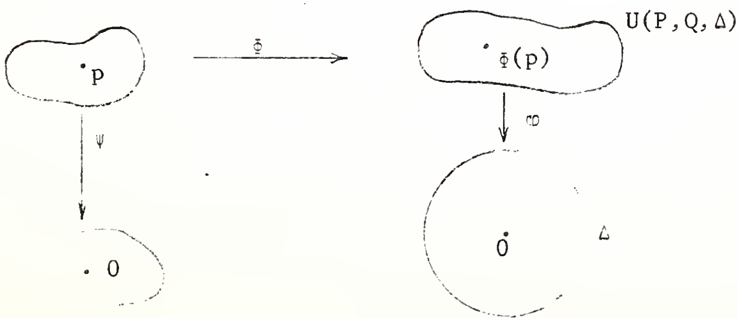
$$g = V \circ \Phi.$$

It is convenient to draw a diagram to indicate these two equations:

$$\begin{tikzcd}
S \arrow[rr, dashed, "\Phi"] \arrow[dr, dashed, "{f,g}"'] & & \overline{M} \arrow[dl, "{\pi,V}"] \\
& \hat{C} \times \hat{C} &
\end{tikzcd}$$

The statement of the theorem is then exactly that there exists an analytic Φ making this diagram commutative.

Proof: Uniqueness: Suppose $p \in S$ and that $m_f(p) = 1$. Since $f = \pi \circ \Phi$, it follows that $m_\Phi(p) = 1$. Let ψ be any compatible chart in a neighborhood of p. Suppose $\Phi(p) = e(P,Q)$ and let $\varphi: U(P,Q,\Delta) \to \Delta$ be a canonical chart. We assume $\psi(p) = 0$.



Recall from p. 126 that

$$P = \pi \circ \varphi^{-1},$$

$$Q = V \circ \varphi^{-1} .$$

Now the mapping $\varphi \circ \Phi \circ \psi^{-1}$ is a <u>parameter change</u>, since it is one-to-one near 0 and maps 0 to 0. Thus, we consider

$$P \circ (\varphi \circ \Phi \circ \psi^{-1}) = \pi \circ \Phi \circ \psi^{-1} = f \circ \psi^{-1},$$

$$Q \circ (\varphi \circ \Phi \circ \psi^{-1}) = V \circ \Phi \circ \psi^{-1} = g \circ \psi^{-1} ,$$

and we thus have

$$(P,Q) \sim (f \circ \psi^{-1}, g \circ \psi^{-1}).$$

Therefore, if $m_f(p) = 1$ we have

$$\Phi(p) = e(f \circ \psi^{-1}, g \circ \psi^{-1}).$$

This proves that Φ is uniquely determined except on the discrete set where the multiplicity of f is greater than 1. Since Φ is continuous on S, then Φ is also uniquely determined everywhere.

<u>Existence</u>: We already know how to define Φ at points where $m_f = 1$. Therefore, we so define Φ at those points, just noting that the definition $\Phi(p) = e(f \circ \psi^{-1}, g \circ \psi^{-1})$ really is independent of the particular chart ψ; a different selection of the chart merely gives a

parameter change.

Now suppose $p_o \in S$ and $m_f(p_o) = m$. Choose a chart ψ near p_o such that $\psi(p_o) = 0$ and

$$f\circ\psi^{-1}(t) = f(p_o) + t^m$$

if $f(p_o) \neq \infty$. As usual, if $f(p_o) = \infty$ we have instead

$$f\circ\psi^{-1}(t) = t^{-m}.$$

Then consider the Laurent expansion of $g\circ\psi^{-1}$:

$$g\circ\psi^{-1}(t) = \sum_{k=-\infty}^{\infty} a_k t^k.$$

Let n be the largest positive integer which is a factor of all k such that $a_k \neq 0$; in case $a_k = 0$ for all k, let $n = m$. Then we can let $k = nj$ in the above series and we obtain

$$g\circ\psi^{-1}(t) = \sum_{j=-\infty}^{\infty} a_{nj}t^{nj} = Q(t^n),$$

where

$$Q(s) = \sum_{j=-\infty}^{\infty} \alpha_j s^j$$

and $\alpha_j = a_{nj}$. Thus, either $\alpha_j = 0$ for all j or there is no common factor of all j with $\alpha_j \neq 0$ except 1 (and -1). Let μ be the positive integer which is the greatest common divisor of m and n and define

$$\Phi(p_o) = e(f(p_o) + t^{\frac{m}{\mu}}, Q(t^{\frac{n}{\mu}}))$$

(replace $f(p_o) + t^{\frac{m}{\mu}}$ by $t^{-\frac{m}{\mu}}$ if $f(p_o) = \infty$). We have to check that this is really a meromorphic element, i.e., that

$$(t^{\frac{m}{\mu}}, Q(t^{\frac{m}{\mu}}))$$

is a pair. If $\alpha_j = 0$ for all j, then $\frac{m}{\mu} = 1$ and it is obvious. Otherwise, Lemma 1 applied to the relatively prime integers $\frac{m}{\mu}$ and $\frac{n}{\mu}$ shows that we do have a pair. Thus, $\Phi(p_o)$ makes sense; we do not need to check that we have defined it independently of the choice of ψ (there are only m choices to make) since we can regard the choice of ψ to be an arbitrary "function" of p_o.

We now observe that this definition of $\Phi(p_o)$ works even when $m = 1$; then $\mu = 1$ and the definition agrees with the earlier definition of Φ at points where f has multiplicity 1. Note that obviously

$$\pi \circ \Phi(p_o) = f(p_o),$$

$$V \circ \Phi(p_o) = Q(0) = g \circ \psi^{-1}(0) = g(p_o).$$

Thus, the required commutativity of the diagram is proved. We thus need to check the analyticity of Φ in order to finish the proof. We prove that Φ is analytic in a neighborhood of p_o, using the above notation.

Let z be near 0, $z \neq 0$, and let $p = \psi^{-1}(z)$. Then for sufficiently small z, p is a point where f has multiplicity 1. Define the chart $\psi_1 = \psi - z$, so that $\psi_1(p) = 0$ and

$$\psi_1^{-1}(t) = \psi^{-1}(z+t).$$

Therefore, according to our first definition of Φ,

$$\begin{aligned}\Phi(\psi^{-1}(z)) &= e(f\circ\psi_1^{-1}, g\circ\psi_1^{-1}) \\ &= e(f\circ\psi^{-1}(z+t), g\circ\psi^{-1}(z+t)) \\ &= e(f(p_o) + (z+t)^m, Q((z+t)^n)).\end{aligned}$$

Now we introduce the canonical chart near $\Phi(p_o)$: call it φ: $U \to \Delta$, where

$$\varphi^{-1}(t_o) = e(f(p_o) + (t_o+t)^{\frac{m}{\mu}}, Q((t_o+t)^{\frac{n}{\mu}})).$$

We introduce next the parameter change ρ defined by

$$\rho(t) = (z+t)^{\mu} - z^{\mu};$$

since $z \neq 0$, this <u>is</u> a parameter change. And we have

$$(z+t)^m = (z^{\mu}+\rho(t))^{\frac{m}{\mu}}$$ (and a similar formula with n),

showing that

$$\begin{aligned}\Phi(\psi^{-1}(z)) &= e(f(p_o)+(z^{\mu}+\rho(t))^{\frac{m}{\mu}}, Q((z^{\mu}+\rho(t))^{\frac{n}{\mu}})) \\ &= e(f(p_o)+(z^{\mu}+t)^{\frac{m}{\mu}}, Q((z^{\mu}+t)^{\frac{n}{\mu}})) \\ &= \varphi^{-1}(z^{\mu}).\end{aligned}$$

Thus,

$$\omega\circ\Phi\circ\psi^{-1}(z) = z^{\mu},$$

so $\omega\circ\Phi\circ\psi^{-1}$ is holomorphic near 0. This proves that Φ is analytic near p_o.

QED

COROLLARY. *Let* S *be any compact connected Riemann surface and* f,g *meromorphic functions on* S *such that* $f \not\equiv$ *constant. Then there exist a unique compact analytic configuration* T *and analytic function* Φ *from* S *onto* T *such that the diagram commutes*:

$$\begin{tikzcd} S \arrow[rr, "\Phi"] \arrow[dr, "{f,g}"'] & & T \arrow[dl, "{\pi,V}"] \\ & \hat{\mathbb{C}} \times \hat{\mathbb{C}} & \end{tikzcd}$$

Proof: This is trivial. We just let $T = \Phi(S)$. Since S is compact and connected and Φ is continuous, T is also compact and connected. Since Φ is analytic and nonconstant, Proposition 4 of Chapter II implies T is an open subset of $\overline{M}$. As T is thus closed and open and connected, it is an analytic configuration.

QED

COROLLARY. *Under the hypothesis of the previous corollary, there exists a unique irreducible polynomial* A(z,w) *such that*

$$A(f(p), g(p)) = 0 \text{ for } p \in S.$$

Proof: We first prove uniqueness. If A has the required properties, then $A(\pi(\Phi(p)), V(\Phi(p))) = 0$ for $p \in S$. Since Φ is onto, this implies $A(\pi(e), V(e)) = 0$ for $e \in T$. Therefore the Riemann surface for A satisfies $S_A \supset T$. Since T is a component of $\overline{M}$ and S_A is connected, $S_A = T$. Thus, Theorem 3 of Chapter V shows A is unique.

Existence is trivial. Simply let A be chosen by Theorem 3 of Chapter V such that $S_A = T$. The above argument worked the other direction proves $A(f,g) = 0$.

QED

We are most interested in the possibility that the mapping Φ of S onto T is also one-to-one. For then we will have an analytic equivalence of the compact Riemann surface S with an analytic configuration. The next theorem gives some equivalent conditions.

THEOREM 2. Let S be a compact connected Riemann surface and f,g meromorphic functions on S such that $f \neq$ constant. Let Φ, T, A be the objects of the two previous corollaries. Assume that f takes every value n times. The the following conditions are equivalent.

1. Φ is an analytic equivalence of S onto T.

2. There exists a point $z \in \hat{C}$ such that $\{g(p): f(p) = z\}$ has n points.

3. For all except finitely many $z \in \hat{C}$, $\{g(p): f(p) = z\}$ has n points.

4. <u>The polynomial</u> A <u>has degree</u> n <u>in</u> w.

<u>Proof</u>: <u>$1 \Rightarrow 4$</u>: Since Φ is an analytic equivalence and $f = \pi \circ \Phi$, π also takes every value n times. The results of Chapter V, especially Theorem 3, imply that A has degree n in w.

<u>$4 \Rightarrow 3$</u>: If z is a regular point for A (and this is true for all but the finitely many critical points), then $T \cap \pi^{-1}(\{z\}) = \{e_1,\ldots,e_n\}$ and the numbers $V(e_1),\ldots,V(e_n)$ are distinct, being the solutions of $A(z,w) = 0$. Since Φ is onto, there exist $p_1,\ldots,p_n \in S$ such that $\Phi(p_k) = e_k$. Then $g(p_k) = V(e_k)$ and $f(p_k) = \pi(e_k) = z$, so

$$\{g(p): f(p) = z\}$$

has at least n points $V(e_1),\ldots,V(e_n)$. Since f takes every value n times, this set can contain no more than n points.

<u>$3 \Rightarrow 2$</u>: Trivial.

<u>$2 \Rightarrow 1$</u>: Finally we come to an interesting proof. By hypothesis there are n points $p_1,\ldots,p_n$ such that $f(p_k) = z$ and the numbers $g(p_1),\ldots,g(p_n)$ are distinct. In particular, since f takes the value z n times, $m_f(p_1) = 1$ Since $f = \pi \circ \Phi$, $m_\Phi(p_1) = 1$. Let $e = \Phi(p_1)$. We shall show that Φ takes the value e <u>one</u> time. Suppose then that $\Phi(p) = e$. Then $f(p) = \pi(e) = \pi \circ \Phi(p_1) = f(p_1) = z$, so

$p = p_k$ for some k. Then $g(p_k) = V(e) = V\circ\Phi(p_1) = g(p_1)$, so $p_k = p_1$. Thus $\Phi(p) = e$ if and only if $p = p_1$. Since moreover $m_\Phi(p_1) = 1$, we have now proved that Φ takes the value e one time. But Proposition 9.1 of Chapter II implies Φ takes <u>every</u> value one time. That is, Φ is one-to-one. Thus, Φ is an analytic equivalence.

<u>QED</u>

Let us comment on 4. Suppose that the mapping Φ takes every value k times, and that $\pi: T \to \hat{\mathbb{C}}$ takes every value m times. Then since $f = \pi\circ\Phi$, it is easy to see that f takes every value mk times. In the notation of Theorem 2, this means $n = mk$ and A has degree m in w. Thus, in general the degree of A is a factor of n. For example, consider the trivial case in which $S = \hat{\mathbb{C}}$, $f(z) = z^4$, and $g(z) = z^2$. The uniqueness assertion of the previous corollary implies

$$A(z,w) = w^2 - z.$$

For, A is irreducible and $A(f(z),g(z)) = g(z)^2 - f(z) = z^4 - z^4 = 0$. Thus, f takes every value 4 times, the degree of A is 2, so we conclude that Φ takes every value 2 times. In fact, our explicit construction shows that for $z \neq 0,\infty$,

$$\Phi(z) = e((z+t)^4,\ (z+t)^2).$$

$$\begin{aligned}\Phi(-z) &= e((-z+t)^4,\ (-z+t)^2)\\ &= e((z-t)^4,\ (z-t)^2)\end{aligned}$$

$$= e((z+t)^4, (z+t)^2)$$
$$= \Phi(z)$$

by the parameter change $t \to -t$.

Now we state the main theorem of this section and show how it can be used to produce functions f and g which satisfy the criteria of Theorem 2. Note that we must at least produce a nonconstant meromorphic function f on S; the following theorem allows us to do even better.

<u>THEOREM 3</u>. <u>Let</u> S <u>be any connected Riemann surface and let</u> $p, q \in S$, $p \neq q$. <u>Then there exists a meromorphic function</u> f <u>on</u> S <u>such that</u> $f(p) \neq f(q)$.

We are nowhere near being able to prove this yet. But assuming its validity for the moment we prove

<u>COROLLARY</u>. <u>Let</u> S <u>be a compact connected Riemann surface</u>. <u>Then</u> S <u>is analytically equivalent to an analytic configuration</u>.

<u>Proof</u>: First apply Theorem 3 to find a nonconstant meromorphic f on S. Now we show how to construct another meromorphic g on S which satisfies criterion 2 of Theorem 2. Assume that f takes every value n times.

If $n = 1$, take $g = 0$. Suppose $n > 1$. Since the points of S where the multiplicity of f is greater than 1 are isolated, there exists $z \in \hat{\mathbb{C}}$ such that $f^{-1}(\{z\})$ consists of n distinct points $p_1, \ldots, p_n$. Theorem 3 implies that if $j \neq 1$, there exists a meromorphic h on S such that $h(p_j) \neq h(p_1)$. Choose a complex number $\alpha \notin \{h(p_1), \ldots, h(p_n)\}$. Then there exists a Möbius transformation

$$F(w) = \frac{aw + b}{cw + d}$$

such that $F(h(p_1)) = 1$, $F(h(p_j)) = 0$, $F(\alpha) = \infty$. Thus, there exists a meromorphic $h_j = F \circ h$ such that

$$\begin{aligned} &h_j(p_1) = 1, \\ &h_j(p_j) = 0, \\ &h_j(p_k) \text{ is in } \mathbb{C} \text{ for } 1 \leq k \leq n. \end{aligned}$$

Define $g_1 = \prod_{j=2}^{n} h_j$. Then g_1 is meromorphic on S and

$$\begin{aligned} &g_1(p_1) = 1, \\ &g_1(p_j) = 0, \ 2 \leq j \leq n. \end{aligned}$$

Repeating this construction, there exist meromorphic functions $g_1, \ldots, g_n$ on S such that

$$\begin{aligned} &g_k(p_k) = 1, \\ &g_k(p_j) = 0 \text{ if } j \neq k. \end{aligned}$$

Now define

$$g = \sum_{k=1}^{n} k g_k.$$

Then g is meromorphic on S and $g(p_k) = k, 1 \le k \le n$. Therefore,

$$\{g(p): f(p) = z\} = \{1,2,\ldots,n\}$$

has n points. So criterion 3 of Theorem 2 is satisfied and therefore S is analytically equivalent to an analytic configuration (criterion 1 of Theorem 2).

QED

Now we shall begin to introduce the machinery needed to prove Theorem 3. The basis is the idea of harmonic functions on Riemann surfaces. First, we recall that a function u on an open set in $\mathbb{C}$ is said to be harmonic if u is of class C^2 and $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$. A convenient way of discussing this is to define the differential operators

$$\partial u \equiv \frac{1}{2}\frac{\partial u}{\partial x} + \frac{1}{2i}\frac{\partial u}{\partial y},$$
$$\bar{\partial} u \equiv \frac{1}{2}\frac{\partial u}{\partial x} - \frac{1}{2i}\frac{\partial u}{\partial y}.$$

Then

$$\partial\bar{\partial} u = \bar{\partial}\partial u = \tfrac{1}{4}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right).$$

Now the equation $\bar{\partial} f = 0$ is exactly the Cauchy-Riemann equation. Thus, f is holomorphic if and only if $\bar{\partial} f = 0$; moreover, in this case $\partial f = f'$, the ordinary complex derivative of f. Thus, if u is of class C^2, then u is harmonic if and only if ∂u is holomorphic. In particular, if u is harmonic then ∂u has derivatives of all orders. Likewise, $\bar{u}$ is harmonic so that $\bar{\partial} u = \overline{\partial \bar{u}}$ has derivatives of all orders. Thus, $\frac{\partial u}{\partial x} = \partial u + \bar{\partial} u$ has derivatives of all orders, and the same is true for $\frac{\partial u}{\partial y}$. Thus, harmonic functions have derivatives of all orders.

Chain rule. There is a chain rule for these differential operators, which we now describe. Suppose V and W are open sets in $\mathbb{C}$ and $h: V \to W$ is a class C^1 mapping of V into W.

Let $u: W \to \mathbb{C}$ be of class C^1. Then $u\circ h$ is also of class C^1 and if we let D_1 denote partial differentiation with respect to the first argument and D_2 with respect to the second argument, the usual chain rule reads

$$D_1(u\circ h) = (D_1 u)\circ h\, D_1(\mathrm{Re} h) + (D_2 u)\circ h\, D_1(\mathrm{Im} h).$$
$$D_2(u\circ h) = (D_1 u)\circ h\, D_2(\mathrm{Re} h) + (D_2 u)\circ h\, D_2(\mathrm{Im} h).$$

Therefore,

$$\begin{aligned}
\partial(u\circ h) &= (D_1 u)\circ h\ \partial(\mathrm{Re} h) + (D_2 u)\circ h\ \partial(\mathrm{Im} h) \\
&= (D_1 u)\circ h\ \frac{\partial h + \partial \bar{h}}{2} + (D_2 u)\circ h\ \frac{\partial h - \partial \bar{h}}{2i} \\
&= (\tfrac{1}{2} D_1 u + \tfrac{1}{2i} D_2 u)\circ h\ \partial h + (\tfrac{1}{2} D_1 u - \tfrac{1}{2i} D_2 u)\circ h\, \partial \bar{h} \\
&= (\partial u)\circ h\ \partial h + (\bar{\partial} u)\circ h\ \partial \bar{h}.
\end{aligned}$$

The corresponding formula for $\bar{\partial}(u\circ h)$ follows the same way. We thus obtain

$$\partial(u\circ h) = (\partial h)\circ h\ \partial h + (\bar{\partial}u)\circ h\ \partial\bar{h},$$
$$\bar{\partial}(u\circ h) = (\partial u)\circ h\ \bar{\partial}h + (\bar{\partial}u)\circ h\ \bar{\partial}\bar{h}.$$

As a special case, suppose h is holomorphic. Then $\partial h = h'$, $\bar{\partial}h = 0$, so we obtain

$$\begin{aligned} \partial(u\circ h) &= (\partial u)\circ h\ h', \\ \bar{\partial}(u\circ h) &= (\bar{\partial}u)\circ h\ \overline{h'}\ . \end{aligned} \tag{1}$$

Now define $\Delta = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}$ (the Laplacian); as we have seen, $\Delta = 4\partial\bar{\partial} = 4\bar{\partial}\partial$. Thus, if h is holomorphic, the above chain rule implies

$$\begin{aligned} \Delta(u\circ h) &= 4\bar{\partial}[(\partial u)\circ h\ h'] \\ &= 4\bar{\partial}[(\partial u)\circ h]h' + 4(\partial u)\circ h\ \bar{\partial}h' \quad \text{(Leibniz' rule)} \\ &= 4(\bar{\partial}\partial u)\circ h\ \overline{h'}h' + 0, \end{aligned}$$

so we obtain

$$\Delta(u\circ h) = (\Delta u)\circ h\ |h'|^2\ . \tag{2}$$

We need one more formula involving ∂. Suppose f is holomorphic. Then

$$\partial\ \mathrm{Re} f = \frac{\partial f + \partial\bar{f}}{2} = \frac{f' + 0}{2}\ ,$$

so we have

(3) $$\partial \operatorname{Re} f = \tfrac{1}{2} f'.$$

<u>DEFINITION 1</u>. Let u be a real-valued function defined on a Riemann surface S. Then u is <u>harmonic</u> if for every chart $\varphi: U \to W$ in the complete analytic atlas for S, $u \circ \varphi^{-1}$ is harmonic on W.

<u>PROPOSITION 1</u>. *Let S be a Riemann surface and $u: S \to \mathbb{R}$. Then the following conditions are equivalent.*

1. *u is harmonic.*
2. *For each $p \in S$ there exists a chart $\varphi: U \to W$ in the complete analytic atlas for S such that $p \in U$ and $u \circ \varphi^{-1}$ is harmonic in a neighborhood of $\varphi(p)$.*
3. *In a neighborhood of each point of S there exist holomorphic functions f and g such that $u = f + \bar{g}$.*
4. *In a neighborhood of each point of S there exists a holomorphic function F such that $u = \operatorname{Re} F$.*

<u>Proof</u>: <u>$1 \Rightarrow 2$</u>: Trivial.

<u>$2 \Rightarrow 1$</u>: If $u \circ \varphi^{-1}$ is harmonic near $\varphi(p)$ as in condition 2, and if ψ is any compatible chart near p, then

$$u \circ \psi^{-1} = u \circ \varphi^{-1} \circ (\varphi \circ \psi^{-1}),$$

so $u \circ \psi^{-1}$ is also harmonic by formula (2) on p. 164 . This proves 1.

2 ⇒ 3: Since $u\circ\varphi^{-1}$ is harmonic, $\partial(u\circ\varphi^{-1})$ is holomorphic. Locally, any holomorphic function has a primitive, so there exists a holomorphic function f near p such that near $\varphi(p)$

$$\partial(u\circ\varphi^{-1}) = (f\circ\varphi^{-1})'.$$

Define $\bar{g} = u-f$. Then

$$\bar{\partial}(g\circ\varphi^{-1}) = \overline{\partial(\bar{g}\circ\varphi^{-1})} = \overline{\partial(u\circ\varphi^{-1}) - \partial(f\circ\varphi^{-1})}$$

$$= \overline{\partial(u\circ\varphi^{-1}) - (f\circ\varphi^{-1})'}$$

$$= 0.$$

Thus, $g\circ\varphi^{-1}$ is holomorphic, proving g is holomorphic.

3 ⇒ 4: Using $u = f + \bar{g}$, we have since u is real, $u = \operatorname{Re} u = \operatorname{Re} f + \operatorname{Re} \bar{g} = \operatorname{Re} f + \operatorname{Re} g$, so we merely take $F = f + g$.

4 ⇒ 2: We have $u\circ\varphi^{-1} = \operatorname{Re}(F\circ\varphi^{-1})$ is harmonic near $\varphi(p)$.

QED

PROPOSITION 2. *Let S and T be Riemann surfaces, $F: S \to T$ an analytic mapping. If u is a harmonic function on T, then $u\circ F$ is harmonic on S.*

Proof: If φ is a chart on S and ψ a chart on T, then we must investigate $(u\circ F)\circ\varphi^{-1}$. This is

$$(u\circ F)\circ\varphi^{-1} = u\circ\psi^{-1}\circ(\psi\circ F\circ\varphi^{-1})$$

and we know $u\circ\psi^{-1}$ is harmonic and $\psi\circ F\circ\varphi^{-1}$ is holomorphic. Therefore, formula (2) of p. 164 obtains.

QED

PROPOSITION 3. Let S be a connected Riemann surface and u a harmonic function on S. If u vanishes on a neighborhood of some point of S, then $u \equiv 0$.

Proof: Define $A = \{p \in S: u \equiv 0$ in a neighborhood of $p\}$. Then A is open by definition and $A \neq \emptyset$ by hypothesis. We now demonstrate that A is closed: suppose p_o is a limit point of A. By criterion 4 of Proposition 1, there exists a holomorphic function F near p_o such that $u = \mathrm{Re}F$ near p_o. Thus, ReF vanishes on some open set near p_o, namely on the intersection of A with any neighborhood of p_o where F is defined. But then F must be constant on this open set and by the uniqueness of analytic continuation F is constant. Thus, u is constant near p_o and thus $p_o \in A$. Since A is open and closed and not empty, and since S is connected, $A = S$.

QED

The fundamental Theorem 3 actually follows from a theorem on the existence of harmonic functions, which we now state.

THEOREM 4. *Let S be any connected Riemann surface and let* $p \in S$. *Let* $\varphi: U \to W$ *be a chart in the complete analytic atlas for* S *with* $p \in U$ *and* $\varphi(p) = 0$. *Let* n *be a positive integer. Then there exists a harmonic function* u *on* $S - \{p\}$ *such that for* z *near* 0

$$u \circ \varphi^{-1}(z) = c \log|z| + \operatorname{Re} f(z),$$

where c *is some real constant and* f *is meromorphic in a neighborhood of* 0 *and has a pole of order* n *at* 0.

Thus, Theorem 4 guarantees the existence of a harmonic function on $S - \{p\}$ with prescribed singularity at p. For emphasis, we repeat that the order of the pole of f at 0 is *exactly* n: in the notation of p. 38, $\partial_f(0) = -n$.

Now we shall indicate how the knowledge of Theorem 4 leads to a proof of Theorem 3. Let p_o, q_o be the distinct points on S mentioned in the hypothesis of Theorem 3. Let u be harmonic on $S - \{p_o\}$ with representation near p_o as prescribed by Theorem 4 with (say) $n = 1$:

$$u \circ \varphi^{-1}(z) = c \log|z| + \operatorname{Re} f(z),$$

$$f(z) = \frac{\alpha}{z} + \ldots \text{ (Laurent expansion near 0)}, \ \alpha \neq 0.$$

Let ψ be a chart near q_o, $\psi(q_o) = 0$, and let v be a harmonic function on $S - \{q_o\}$ with expansion near 0 of the form

$$v\circ\psi^{-1}(z) = d \log|z| + \mathrm{Reg}(z),$$

$$g(z) = \frac{\beta}{z} + \ldots, \ \beta \neq 0.$$

Using these two harmonic functions we shall construct the meromorphic function required in Theorem 3. Here is how it is done: let $p_1 \in S$ and let σ be a chart near p_1 (in the complete analytic atlas for S). Near p_1 we define

$$F(p) = \frac{\partial(u\circ\sigma^{-1})(\sigma(p))}{\partial(v\circ\sigma^{-1})(\sigma(p))} .$$

First, we show this definition to be independent of σ. Let σ_1 be another chart near p_1 and let $h = \sigma\circ\sigma_1^{-1}$, so that h is holomorphic and has a holomorphic inverse. Then formula (1) of p.164 implies

$$\begin{aligned}\partial(u\circ\sigma_1^{-1})(\sigma_1(p)) &= \partial(u\circ\sigma^{-1}\circ h)(\sigma_1(p))\\ &= \partial(u\circ\sigma^{-1})(h(\sigma_1(p)))h'(\sigma_1(p))\\ &= \partial(u\circ\sigma^{-1})(\sigma(p))h'(\sigma_1(p)).\end{aligned}$$

Therefore,

$$\frac{\partial(u\circ\sigma_1^{-1})(\sigma_1(p))}{\partial(v\circ\sigma_1^{-1})(\sigma_1(p))} = \frac{\partial(u\circ\sigma^{-1})(\sigma(p))}{\partial(v\circ\sigma^{-1})(\sigma(p))} ,$$

since the common nonzero factor $h'(\sigma_1(p))$ cancels after division. Thus, the definition of F is independent of the choice of chart.

Next, since $u \circ \sigma^{-1}$ and $v \circ \sigma^{-1}$ are harmonic, the functions $\partial(u \circ \sigma^{-1})$ and $\partial(v \circ \sigma^{-1})$ are holomorphic, and not identically zero since otherwise e.g. $\overline{v \circ \sigma^{-1}}$ would be holomorphic (Cauchy-Riemann equation) and thus constant (since it is real-valued). But then Proposition 3 would imply that v is constant on $S - \{q_o\}$, which manifestly contradicts its singular behavior near q_o. Thus, the zeros of $\partial(v \circ \sigma^{-1})$ are isolated, so the formula for F exhibits F as the quotient of two holomorphic functions near p_1, the denominator not vanishing identically, and thus F is meromorphic near p_1. Thus, F is meromorphic on $S - \{p_o\} - \{q_o\}$.

Finally, we must examine the behavior of F near p_o and q_o. Near p_o we use the chart φ and compute according to (3) of p. 165

$$\partial(u \circ \varphi^{-1}(z)) = \frac{c}{2z} + \frac{1}{2}f'(z)$$

$$= -\frac{\alpha}{2z^2} + \dots \quad ,$$

so that $F \circ \varphi^{-1}$ has a pole of order at least 2 at 0. Thus, $F(p_o) = \infty$. Likewise, near q_o we use the chart ψ and compute

$$\partial(v \circ \psi^{-1}(z)) = -\frac{\beta}{2z^2} + \dots,$$

so that $F \circ \psi^{-1}$ has a zero of order at least 2 at 0.

Thus, $F(q_o) = 0$. This concludes the proof of Theorem 3.

We have therefore finally reduced the problem to that of demonstrating Theorem 4. It will take a considerable amount of machinery and technique in the area of harmonic function theory to accomplish this, so we now begin a discussion of the relevant properties we need.

<u>Proposition 4</u>. <u>Let</u> u <u>be continuous on</u> $\bar{\Delta}$, <u>the closure of an open disk</u> $\Delta \subset \mathbb{C}$, <u>and harmonic in</u> Δ. <u>Suppose</u> Δ <u>has center</u> z_o <u>and radius</u> r. <u>Then</u>

$$u(z_o) = \frac{1}{2\pi}\int_0^{2\pi} u(z_o + re^{i\theta})d\theta.$$

<u>Proof</u>: Let $0 < \rho < r$. Then the divergence theorem implies

$$0 = \int_{|z-z_o|<\rho} \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)dxdy = \int_{|z-z_o|=\rho} \frac{\partial u}{\partial \nu}\, dS,$$

where dS is the element of arc length on the circle $|z-z_o| = \rho$ and $\frac{\partial u}{\partial \nu}$ is the directional derivative in the direction of the outer normal. Another way of writing this is

$$0 = \int_0^{2\pi} \left(\frac{\partial}{\partial \rho}\, u(z_o + \rho e^{i\theta})\right)\rho d\theta.$$

Dividing by ρ and then moving $\frac{\partial}{\partial\rho}$ outside the sign of integration implies

$$0 = \frac{\partial}{\partial\rho}\int_0^{2\pi} u(z_o + \rho e^{i\theta})d\theta.$$

Therefore, the continuous function of $\rho \in [0,r]$ given by

$$\rho \to \frac{1}{2\pi}\int_0^{2\pi} u(z_o + \rho e^{i\theta})d\theta$$

is constant. Since its value at $\rho = 0$ is $u(z_o)$, the result follows.

QED

Now we show how to apply this simple property of harmonic functions to obtain a representation of u in all of Δ, not just at the center. First, we take Δ to be the unit disk $\{z: |z| < 1\}$ for simplicity of computations. Let $a \in \Delta$ and consider the Möbius transformation

$$T(z) = \frac{z-a}{1-\bar{a}z} ;$$

T maps Δ onto Δ conformally, $\overline{\Delta}$ onto $\overline{\Delta}$, and $T(a) = 0$. Thus $u\circ T^{-1}$ is harmonic on Δ, continuous on $\overline{\Delta}$, so that Proposition 4 implies

$$u\circ T^{-1}(0) = \frac{1}{2\pi}\int_0^{2\pi} u\circ T^{-1}(e^{i\varphi})d\varphi.$$

Now we introduce the change of variable

$$e^{i\theta} = T^{-1}(e^{i\varphi}).$$

Then $e^{i\varphi} = T(e^{i\theta})$, so that a simple computation yields

$$\frac{d\varphi}{d\theta} = \frac{e^{i\theta}}{e^{i\theta} - a} + \frac{\bar{a}e^{i\theta}}{1-\bar{a}e^{i\theta}}$$

$$= \frac{e^{i\theta}}{e^{i\theta} - a} + \frac{\bar{a}}{e^{-i\theta}-\bar{a}}$$

$$= \frac{1-\bar{a}e^{i\theta}+\bar{a}e^{i\theta}-\bar{a}a}{(e^{i\theta}-a)(e^{-i\theta}-\bar{a})}$$

$$= \frac{1-|a|^2}{|e^{i\theta}-a|^2}.$$

Therefore,

$$u(a) = \frac{1}{2\pi}\int_0^{2\pi} \frac{1-|a|^2}{|e^{i\theta}-a|^2}\, u(e^{i\theta})d\theta.$$

Define

(4) $$P(z,e^{i\theta}) = \frac{1}{2\pi}\,\frac{1-|z|^2}{|e^{i\theta}-z|^2}, \quad |z| < 1;$$

this is the so-called Poisson kernel. We want to observe certain things about it:

1. $P \geq 0$;
2. $\int_0^{2\pi} P(z,e^{i\theta})d\theta = 1$, $|z| < 1$;
3. $P(z,e^{i\theta})$ is a harmonic function of z;
4. for any $\delta > 0$,

$$\lim_{z\to e^{i\theta_0}} \int_{|e^{i\theta}-e^{i\theta_0}|\geq\delta} P(z,e^{i\theta})d\theta = 0.$$

The first property is obvious and the second follows from formula (4) applied to the harmonic function $u \equiv 1$. The third follows from the formula for $\frac{d\varphi}{d\theta}$, which reads

$$2\pi P(z,e^{i\theta}) = \frac{e^{i\theta}}{e^{i\theta}-z} + \frac{\bar{z}e^{i\theta}}{1-\bar{z}e^{i\theta}} \quad ,$$

exhibiting P as a sum of two harmonic functions of z. To prove the fourth, assume $|z-e^{i\theta_0}| < \delta/2$. Then $|e^{i\theta}-z| \geq |e^{i\theta}-e^{i\theta_0}|-|e^{i\theta_0}-z| > \delta-\delta/2 = \delta/2$, so that

$$\int_{|e^{i\theta}-e^{i\theta_0}|\geq\delta} P(z,e^{i\theta})d\theta \leq \frac{1}{2\pi}\,\frac{1-|z|^2}{(\delta/2)^2}\cdot 2\pi < \frac{8}{\delta^2}(1-|z|),$$

and this clearly tends to zero as $z \to e^{i\theta_0}$.

These four properties are all we need to establish the following converse to formula (4).

PROPOSITION 5. *Let f be a continuous function on the circle $|z| = 1$. Define*

$$u(z) = \begin{cases} \int_0^{2\pi} P(z,e^{i\theta})f(e^{i\theta})d\theta, & |z| < 1, \\ f(z), & |z| = 1. \end{cases}$$

Then u is harmonic for $|z| < 1$ and continuous for $|z| \leq 1$.

Proof: The fact that u is harmonic for $|z| < 1$ follows from 3 by differentiation under the integral sign. Clearly, we need only prove that $\lim u(z) = f(e^{i\theta_0})$ for $|z| < 1$, $z \to e^{i\theta_0}$, in order to finish the proof. Let $\varepsilon > 0$. By continuity of f at $e^{i\theta_0}$, there exists $\delta > 0$ such that $|f(e^{i\theta}) - f(e^{i\theta_0})| < \frac{\varepsilon}{2}$ if $|e^{i\theta} - e^{i\theta_0}| < \delta$.

Now 2 implies

$$u(z) - f(e^{i\theta_0}) = \int_0^{2\pi} P(z,e^{i\theta})[f(e^{i\theta})-f(e^{i\theta_0})]d\theta.$$

Choose a constant C such that $|f(e^{i\theta})| \le C$ for all θ.

Then

$$|u(z)-f(e^{i\theta_0})| \le \frac{\varepsilon}{2} \int_{|e^{i\theta}-e^{i\theta_0}|<\delta} P(z,e^{i\theta})d\theta$$

$$+ 2C \int_{|e^{i\theta}-e^{i\theta_0}|\ge\delta} P(z,e^{i\theta})d\theta.$$

Since the first integral is bounded by 1, and property 4 implies there exists $\delta' > 0$ such that the second integral is bounded by $\frac{\varepsilon}{4C}$ if $|z-e^{i\theta_0}| < \delta'$, we obtain

$$|u(z)-f(e^{i\theta_0})| < \varepsilon$$

if $|z-e^{i\theta_0}| < \delta'$.

QED

Of course, it is not necessary to restrict our attention to the unit disk. If we consider functions in the disk Δ of center z_o and radius r, the formula analogous to that of Proposition 5 is

$$u(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{r^2-|z-z_o|^2}{|re^{i\theta}-(z-z_o)|^2} f(z_o + re^{i\theta})d\theta.$$

This can be derived in the same manner, or merely by considering the change of variable $z-z_o = rw$ and using Proposition 5 as it stands.

The Poisson integral formula we have just derived has several immediate applications which will be of great importance to us. For example, we have

PROPOSITION 6. *Let* D *be an open set in* $\mathbb{C}$ *and* K *a compact subset of* D. *Then there exists a constant* C *which depends only on* K *and* D *such that if* u *is harmonic in* D *then*

$$\sup_K\left|\frac{\partial u}{\partial x}\right| \le C \sup_D |u|.$$

A similar result holds with $\frac{\partial}{\partial x}$ *replaced by any derivative of any order.*

Proof: For any $z_o \in K$ there exists a disk Δ of center z_o and radius r such that the closure of Δ is contained in D. For $|z-z_o| < \frac{1}{2}r$, the Poisson integral formula implies

$$\left|\frac{\partial u}{\partial x}(z)\right| \le c \sup_D|u|,$$

where

$$c = \sup_{\substack{|z-z_o|<\frac{1}{2}r \\ 0\le\theta\le 2\pi}} \left| \frac{\partial}{\partial x} \frac{r^2-|z-z_o|^2}{|re^{i\theta}-(z-z_o)|^2} \right|$$

is easily seen to be finite. Since K can be covered by finitely many such disks as $\{z: |z-z_o| < \frac{1}{2} r\}$, the result follows.

QED

PROPOSITION 7. Let D be an open set in $\mathbb{C}$ and $u_1, u_2, \ldots$ a sequence of harmonic functions in D which converge uniformly on compact subsets of D to a function u. Then u is harmonic in D and the sequence $\frac{\partial u_n}{\partial x}$ converges to $\frac{\partial u}{\partial x}$, also uniformly on compact sets in D.

Proof: If Δ is a disk whose closure is contained in D, then u_n has a Poisson integral representation in Δ of the form

$$u_n(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{r^2-|z-z_o|^2}{|re^{i\theta}-(z-z_o)|^2} u_n(z_o + re^{i\theta})d\theta.$$

For fixed $z \in \Delta$ let $n \to \infty$ in this formula and use the uniform convergence to pass the limit under the integral sign to obtain

$$u(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{r^2-|z-z_o|^2}{|re^{i\theta}-(z-z_o)|^2} u(z_o+re^{i\theta})d\theta.$$

Therefore, u is harmonic in Δ. Therefore, u is harmonic in D.

Now suppose K is a compact subset of D. Choose an open set D_1 such that $K \subset D_1$ and the closure of D_1 is a compact subset of D. Let C be the constant of Proposition 6 relative to K and D_1. Then

$$\sup_K \left|\frac{\partial u_n}{\partial x} - \frac{\partial u}{\partial x}\right| \le C \sup_{D_1} |u_n - u| \quad .$$

By hypothesis, $\sup_{D_1} |u_n - u| \to 0$ as $n \to \infty$, and therefore $\frac{\partial u_n}{\partial x} \to \frac{\partial u}{\partial x}$ uniformly on K. As K is arbitrary, the result follows.

QED

PROPOSITION 8. Let D be an open set in $\mathbb{C}$ and $u_1 \cdot u_2, \ldots$ a sequence of harmonic functions in D which are uniformly bounded on every compact subset of D. Then there exists a subsequence $n_1 < n_2 < \ldots$ such that

$$\lim_{k \to \infty} u_{n_k}$$

exists uniformly on compact subsets of D.

Proof: If Δ is a disk such that its closure is a compact subset of D, then there exists a constant C depending only on Δ such that $|u_n(z)| \le C$ for $z \in \Delta$, $n \ge 1$. Therefore, Proposition 6 implies that if $\frac{1}{2}\Delta$ is the concentric disk with half the radius of Δ, then for some other constant C_1

$$\left|\frac{\partial u_n}{\partial x}\right| \le C_1, \quad \left|\frac{\partial u_n}{\partial y}\right| < C_1 \text{ on } \tfrac{1}{2}\Delta.$$

Now we apply the mean value theorem on the disk $\frac{1}{2}\Delta$ (details omitted) to conclude that for $z, z' \in \frac{1}{2}\Delta$,

$$|u_n(z)-u_n(z')| \leq 2C_1|z-z'|.$$

This proves that the family of functions $u_1, u_2, \ldots$ is equicontinuous on $\frac{1}{2}\Delta$. Since Δ was arbitrary, it follows that the family $u_1, u_2, \ldots$ is equicontinuous on each compact subset of D. By the Arzela-Ascoli theorem, there exists a subsequence with the required property that u_{n_k} converges uniformly on compact subsets of D.

QED

DEFINITION 2. An open subset D of a Riemann surface is an analytic disk if there exists a chart φ: U → W in the complete analytic atlas such that $\varphi(D)$ is a disk whose closure is a (compact) subset of W.

Notation. If A is a subset of a topological space, A^- denotes the closure of A and ∂A denotes the boundary of A.

PROPOSITION 9. Let D be an analytic disk in a Riemann surface S and let f: $\partial D \to \mathbb{R}$ be continuous. Then there exists a unique function P_f on D^- such that P_f is continuous on D^-, harmonic in D, and $P_f \equiv f$ on ∂D.

Proof: Let φ: U → W be a chart in the complete analytic atlas for S satisfying the condition of Definition 2. If $\varphi(D) = \{z: |z-z_o| < r\}$, then

$P_f \circ \varphi^{-1}$ must be continuous on $\varphi(D)^-$, harmonic on $\varphi(D)$, and $P_f \circ \varphi^{-1} \equiv f \circ \varphi^{-1}$ on $\partial\varphi(D)$ $(= \varphi(\partial D))$. Thus, if $z \in \varphi(D)$, then

$$P_f \circ \varphi^{-1}(z) = \frac{1}{2\pi} \int_0^{2\pi} \frac{r^2 - |z-z_o|^2}{|re^{i\theta} - (z-z_o)|^2} f \circ \varphi^{-1}(z_o + re^{i\theta}) d\theta.$$

Therefore, P_f is uniquely determined and Proposition 5 implies that P_f as defined by this formula satisfies the conditions of Proposition 9.

QED

DEFINITION 3. If D is an analytic disk in a Riemann surface S and if $u: S \to \mathbb{R}$ is continuous, u_D is the unique continuous function on S which agrees with u on S-D and is harmonic in D. The existence and uniqueness of u_D are guaranteed by Proposition 9.

LEMMA 2. *Let S be a Riemann surface and $p_o \in S$. Let $\varepsilon > 0$. Then there exists a neighborhood U of p_o such that for all functions u which are harmonic and nonnegative on S, and for all $p,q \in U$*

$$u(p) \leq (1+\varepsilon)u(q).$$

Proof: There exists a chart $\varphi: U_o \to W$ in the complete analytic atlas for S such that W contains $\{z: |z| \leq 1\}$ and $\varphi(p_o) = 0$. This can obviously be achieved by composing an arbitrary chart with a suitable linear transformation of $\mathbb{C}$ onto itself. Let $v = u \circ \varphi^{-1}$. Then according to p. 173,

$$v(z) = \int_0^{2\pi} P(z,e^{i\theta})v(e^{i\theta})d\theta, \quad |z| < 1.$$

Now $1-|z| \le |e^{i\theta}-z| \le 1 + |z|$, so we obtain

$$\frac{1-|z|}{1+|z|} = \frac{1-|z|^2}{(1+|z|)^2} \le \frac{1-|z|^2}{|e^{i\theta}-z|^2} \le \frac{1-|z|^2}{(1-|z|)^2}$$

$$= \frac{1+|z|}{1-|z|} .$$

Therefore, since $v(e^{i\theta}) \ge 0$,

$$\frac{1-|z|}{1+|z|} \cdot \frac{1}{2\pi} \int_0^{2\pi} v(e^{i\theta})d\theta \le v(z) \le \frac{1+|z|}{1-|z|} \frac{1}{2\pi} \int_0^{2\pi} v(e^{i\theta})d\theta.$$

By Proposition 4 this pair of inequalities can be written in the form

$$\frac{1-|z|}{1+|z|} v(0) \le v(z) \le \frac{1+|z|}{1-|z|} v(0).$$

If $0 < \delta < 1$ and $|z| \le \delta$, we obtain

$$\frac{1-\delta}{1+\delta} v(0) \le v(z) \le \frac{1+\delta}{1-\delta} v(0) .$$

Therefore, if $|z| \le \delta$ and $|w| \le \delta$,

$$v(z) \le \frac{1+\delta}{1-\delta} v(0) \le \left(\frac{1+\delta}{1-\delta}\right)^2 v(w).$$

Pick δ such that

$$\left(\frac{1+\delta}{1-\delta}\right)^2 \leq 1 + \varepsilon.$$

Then let $U = \varphi^{-1}(\{z: |z| < \delta\})$. This is a neighborhood of $p_o = \varphi^{-1}(0)$ and for $p,q \in U$,

$$u(p) = v(\varphi(p)) \leq (1+\varepsilon)v(\varphi(q)) = (1+\varepsilon)u(q).$$

QED

Harnack's Inequality. *Let* S *be a connected Riemann surface and* K *a compact subset of* S. *Then there exists a constant* C *depending only on* K *and* S *such that for all nonnegative harmonic functions* u *on* S *and all* $p,q \in K$,

$$u(p) \leq Cu(q).$$

Proof: It obviously suffices to consider the class $\mathcal{H}$ of functions which are harmonic and positive on S; if u is harmonic and $u \geq 0$, then for every $\varepsilon > 0$, $u+\varepsilon \in \mathcal{H}$ and if the inequality is true for functions in $\mathcal{H}$ then $u(p) + \varepsilon \leq C(u(q) + \varepsilon)$. Then let $\varepsilon \to 0$. Of course, we are debating a triviality anyway, because Harnack's inequality implies that if $u \geq 0$ and u is harmonic, then either $u \equiv 0$ or $u > 0$. Now choose some fixed point $p_o \in S$ and define

$$F(p) = \sup \; \max \left(\frac{u(p)}{u(p_o)}, \frac{u(p_o)}{u(p)}\right): u \in \mathcal{H}.$$

We are going to prove F is continuous. Let $p_1 \in S$ and let $\varepsilon > 0$. Let U be a neighborhood of p_1 satisfying the condition of Lemma 2. Then for $u \in \mathcal{H}$ and $p,q \in U$,

$$\frac{u(p)}{u(p_o)} \le (1+\varepsilon)\frac{u(q)}{u(p_o)} \le (1+\varepsilon)F(q),$$

$$\frac{u(p_o)}{u(p)} \le (1+\varepsilon)\frac{u(p_o)}{u(q)} \le (1+\varepsilon)F(q).$$

Therefore,

(5) $\qquad F(p) \le (1+\varepsilon)\,F(q) \quad \text{for } p,q \in U.$

In particular, if $F(p_1) < \infty$ we choose $q = p_1$ to conclude that $F(p) < \infty$ for all $p \in U$; if $F(p_1) = \infty$ we choose $p = p_1$ to conclude that $F(q) = \infty$ for all $q \in U$. Therefore, the sets

$$\{p \in S\colon F(p) < \infty\},\ \{p \in S\colon F(p) = \infty\}$$

are open. As they are obviously disjoint and their union is obviously S, the connectedness of S implies one of these sets is S, the other empty. Since $F(p_o) = 1$, p_o belongs to the first of the sets, and thus we have proved that $F < \infty$ everywhere on S.

Now we obtain the continuity. Taking $q = p_1$ in (5),

$$F(p)-F(p_1) \le \varepsilon F(p_1) \text{ if } p \in U;$$

taking $p = p_1$,

$$-\varepsilon F(p_1) \le (1+\varepsilon)(F(q)-F(p_1)) \text{ if } q \in U.$$

Thus, we obtain

$$-\frac{\epsilon}{1+\epsilon}F(p_1) \le F(p)-F(p_1) \le \epsilon F(p_1) \text{ if } p \in U,$$

and since ϵ is arbitrary, this proves that F is continuous at p_1.

Since F is continuous on S and K is compact, there exists a constant c such that $F(p) \le c$ for $p \in K$. Therefore, if $p,q \in K$

$$u(p) \le F(p)u(p_o) \le F(p)F(q)u(q) \le c^2u(q).$$

QED

Harnack's Convergence Theorem. *Let S be a connected Riemann surface and $\mathfrak{F}$ a nonvoid family of harmonic functions on S which is directed upwards, i.e., if $u,v \in \mathfrak{F}$, there exists $w \in \mathfrak{F}$ such that $w \ge u$, $w \ge v$. Let $U = \sup \mathfrak{F}$, i.e., for $p \in S$*

$$U(p) = \sup\{u(p): u \in \mathfrak{F}\}.$$

Then there exists a sequence $u_1 \le u_2 \le u_3 \le \cdots$ such that $u_n \in \mathfrak{F}$ for all n and $u_n \to U$ uniformly on compact subsets of S. Moreover, either $U \equiv \infty$ or U is harmonic on S.

Proof: Let u_o be an arbitrary function in $\mathfrak{F}$ and let

$$\mathfrak{F}' = \{u \in \mathfrak{F}: u \ge u_o\}.$$

Then $\sup \mathfrak{F}' = \sup \mathfrak{F}$. Indeed, since $\mathfrak{F}' \subset \mathfrak{F}$ the inequality $\sup \mathfrak{F}' \le \sup \mathfrak{F}$ is obvious, and if $u \in \mathfrak{F}$ then there exists $v \in \mathfrak{F}$ such that $v \ge u$ and $v \ge u_o$; therefore, $v \in \mathfrak{F}'$ and

$v \geq u$, proving that $\sup \mathfrak{F}' \geq \sup \mathfrak{F}$.

For any compact set $K \subset S$ let C_K be the corresponding constant in the conclusion of Harnack's inequality. Then for $u \in \mathfrak{F}'$, $u-u_o$ is a nonnegative harmonic function, so for all $p,q \in K$ it follows that

$$\begin{aligned} u(p) - u_o(p) &\leq C_K(u(q) - u_o(q)) \\ &\leq C_K(U(q) - u_o(q)). \end{aligned}$$

Taking the supremum over all $u \in \mathfrak{F}'$ implies

$$(6) \qquad U(p) - u_o(p) \leq C_K(U(q) - u_o(q)).$$

It follows that if $U(q) < \infty$, then $U(p) < \infty$, and here p,q can be any points in S (just take K to be the compact set $\{p,q\}$). Therefore, either $U \equiv \infty$ or $U < \infty$.

Let $p_o \in S$ be fixed and choose a sequence $u_1', u_2', u_3', \ldots$ from $\mathfrak{F}$ such that $u_n'(p_o) \to U(p_o)$. By hypothesis, we can let $u_1 = u_1'$ and then find inductively $u_n \in \mathfrak{F}$ such that

$$u_n \geq u_n', \; u_n \geq u_{n-1}.$$

Then we have a sequence $u_1 \leq u_2 \leq u_3 \leq \ldots$ from $\mathfrak{F}$ such that $u_n(p_o) \to U(p_o)$. Note that (6) holds for an arbitrary $u_o \in \mathfrak{F}$, so we can take $u_o = u_n$ in (6). If $U(p_o) = \infty$, then for any compact set K containing p_o Harnack's inequality implies

$$u_n(p_o) - u_1(p_o) \leq C_K(u_n(q) - u_1(q)), \; q \in K,$$

and therefore $u_n(q) \to \infty$ uniformly on K. This proves the result in case $U \equiv \infty$. If $U(p_o) < \infty$, then (6) implies

$$U(p) - u_n(p) \leq C_K(U(p_o) - u_n(p_o)), \quad p \in K,$$

and therefore $u_n(p) \to U(p)$ uniformly for $p \in K$. Finally, Proposition 7 shows that U is harmonic in this case.

QED

Now we need to introduce the basic building block other than the Poisson integral, which is *subharmonic* functions. The basic theory is contained in the following proposition.

PROPOSITION 10. *Let* u *be a continuous real-valued function on a connected Riemann surface* S. *Then the following conditions are equivalent.*

1. *For every analytic disk* $D \subset S$, $u \leq u_D$. (Cf. Definition 3.)
2. *If* $\mathcal{O}$ *is a proper open subset of* S, *if* $\mathcal{O}^-$ *is compact, if* h *is continuous on* $\mathcal{O}^-$ *and harmonic in* $\mathcal{O}$, *and if* $u \leq h$ *on* $\partial\mathcal{O}$, *then* $u \leq h$ *in* $\mathcal{O}$.
3. *For each* $p \in S$ *there exists a chart* $\varphi\colon U \to W$ *in the complete analytic atlas for* S *such that* $p \in U$ *and*

$$u(p) \leq \frac{1}{2\pi}\int_0^{2\pi} u\circ\varphi^{-1}(\varphi(p)+re^{i\theta})\,d\theta$$

for small positive r.

4. <u>For each</u> $p \in S$ <u>and every chart</u> $\varphi: U \to W$ <u>in the complete analytic atlas for</u> S <u>such that</u> $p \in U$ <u>and every</u> z <u>sufficiently close to</u> $\varphi(p)$,

$$u \circ \varphi^{-1}(z) \leq \frac{1}{2\pi} \int_0^{2\pi} u \circ \varphi^{-1}(z+re^{i\theta})\,d\theta$$

<u>for small positive</u> r.

<u>Proof</u>: We are going to establish four implications, three of which are absolutely trivial.

<u>2⇒1</u>: Assume that 2 holds and let D be an analytic disk. Use 2 with $\mathcal{O} = D$ and $h = u_D$ restricted to $\bar{\mathcal{O}}$. Then $u = h$ on ∂D so we obtain $u \leq h$ in D, i.e., $u \leq u_D$ in D. Since $u = u_D$ outside D, 1 follows.

<u>1⇒4</u>: Assume that 1 holds and consider the analytic disk

$$D = \varphi^{-1}(\{w: |z-w| < r\})$$

for sufficiently small r. Then $u \leq u_D$, so in particular

$$u \circ \varphi^{-1}(z) \leq u_D \circ \varphi^{-1}(z).$$

Since $u_D \circ \varphi^{-1}$ is continuous on $\{w: |z-w| \leq r\}$ and harmonic in the interior, the mean value property of Proposition 4 implies

$$\begin{aligned} u_D \circ \varphi^{-1}(z) &= \frac{1}{2\pi} \int_0^{2\pi} u_D \circ \varphi^{-1}(z+re^{i\theta})\,d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} u \circ \varphi^{-1}(z+re^{i\theta})\,d\theta \end{aligned}$$

since $u_D = u$ on ∂D. Thus, 4 follows.

4⇒3: Completely trivial: we allow every chart in 4 and moreover 3 is just 4 at the single point $z = \varphi(p)$.

3⇒2: Finally here is something which requires thought. Assume that 3 holds and assume we have the hypothesis of 2. Define $v = u-h$ in $\bar{\mathcal{O}}$. Let $M = \sup_{\bar{\mathcal{O}}} v$. Since $\bar{\mathcal{O}}$ is compact and v is continuous, the supremum is attained, so the set

$$A = \{p \in \mathcal{O}\colon v(p) = M\}$$

is either nonvoid or $v = M$ somewhere on $\partial\mathcal{O}$. In the latter case, since $v \le 0$ on $\partial\mathcal{O}$ we obtain $M \le 0$ and the result follows. So we assume $v < M$ everywhere on $\partial\mathcal{O}$, in which case A is not empty. Since v is continuous, the set A is closed relative to $\mathcal{O}$. We use 3 to show that A is open: suppose $p \in A$. Pick a chart φ according to 3 with respect to the point p. Then for small positive r

$$u(p) \le \frac{1}{2\pi}\int_0^{2\pi} u\circ\varphi^{-1}(\varphi(p) + re^{i\theta})d\theta.$$

Since h is harmonic, it satisfies the similar relation with equality instead of inequality (Proposition 4) and thus we obtain by subtraction

$$v(p) \le \frac{1}{2\pi}\int_0^{2\pi} v\circ\varphi^{-1}(\varphi(p) + re^{i\theta})d\theta.$$

But the left side is $v(p) = M$ and the integrand

$v\circ\varphi^{-1}(\varphi(p)+re^{i\theta}) \le M$ by definition of M. Thus, we conclude that equality holds everywhere, so

$$v\circ\varphi^{-1}(\varphi(p) + re^{i\theta}) \equiv M$$

for $0 \le \theta \le 2\pi$ and small positive r. Thus, $v \equiv M$ near p, so A is open.

It is now a simple topological argument to show that ∂A and $\partial\mathfrak{G}$ have a point in common. To see this, let $p_0 \in A$ and $p_1 \in S-\mathfrak{G}$ be chosen arbitrarily, and use the connectedness of S to conclude that there exists a path γ in S from p_0 to p_1. Since $p_0 \in A$ and $p_1 \notin A$ and the image of γ is connected, there exists a point p_2 in the image of γ such that $p_2 \in \partial A$. If a neighborhood of p_2 were disjoint from $\mathfrak{G}$, it would also be disjoint from A, contradicting $p_2 \in A^-$. Therefore, $p_2 \in \mathfrak{G}^-$. If $p_2 \in \mathfrak{G}$, then since A is closed in $\mathfrak{G}$, $p_2 \in A$; since A is open, this contradicts the fact that $p_2 \in (S-A)^-$. Thus $p_2 \in \partial\mathfrak{G}$. Since $p_2 \in \partial A$, $v(p_2) = M$ by continuity of v. Since $p_2 \in \partial\mathfrak{G}$, $v(p_2) \le 0$ by hypothesis. Therefore, $M \le 0$, and the conclusion of 2 follows.

QED

DEFINITION 4. A continuous real-valued function u on a Riemann surface S is *subharmonic* if it satisfies condition 4 of Proposition 10.

Strong Maximum Principle. *Let u be a subharmonic function on a connected Riemann surface S such that $u \le 0$. Then either $u < 0$ on S or $u \equiv 0$ on S.*

Proof: This is contained in the proof of $3 \Rightarrow 2$ of Proposition 10. For the set $A = \{p \in S: u(p) = 0\}$ is closed since u is continuous and is open by Condition 4 of Proposition 10, and thus either $A = S$ or A is empty.

QED

Weak Maximum Principle. *Let S be a connected Riemann surface, and $\mathcal{O}$ a proper open subset of S such that $\mathcal{O}^-$ is compact. Let u be continuous on $\mathcal{O}^-$ and subharmonic on $\mathcal{O}$. Assume $u \le 0$ on $\partial\mathcal{O}$. Then $u \le 0$.*

Proof: This is again contained in the proof of $3 \Rightarrow 2$ of Proposition 10. If $M = \max_{\mathcal{O}^-} u$ and if $M > 0$, let $A = \{p \in \mathcal{O}: u(p) = M\}$. Then the argument proving $3 \Rightarrow 2$ shows that ∂A and $\partial\mathcal{O}$ have a point in common and thus $M \le 0$, a contradiction.

QED

COROLLARY. *If $\mathcal{O}$ is a proper open subset of a connected Riemann surface S such that $\mathcal{O}^-$ is compact, and if u is harmonic on $\mathcal{O}$ and continuous on $\mathcal{O}^-$, then*

$$\sup_{\mathcal{O}^-} |u| = \sup_{\partial\mathcal{O}} |u|.$$

Proof: Let $M = \sup_{\partial\mathfrak{G}}|u|$. Then $-M+u$ and $-M-u$ are subharmonic in $\mathfrak{G}$ and nonpositive on $\partial\mathfrak{G}$, so the weak maximum principle implies $-M+u \le 0$ and $-M-u \le 0$ in $\mathfrak{G}$. That is, $-M \le u \le M$.

QED

PROPOSITION 11. *Let* u *be a continuous real-valued function on a Riemann surface* S. *Then* u *is harmonic if and only if* u *and* $-u$ *are subharmonic*.

Proof: We only have to prove the "if" part of the assertion. Since the proposition deals with local properties, we can assume S is connected. By part 1 of Proposition 10, if D is an analytic disk, then $u \le u_D$ and $-u \le (-u)_D$. But clearly $(-u)_D = -u_D$, so we have $u \le u_D$ and $-u \le -u_D$. Thus, $u \equiv u_D$. Therefore, u is harmonic in D. Since every point of S is contained in an analytic disk, u is harmonic in S.

QED

PROPOSITION 12. *Let* $u, u_1, \ldots, u_n$ *be subharmonic on a Riemann surface* S, *and let* $a_1, a_2, \ldots, a_n$ *be nonnegative real numbers*. *Then the functions*

$$a_1u_1 + \ldots + a_nu_n,$$

$$\max(u_1, \ldots, u_n)$$

are subharmonic. *Also*, *if* D *is an analytic disk*, u_D *is subharmonic*.

Proof: This follows directly from the definition. Condition 4 of Proposition 10 asserts in that notation that for small positive r

$$u_k \circ \varphi^{-1}(z) \le \frac{1}{2\pi} \int_0^{2\pi} u_k \circ \varphi^{-1}(z+re^{i\theta})\,d\theta.$$

Multiplying by a_k and adding, the function $a_1u_1+\ldots+a_nu_n$ is seen to satisfy condition 4. If $u = \max(u_1,\ldots,u_n)$, then we have

$$u_k \circ \varphi^{-1}(z) \le \frac{1}{2\pi} \int_0^{2\pi} u \circ \varphi^{-1}(z+re^{i\theta})\,d\theta, \quad 1 \le k \le n.$$

Therefore,

$$u \circ \varphi^{-1}(z) \le \frac{1}{2\pi} \int_0^{2\pi} u \circ \varphi^{-1}(z+re^{i\theta})\,d\theta,$$

proving that u is subharmonic. The last statement will be proved on p. 196.

QED

The basic theorem we need is the following.

THEOREM 5. Let S be a connected Riemann surface and $\mathfrak{F}$ a nonempty family of subharmonic functions on S such that

1. *if* $u,v \in \mathfrak{F}$, *then* $\max(u,v) \in \mathfrak{F}$,
2. *if* $u \in \mathfrak{F}$ *and D is an analytic disk in S, then* $u_D \in \mathfrak{F}$.

Then $\sup\mathfrak{F}$ *is either harmonic in S or* $\sup \mathfrak{F} \equiv \infty$.

<u>Proof</u>: Let $U = \sup \mathfrak{F}$. If D is an analytic disk in S, let $\mathfrak{F}_D$ be the functions <u>on D</u> defined by

$$\mathfrak{F}_D = \{u_D : u \in \mathfrak{F}\}.$$

Then $\mathfrak{F}_D$ is a family of harmonic functions on D and

$$\sup \mathfrak{F}_D = U \text{ in } D.$$

For, $u \in \mathfrak{F}$ implies $u_D \in \mathfrak{F}$, so that any function in $\mathfrak{F}_D$ is the restriction to D of a function in $\mathfrak{F}$ and thus $\sup \mathfrak{F}_D \le U$ in D. On the other hand, $u \in \mathfrak{F}$ implies $u \le u_D$ by Proposition 10.1. Therefore, $U \le \sup \mathfrak{F}_D$ in D.

Now we apply Harnack's convergence theorem to the family $\mathfrak{F}_D$ on the Riemann surface D. We have to check that $\mathfrak{F}_D$ is directed upwards. So suppose $u,v \in \mathfrak{F}$. Let $w = \max(u,v)$, so that $w \in \mathfrak{F}$ by property 1. Then $u \le w$ implies $u_D \le w_D$ and $v \le w$ implies $v_D \le w_D$, so that we have found $w_D \in \mathfrak{F}_D$ such that $w_D \ge u_D$, $w_D \ge v_D$. Thus $\mathfrak{F}_D$ is directed upwards. Thus, Harnack's convergence theorem implies that either $\sup \mathfrak{F}_D$ is harmonic or $\sup \mathfrak{F}_D \equiv \infty$. Therefore, either U is harmonic on D or $U \equiv \infty$ on D.

Finally, we have the familiar connectivity argument: if $A = \{p \in S : U(p) = \infty\}$ and $B = \{p \in S : U(p) < \infty\}$, then A and B are disjoint open sets with union S. Since

S is connected, either A is empty or $A = S$. Thus, either $U < \infty$ on S or $U \equiv \infty$ on S. If $U < \infty$ on S, we have shown that U is harmonic in every analytic disk in S. Therefore, U is harmonic.

QED

Problem 8. (The Dirichlet problem for an annulus)

1. Prove the Weierstrass approximation theorem for a circle. That is, if f is a continuous complex-valued function on the circle $|z| = 1$ and $\epsilon > 0$, then there exists a *finite* sum

$$g(z) = \Sigma a_n z^n \text{ (positive and negative } n)$$

such that $|f(z)-g(z)| < \epsilon$ for $|z| = 1$.

Hint: Use Proposition 5 and Proposition 1, with the obvious remark that the proof of $1 \Rightarrow 3$ for the disk $|z| < 1$ gives two holomorphic functions defined on the entire disk.

2. Consider the annulus $r < |z| < 1$, where $0 < r < 1$ is fixed. Let n be an integer. Exhibit the (unique) harmonic function which equals z^n for $|z| = 1$ and equals 0 for $|z| = r$.

3. Combine 1 and 2 to conclude that there exists a function u_ε which is harmonic for $r < |z| < 1$, continuous for $r \le |z| \le 1$ (in fact, it will be harmonic for $0 < |z| < \infty$) such that

$$|u_\varepsilon(z)-f(z)| < \varepsilon \text{ for } |z| = 1,$$
$$u_\varepsilon(z) = 0 \text{ for } |z| = r.$$

4. By a limiting argument, prove there exists a function u which is harmonic for $r < |z| < 1$, continuous for $r \le |z| \le 1$, such that

$$u(z) = f(z) \quad \text{for} \quad |z| = 1,$$
$$u(z) = 0 \quad \text{for} \quad |z| = r.$$

5. Use this result and an appropriate conformal mapping to treat any continuous boundary values on $|z| = r$ as well.

Now we state a corollary, and we use the obvious terminology that a function w is <u>superharmonic</u> if $-w$ is subharmonic.

<u>COROLLARY</u>. <u>Let</u> w <u>be a superharmonic function on a connected Riemann surface</u> S. <u>Let</u>

$$\mathfrak{F} = \{v\colon v \text{ } \underline{\text{subharmonic on}} \text{ } S,\ v \le w\}.$$

<u>Then</u> $\sup \mathfrak{F}$ <u>is either harmonic in</u> S <u>or</u> $\sup \mathfrak{F} \equiv -\infty$.

Proof: We have $\sup \mathfrak{F} \equiv -\infty$ if and only if $\mathfrak{F}$ is empty, so we assume from now on that $\mathfrak{F}$ is not empty. We verify the two properties required in Theorem 5. First, if $v_1, v_2 \in \mathfrak{F}$, then clearly $\max(v_1,v_2) \le w$ and Proposition 12 implies $\max(v_1,v_2)$ is subharmonic; thus, $\max(v_1,v_2) \in \mathfrak{F}$. If D is an analytic disk in S, then for $v \in \mathfrak{F}$,

$$v_D \le w_D \le w ,$$

the latter inequality being a consequence of criterion 1 of Proposition 10 for superharmonic functions. So we need only check v_D is subharmonic. This amounts to checking the local criterion 3 of Proposition 10. This mean value criterion clearly holds at any point $p \in D$ (since v_D is harmonic near p) and at any point p in $S-\overline{D}$ (since $v_D \equiv v$ is subharmonic near p). So we consider $p \in \partial D$ and a chart φ in the complete analytic atlas for S, φ defined near p. Since v is subharmonic and $v \le v_D$, we obtain for small positive r

$$\begin{aligned} v_D(p) = v(p) &\le \frac{1}{2\pi}\int_0^{2\pi} v\circ\varphi^{-1}(\varphi(p) + re^{i\theta})d\theta \\ &< \frac{1}{2\pi}\int_0^{2\pi} v_D\circ\varphi^{-1}(\varphi(p) + re^{i\theta})d\theta, \end{aligned}$$

establishing the criterion in this case as well. Thus, v_D is subharmonic on S. Now Theorem 5 implies $\sup \mathfrak{F} \equiv \infty$ (which is impossible in this case) or $\sup \mathfrak{F}$ is harmonic.

QED

DEFINITION 5. Let w be a superharmonic function on a connected Riemann surface. The function

$$u = \sup\{v: v \text{ subharmonic on } S,\ v \le w\}$$

is called the greatest harmonic minorant of w. This terminology agrees with the obvious fact that if v is harmonic on S and $v \le w$, then $v \le u$. Moreover, if w has any harmonic minorant at all, then u is harmonic (not $\equiv -\infty$). Of course, our corollary shows that actually u is the greatest subharmonic minorant of w, and is itself harmonic. We shall use the abbreviation

$$u = \text{GHM of } w.$$

As an application of these ideas, we show how to solve a certain kind of Dirichlet problem.

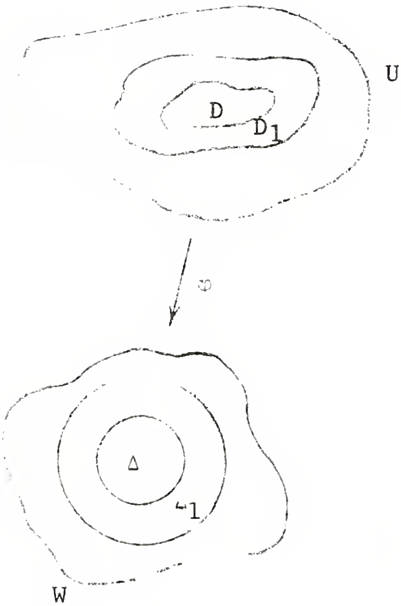
PROPOSITION 13. *Let D be an analytic disk in a connected Riemann surface S and $f: \partial D \to \mathbb{R}$ a continuous function. Then there exists a function u which is continuous in $S-D$, harmonic in $S-D^-$, and such that $u \equiv f$ on ∂D. Moreover, we can assume*

$$\sup_{S-D} u = \sup_{\partial D} f ,$$

$$\inf_{S-D} u = \inf_{\partial D} f .$$

Remark. Nothing is claimed about the uniqueness of u. As we shall see, u is unique for certain S and not unique for other S.

Proof: By Definition 2 of analytic disk, there exists a chart $\varphi: U \to W$ in the complete analytic atlas for S such that $\varphi(D)$ is a disk Δ. Since $\Delta^- \subset W$, there exists a concentric disk Δ_1 with $\Delta^- \subset \Delta_1$, $\Delta_1^- \subset W$.



Let $D_1 = \varphi^{-1}(\Delta_1)$. If $c \in \mathbb{R}$ then by Problem 8 there exists a unique function h_c which is continuous on $\Delta_1^- - \Delta$, harmonic in $\Delta_1 - \Delta^-$, and such that

$$h_c \equiv c \text{ on } \partial\Delta_1,$$
$$h_c \equiv f \circ \varphi^{-1} \text{ on } \partial\Delta.$$

Define a function v_c on $S-D$ by the formula

$$v_c = \begin{cases} h_c \circ \varphi & \text{in } D_1^- - D, \\ c & \text{in } S - D_1 \end{cases}.$$

Then v_c is continuous on $S-D$, $v_c \equiv f$ on ∂D, and v_c is harmonic in $S-D_1^-$ and in D_1-D^-. If $c \le \inf_{\partial D} f$, then v_c is subharmonic on $S-D^-$; for, the only points where we need to check the mean value criterion 3 of Proposition 10 are on ∂D_1, and there v_c takes the value c. But the minimum principle implies $h_c \ge c$ in $\Delta_1^- - \Delta$, and thus $v_c \ge c$ in $S-D$. Therefore, criterion 3 of Proposition 10 is trivially satisfied at a point of ∂D_1. Thus v_c is subharmonic in $S-D^-$. Likewise, if $c \ge \sup_{\partial D} f$, then v_c is superharmonic in $S-D^-$.

Let $A = \inf_{\partial D} f$, $B = \sup_{\partial D} f$. Then v_A is subharmonic in $S-D^-$, v_B is superharmonic in $S-D^-$, and $v_A \le v_B$ in $S-D$. This last inequality follows from the maximum principle, since $h_B - h_A$ is continuous in $\Delta_1^- - \Delta$, harmonic in $\Delta_1 - \Delta^-$, $\equiv B-A$ on $\partial\Delta_1$, $\equiv 0$ on $\partial\Delta$, and thus $h_B - h_A \ge 0$. Let u be GHM of v_B. Then since v_A is a subharmonic minorant of v_B, we have

$$v_A \le u \le v_B \quad , \tag{7}$$

and u is defined and harmonic on $S-D^-$. We have of course applied the corollary of Theorem 5 to the connected Riemann surface $S-D^-$, which is why u is defined only on $S-D^-$. But the inequalities (7) imply that u can be extended to a continuous function on $S-D$ in exactly one way, namely by taking $u \equiv v_A \equiv v_B \equiv f$ on ∂D.

Finally the last assertion of the proposition follows from

$$A \le v_A \le u < v_B \le B.$$

<u>QED</u>

<u>Remark</u>. The above analysis is typical in the sense that even when we wish to have boundary values for a certain harmonic function, the corollary to Theorem 5 does not by itself give anything more than a harmonic

function on an open set. Some other consideration, e.g. (7), is needed to obtain information about the function at the boundary. We shall see more instances of this phenomenon later.

To complete the preliminary material, we need to obtain a representation for harmonic functions in an annulus, analogous to the Laurent expansion of a holomorphic function.

PROPOSITION 14. *Let* u *be a real harmonic function in an annulus* $a < |z| < b$, *where* $0 \le a < b \le \infty$. *Then there exist unique complex numbers* c, $\{a_n\}$, *such that for* $a < |z| < b$

$$u(z) = c \log|z| + \operatorname{Re}\left(\sum_{-\infty}^{\infty} a_n z^n\right),$$

and a_0 *is real. Furthermore, if* $a < a' < b' < b$, *then there exist constants* K *and* $\{K_n\}$ *depending only on* a' *and* b' *(and* n*) such that*

$$|c| \le K \sup\{|u(z)| : a' \le |z| \le b'\},$$

$$|a_n| \le K_n \sup\{|u(z)| : a' \le |z| \le b'\}.$$

Proof: The discussion on p. 162 implies ∂u is holomorphic for $a < |z| < b$. Therefore, the Laurent expansion of ∂u exists, say

$$\partial u = \sum_{-\infty}^{\infty} c_n z^n, \qquad a < |z| < b.$$

By formula (3) of p. 165 , for $n \neq -1$ we have

$$c_n z^n = \frac{d}{dz} c_n \frac{z^{n+1}}{n+1} = 2\, \partial\, \mathrm{Re}\left(\frac{c_n z^{n+1}}{n+1}\right),$$

$$c_{-1} z^{-1} = c_{-1} \frac{d}{dz} \log z = c_{-1} 2 \partial \log|z|.$$

Now the Laurent expansion for ∂u converges uniformly on compact subsets of the annulus, and therefore the same is true for the integrated series, so we obtain

$$\partial u = 2c_{-1} \partial \log|z| + \sum_{n\neq -1} 2\partial\, \mathrm{Re}\left(\frac{c_n z^{n+1}}{n+1}\right)$$

$$= 2\partial(c_{-1} \log|z|) + \partial \sum_{n\neq 0} 2\mathrm{Re}\left(\frac{c_{n-1} z^n}{n}\right)$$

$$= \partial(c \log|z| + \mathrm{Re} \sum_{n\neq 0} a_n z^n),$$

where $c = 2c_{-1}$ and $a_n = \frac{2c_{n-1}}{n}$. The Cauchy-Riemann equation

$$\bar\partial(u - \bar c \log|z| - \mathrm{Re} \sum_{n\neq 0} a_n z^n) = 0$$

follows and shows there exists a function g holomorphic in $a < |z| < b$ such that

$$u = \bar c \log|z| + \mathrm{Re} \sum_{n\neq 0} a_n z^n + g(z).$$

Taking imaginary parts,

$$\mathrm{Im}\, g = (\mathrm{Im}\, c) \log|z|.$$

Since $i \log z = -\arg \dot z + i \log |z|$, this shows that $g = i(\mathrm{Im}\, c)\log z$, and thus g is defined only if $\mathrm{Im}\, c = 0$. Then g is a <u>real</u> holomorphic function, and thus is constant. Thus, if $g \equiv a_0$, we have

$$u(z) = c \log|z| + \operatorname{Re} \sum_{-\infty}^{\infty} a_n z^n, \quad c,\ a_0 \text{ real},$$

a representation of the desired form.

Now we obtain the uniqueness: if $a < r < b$, then

$$\begin{aligned} u(re^{i\theta}) &= c \log r + \frac{1}{2} \sum_{-\infty}^{\infty} a_n r^n e^{in\theta} + \frac{1}{2} \sum_{-\infty}^{\infty} \bar{a}_n r^n e^{-in\theta} \\ &= c \log r + \frac{1}{2} \sum_{-\infty}^{\infty} (a_n r^n + \overline{a_{-n}} r^{-n}) e^{in\theta}. \end{aligned}$$

Since this series converges uniformly for $0 \le \theta \le 2\pi$, we obtain by integration

$$(8) \quad \begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} u(re^{i\theta})\, d\theta &= c \log r + a_0, \\ \frac{1}{\pi} \int_0^{2\pi} u(re^{i\theta}) e^{-im\theta}\, d\theta &= a_m r^m + \overline{a_{-m}}\, r^{-m}, \quad m \ne 0. \end{aligned}$$

Here we have used the orthogonality relation

$$\frac{1}{2\pi} \int_0^{2\pi} e^{in\theta} e^{-im\theta}\, d\theta = \begin{cases} 0 & \text{if } m \ne n, \\ 1 & \text{if } m = n. \end{cases}$$

If we use the relations (8) for two different values $r, r \in (a,b)$, we can solve for all coefficients:

$$c = \frac{1}{\log r - \log r'} \frac{1}{2\pi} \int_0^{2\pi} [u(re^{i\theta}) - u(r'e^{i\theta})]\, d\theta,$$

$$a_0 = \frac{1}{\log r - \log r'} \frac{1}{2\pi} \int_0^{2\pi} [u(r'e^{i\theta}) \log r - u(re^{i\theta}) \log r']\, d\theta,$$

$$a_m = \frac{1}{\left(\frac{r}{r'}\right)^m - \left(\frac{r'}{r}\right)^m} \frac{1}{\pi} \int_0^{2\pi} [u(re^{i\theta}) r'^{-m} - u(r'e^{i\theta}) r^{-m}] e^{-im\theta}\, d\theta,$$

$m \neq 0$. This proves that the coefficients are uniquely determined by u, and at the same time shows easily how to obtain the estimates stated in the last half of the proposition.

QED

COROLLARY. "Removable singularity theorem" *Let u be harmonic and bounded in an annulus* $0 < |z| < b$. *Then there exists a harmonic function in the disk* $|z| < b$ *which agrees with* u *in the annulus* $0 < |z| < b$.

Proof: By Proposition 14 we have

$$u(z) = c \log |z| + \mathrm{Re}\left(\sum_{-\infty}^{\infty} a_n z^n\right), \quad 0 < |z| < b.$$

In formula (8) we let $r \to 0$ and we read off the relations

$$c \log r \text{ is bounded,}$$

$$a_m r^m + \overline{a_{-m}}\, r^{-m} \text{ is bounded, } m \neq 0.$$

Therefore, $c = 0$ and $m < 0$ implies $a_m = 0$. Thus the expansion for u reads

$$u(z) = \mathrm{Re}\left(\sum_{0}^{\infty} a_n z^n\right), \quad 0 < |z| < b.$$

The right side of this expression is harmonic in the disk $|z| < b$.

QED

COROLLARY. *If* u *is harmonic and bounded for* $a < |z| < \infty$ *and continuous for* $a \leq |z| < \infty$, *and if*

$u(z) \equiv 0$ for $|z| = a$, then $u \equiv 0$.

Proof: In formula (8) let $r \to a$ to obtain $c \log a + a_0 = 0$, $a_m a^m + \overline{a_{-m}}\, a^{-m} = 0$. By the reasoning given in the previous corollary, $a_m = 0$ for $m > 0$ and $c = 0$. Therefore, $a_0 = 0$ and $a_m = 0$ for $m < 0$. Thus, $u \equiv 0$.

QED

We are now almost ready to prove Theorem 4. But something rather strange will arise in the proof. Namely, we shall see that there is a certain dichotomy of Riemann surfaces which requires that the proof of Theorem 4 be quite dependent on this classification, although the statement of the theorem is the same in both cases. We present this phenomenon in the form of a proposition.

PROPOSITION 15. The following conditions on a connected Riemann surface S are equivalent.

1. Every bounded subharmonic function on S is constant.
2. If D is any analytic disk and u is a bounded continuous nonnegative function in S-D which is harmonic in $S-D^-$ and which vanishes identically on ∂D, then $u \equiv 0$.

3. <u>If</u> D <u>is any analytic disk and</u> u <u>is a bounded continuous function in</u> S-D <u>which is harmonic in</u> $S\text{-}D^-$, <u>then</u>

$$\sup_{S-D} u = \sup_{\partial D} u.$$

4. <u>Same as</u> 3 <u>with "harmonic" replaced by "subharmonic</u>."

2′. <u>Condition</u> 2 <u>holds for some analytic disk</u> D.

3′. <u>Condition</u> 3 <u>holds for some analytic disk</u> D.

4′. <u>Condition</u> 4 <u>holds for some analytic disk</u> D.

<u>Proof</u>: We shall prove $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4$ and $2' \Rightarrow 3' \Rightarrow 4' \Rightarrow 1$. Since the assertions $2 \Rightarrow 2'$, $3 \Rightarrow 3'$, and $4 \Rightarrow 4'$ are trivial, the proposition will follow. The proof that $2 \Rightarrow 3$ is identical to the proof that $2' \Rightarrow 3'$ and likewise for $3 \Rightarrow 4$ and $3' \Rightarrow 4'$.

<u>$3 \Rightarrow 4$</u>: As in the proof of Proposition 13, we choose a "concentric" analytic disk D_1 with $D^- \subset D_1$. Suppose u is a bounded continuous function in S-D which is subharmonic in $S\text{-}D^-$. Choose a constant C such that $\sup_{S-D} u \leq C$.

Let w be the unique function which is continuous in S-D, harmonic in $D_1\text{-}D^-$, such that

$$w = u \text{ on } \partial D,$$

$$w = C \text{ in } S\text{-}D_1.$$

Then w is superharmonic in $S\text{-}D^-$ and criterion 2 of Proposition 10 implies $u \le w$ in $D_1^-\text{-}D$ and therefore $u \le w$ in $S\text{-}D$. Let v = GHM of w (cf. p. 197). Then v is harmonic in $S\text{-}D^-$ and $u \le v \le w$. Therefore, we can naturally extend v to be continuous in S-D by setting $v = u = w$ on ∂D. Condition 3 applies to v and thus

$$\sup_{S-D} u \le \sup_{S-D} v = \sup_{\partial D} v = \sup_{\partial D} u.$$

<u>4′ ⇒ 1</u>: Suppose u is a bounded subharmonic function on S. Let D be an analtyic disk on S for which condition 4 holds. Then

$$\sup_{S-D} u = \sup_{\partial D} u.$$

Therefore,

$$\sup_{S} u = \sup_{D^-} u ,$$

and since D^- is compact, we see that u assumes its maximum. By the strong maximum principle, $u \equiv$ constant.

<u>1 ⇒ 2</u>: Let u be the function in the hypothesis of 2. Define v on S by

$$v = u \text{ in } S\text{-}D,$$

$$v = 0 \text{ in } D.$$

Then v is subharmonic and bounded on S, so 1 implies $v \equiv$ constant. Thus, $v \equiv 0$ and it follows that $u \equiv 0$.

<u>$2 \Rightarrow 3$</u> Let u be the function in the hypothesis of 3. Define

$$A = \inf_{S-D} u,\ B = \sup_{S-D} u.\ C = \sup_{\partial D} u.$$

Then $A \le C \le B$ and we want to prove $B = C$. As in the proof of Proposition 13 and also the current proof that $3 \Rightarrow 4$ we take a disk D_1 and define v_A to be continuous in $S-D$, harmonic in D_1-D^-, such that

$$v_A = u \text{ on } \partial D,$$

$$v_A = A \text{ in } S-D_1.$$

We define v_B the same way with A replaced by B. Then v_A is subharmonic, v_B is superharmonic in $S-D^-$, and $v_A \le C$, $v_A \le u$, $u \le v_B$, all of which follow from the maximum principle. Let

$$w_1 = \text{GHM of } \min(u,C),$$

$$w_2 = \text{GHM of } v_B.$$

Note that $\min(u,C)$ is superharmonic by Proposition 12. Then the inequalities we have obtained for v_A and v_B show

$$v_A \le w_1 \le u \le w_2 \le v_B.$$

Thus, w_1 and w_2 have continuous extensions to $S-D$ with $w_1 = w_2 = u$ on ∂D. Therefore, w_2-w_1 satisfies the hypothesis of 2, the required boundedness following from $w_2-w_1 \le v_B-v_A \le B-A$. Thus, condition 2 implies $w_2-w_1 \equiv 0$. Thus

$$B = \sup_{S-D} u = \sup_{S-D} w_1 \le C.$$

QED

DEFINITION 6. A noncompact connected Riemann surface satisfying the conditions of Proposition 15 is a parabolic Riemann surface. A noncompact connected Riemann surface not satisfying these condition is hyperbolic.

Examples.

1. If S is compact and connected, S satisfies the conditions of Proposition 15. For suppose u is a bounded subharmonic function on S. Then u assumes its maximum, so the strong maximum principle implies u is constant.

2. $\mathbb{C}$ is parabolic. We verify condition 2′ for $D = \{z: |z| < 1\}$. Suppose u is a function satisfying the hypothesis of Proposition 15, criterion 2′. By the second corollary on p. 203, $u \equiv 0$.

3. If $S = \{z: |z| < 1\}$ has its usual complete

analytic atlas, then S is hyperbolic. This is obvious, a nonconstant subharmonic function which is bounded on S being, for example, $z \to \mathrm{Re}\, z$.

Finally, the stage is set for the proof of Theorem 4. In the statement of the theorem, on p. 168, there is a given chart $\varphi: U \to W$ in the complete analytic atlas for S, where U contains the given point p and $\varphi(p) = 0$. By a simple change of variable, we can assume that

$$\{z: |z| \leq 2\} \subset W.$$

Let $\Delta_r = \{z: |z| < r\}$ and $D_r = \varphi^{-1}(\Delta_r)$ for $0 < r \leq 2$.

Proof of Theorem 4 in case S does not satisfy the conditions of Proposition 15: By criterion 2′, there exists a bounded continuous nonnegative function v in $S-D_1$ which is harmonic in $S-\bar{D}_1$ and which is identically zero on ∂D_1, and yet $v \not\equiv 0$. The strong maximum principle implies that $v > 0$ in $S-\bar{D}_1$. The function φ is holomorphic on U, and therefore $\mathrm{Re}(\varphi^{-n}-\varphi^{n})$ is harmonic on $U-\{p\}$ and in particular is harmonic on D_2. On ∂D_1, $|\varphi| = 1$ so that $\varphi^{-n}=\bar{\varphi}^{n}$ and thus $\varphi^{-n}-\varphi^{n}$ is purely imaginary and thus $\mathrm{Re}(\varphi^{-n}-\varphi^{n}) = 0$. Since $v > 0$ on the compact set ∂D_2, it is bounded below by a positive constant there. Therefore, there exists a constant C such that

$$|\mathrm{Re}(\varphi^{-n}-\varphi^{n})| \le Cv \text{ on } \partial D_2.$$

The same inequality holds trivially on ∂D_1, both sides vanishing, and therefore the weak maximum principle implies

(9) $$|\mathrm{Re}(\varphi^{-n}-\varphi^{n})| \le Cv \text{ in } \bar{D}_2 - D_1.$$

Now we define

$$w_1 = \begin{cases} -Cv \text{ in } S-D_1, \\ \mathrm{Re}(\varphi^{-n}-\varphi^{n}) \text{ in } D_1 - \{p\}, \end{cases}$$

$$w_2 = \begin{cases} Cv \text{ in } S-D_1, \\ \mathrm{Re}(\varphi^{-n}-\varphi^{n}) \text{ in } D_1 - \{p\}. \end{cases}$$

Then w_1 and w_2 are clearly continuous on $S - \{p\}$ and (9) implies w_1 is superharmonic and w_2 is subharmonic in $S - \{p\}$: it suffices to check the mean value property of Proposition 10.3 at points on ∂D_1 and at such a point $w_1 = 0 = \mathrm{Re}(\varphi^{-n}-\varphi^{n})$ = the mean value of $\mathrm{Re}(\varphi^{-n}-\varphi^{n})$ on small circles centered at the point $\ge$ the corresponding mean value of w_1 (since $\mathrm{Re}(\varphi^{-n}-\varphi^{n}) \ge -Cv$ on the part of the circle lying outside D_1). Thus, w_1 is superharmonic, and a similar proof shows w_2 is subharmonic.

Choose a constant $A \ge 2C \sup_{S-D_1} v$. Let u = GHM of

$w_1 + A$. Note that trivially $w_2 - w_1 \le A$, and therefore

$$w_2 \le u \le w_1 + A.$$

We have of course used here the existence of GHM on the Riemann surface $S - \{p\}$. Therefore, u is harmonic on $S - \{p\}$ and our inequalities show

$$0 \le u - \mathrm{Re}(\varphi^{-n} - \varphi^{n}) \le A \text{ in } D_1 - \{p\}.$$

Therefore, the function $u - \mathrm{Re}(\varphi^{-n} - \varphi^{n})$ is harmonic and <u>bounded</u> in $D_1 - \{p\}$, so the removable singularity theorem of p. 203 shows that there is a harmonic function h in D_1 such that

$$u = \mathrm{Re}(\varphi^{-n}) + h \text{ in } D_1.$$

Since $h = \mathrm{Re}F$ for some holomorphic function F in D_1 (Proposition 1.4), we have proved Theorem 4 in this case, and we can even assert that no term $\log|\varphi|$ appears in the representation for u.

<u>Proof of Theorem 4 in case S does satisfy the conditions of Proposition 15</u>: By Proposition 13, there exists for $0 < r < 1$ a bounded continuous function u_r in $S - D_r$ such that $u_r = \mathrm{Re}(\varphi^{-n})$ on ∂D_r and u_r is harmonic in $S - \bar{D}_r$. If u_r is constant on ∂D_1, then Proposition 15.3 implies u_r is constant in $S - D_1$ (apply the criterion 3 both to u_r and $-u_r$) and thus u_r is constant in $S - D_r$ by Proposition 3, which is not true. Thus u_r is not constant on ∂D_1, and therefore there exist unique coefficients α_r and β_r such that if $v_r = \alpha_r u_r + \beta_r$, then

$$\max_{\partial D_1} v_r = 1, \quad \min_{\partial D_1} v_r = 0.$$

Proposition 15.3 implies $0 \le v_r \le 1$ in $S-D_1$.

By Proposition 8 there exists a sequence $r_k \to 0$ such that v_{r_k} converges uniformly on compact subsets of $D_2-\bar{D}_1$. Moreover, Proposition 15.3 applied to $D_{3/2}$ implies that v_{r_k} converges uniformly on $S-D_{3/2}$. If $v = \lim_{k\to\infty} v_{r_k}$, then Proposition 7 implies v is harmonic in $S-\bar{D}_1$.

We now write down the Laurent expansion of Proposition 14 for v_r in the set $D_2-\bar{D}_r$:

$$v_r \circ \varphi^{-1}(z) = c(r)\log|z| + \operatorname{Re}\sum_{-\infty}^{\infty} a_j(r)z^j,$$

$$r < |z| < 2,$$

where $c(r)$ and $a_0(r)$ are real, and formula (8) of p. 202 shows

$$(10)\quad \begin{aligned} c(r)\log s + a_0(r) &= \frac{1}{2\pi}\int_0^{2\pi} v_r \circ \varphi^{-1}(se^{i\theta})\,d\theta, \\ a_j(r)s^j + \overline{a_{-j}(r)}\, s^{-j} &= \frac{1}{\pi}\int_0^{2\pi} v_r \circ \varphi^{-1}(se^{i\theta})e^{-ij\theta}\,d\theta, \end{aligned}$$

$$j \ne 0,$$

for $r < s < 2$.

Now $v_r \circ \varphi^{-1}(re^{i\theta}) = \alpha_r \mathrm{Re}(r^{-n}e^{-in\theta}) + \beta_r$

$$= \alpha_r r^{-n} \frac{e^{in\theta}+e^{-in\theta}}{2} + \beta_r,$$

and therefore if we let $s \to r$ in the second part of (10) we obtain

$$a_j(r)r^j + \overline{a_{-j}(r)}\, r^{-j} = 0 \text{ if } j \geq 1,\ j \neq n.$$

Taking $s = 1$ in (10),

$$|a_j(r) + \overline{a_{-j}(r)}| \leq 2 \text{ if } j \geq 1.$$

Therefore, for $j \geq 1$ and $j \neq n$,

$$2 \geq |a_j(r) + \overline{a_{-j}(r)}| = |a_j(r) - a_j(r)r^{2j}|$$

$$= |a_j(r)|(1-r^{2j}) \geq \tfrac{1}{2}\, |a_j(r)| \text{ if } 0 < r < \tfrac{1}{2};$$

thus, $|a_j(r)| \leq 4$ and therefore

$$|a_{-j}(r)| \leq 4r^{2j}.$$

Now the estimates in Proposition 14 imply that as $r_k \to 0$, the coefficients in the Laurent expansion for v_{r_k} converge to the coefficients in the Laurent expansion for v. Therefore,

$$v \circ \varphi^{-1}(z) = c \log|z| + \mathrm{Re}(a_{-n}z^{-n} + \sum_0^\infty a_j z^j),$$

$$1 < |z| < 2,$$

since $a_{-j}(r_k) \to 0$ for $j \geq 1$, $j \neq n$. Now we define u on $S - \{p\}$ by

$$u = v \text{ in } S\text{-}\bar{D}_1 ,$$

$$u\circ\varphi^{-1}(z) = c \log |z| + \mathrm{Re}(a_{-n}z^{-n} + \sum_0^\infty a_j z^j),$$

$$0 < |z| < 2.$$

It is clear that u is harmonic on $S - \{p\}$. The theorem will be proved once we establish that $a_{-n} \neq 0$. This involves a rather delicate argument.

Suppose that $a_{-n} = 0$. Then the formula for u near p shows that there exists

$$\ell = \lim_{q\to p} u(q),$$

where $-\infty \leq \ell \leq \infty$. By Proposition 15.3 applied to u and -u and small analytic disks containing p, we conclude that $u \equiv \ell$ on $S - \{p\}$; therefore, $-\infty < \ell < \infty$. Now we shall prove that $v_{r_k} \to u$ uniformly on ∂D_1, and therefore that $\max_{\partial D_1} u = 1$, $\min_{\partial D_1} u = 0$, contradicting the fact that u is constant. Again, Proposition 15.3 shows that it is sufficient to prove that $v_{r_k} \to u$ uniformly on $\partial D_{\frac{1}{2}}$. For $|z| = \frac{1}{2}$ and $0 < r < \frac{1}{2}$,

$$|v_r\circ\varphi^{-1}(z) - u\circ\varphi^{-1}(z)| \leq |c(r)-c| \log 2 + |a_{-n}(r)-a_{-n}| 2^n$$

$$+ \sum_0^N |a_j(r)-a_j| + \sum_{N+1}^\infty (4 + 4)2^{-j}$$

$$+ \sum_{\substack{j<0 \\ j\neq -n}} 4r^{-2j}2^{-j}$$

$$\leq C_n[|c(r)-c| + |a_{-n}(r)-a_{-n}| + \sum_0^N |a_j(r)-a_j|]$$

$$+ 8\cdot 2^{-N} + \frac{8r^2}{1-2r^2} \quad .$$

Therefore, if $\varepsilon > 0$ we can choose a fixed N such that $8\cdot 2^{-N} < \frac{\varepsilon}{3}$ and a k_0 such that $\frac{8r_k^2}{1-2r_k^2} < \frac{\varepsilon}{3}$ and $r_k < \frac{1}{2}$ for $k \geq k_0$. Then we choose $k_1 \geq k_0$ such that

$$C_n[|c(r_k)-c| + |a_{-n}(r_k)-a_n| + \sum_0^N |a_j(r_k)-a_j|] < \frac{\varepsilon}{3}$$

if $k \geq k_1$. Therefore, if $k \geq k_1$,

$$|v_r - u| < \varepsilon \quad \text{on} \quad \partial D_{\frac{1}{2}}.$$

<u>QED</u>

Thus, we have completed the proof of Theorem 4. We have already indicated the use of this theorem in establishing the existence of meromorphic functions, shown on pp. 168-171 in the proof of Theorem 3. In the next section we shall give further applications.

<u>Remark</u>. In the above proof in the second case, the second part of (10) in the case j = n was not used. But we get additional information by using this formula:

$$a_n(r)s^n + \overline{a_{-n}(r)}\ s^{-n} = \frac{1}{\pi}\int_0^{2\pi} \psi_r\circ\varphi^{-1}(se^{i\theta})e^{-in\theta}d\theta.$$

Letting $s = r$, we obtain

$$a_n(r)r^n + \overline{a_{-n}(r)}\ r^{-n} = \alpha_r r^{-n}\ \frac{1}{2\pi}\int_0^{2\pi}(e^{in\theta}+e^{-in\theta})e^{-in\theta}d\theta$$

$$+\ \beta_r\ \frac{1}{\pi}\int_0^{2\pi} e^{-in\theta}d\theta$$

$$= \alpha_r r^{-n}\ .$$

We already know from $s = 1$ that

$$|a_n(r) + \overline{a_{-n}(r)}| \le 2.$$

Therefore,

$$|\alpha_r - \overline{a_{-n}(r)} + \overline{a_{-n}(r)}\ r^{2n}| \le 2r^{2n}.$$

Taking imaginary parts, we conclude

$$|\mathrm{Im}\ a_{-n}(r)|\ (1-r^{2n}) \le 2r^{2n},$$

showing that $\mathrm{Im}\ a_{-n}(r) \to 0$ as $r \to 0$. Therefore, letting $r = r_k$, we have

$$a_{-n} = \lim_{k\to\infty} a_{-n}(r_k) \text{ is } \underline{\text{real}}.$$

We obtain from this remark the following result:

<u>COROLLARY TO THEOREM 4</u>. <u>Let</u> S <u>be any connected Riemann surface and let</u> $p \in S$. <u>Let</u> $\varphi: U \to W$ <u>be a chart in the complete analytic atlas for</u> S <u>with</u> $p \in U$ <u>and</u> $\varphi(p)=0$.

Let n be a positive integer. Then there exists a harmonic function u on $S - \{p\}$ such that for z near 0

$$u \circ \varphi^{-1}(z) = \mathrm{Re}\left(\frac{\alpha}{z^n}\right) + \mathrm{Re} f(z),$$

where α is a nonzero complex number and f is holomorphic in a neighborhood of 0. Moreover, u can be taken to be bounded outside a neighborhood of p.

Proof: If S is hyperbolic, our proof on pp. 209-211 already contained this result with $\alpha = 1$. The boundedness of u away from p has also been shown in this case.

If S is compact or parabolic, the assertion about the boundedness of u is automatic. What we must do is eliminate the term involving $\log |z|$. By the previous remark, we have obtained a harmonic function v on $S - \{p\}$ such that near 0

$$v \circ \varphi^{-1}(z) = c \log |z| + \mathrm{Re}\left(\frac{a}{z^n}\right) + \mathrm{Re}\, g(z),$$

where g is holomorphic near 0, c is real, and $a \neq 0$ is real. If $c = 0$, we are through. Otherwise, we replace φ by the chart $\omega\varphi$, where ω is a fixed complex number with $\omega^n = i$. Applying our result in this case, we obtain a harmonic function w on $S - \{p\}$ such that near 0

$$w \circ \varphi^{-1}(z) = d \log|z| + \mathrm{Re}\left(\frac{b}{iz^n}\right) + \mathrm{Re} h(z),$$

where h is holomorphic near 0, d is real, and $b \neq 0$ is real. We have left out a trivial intermediate calculation here. Now define

$$u = w - \frac{d}{c} v.$$

Then u is harmonic on $S - \{p\}$, and near 0

$$u \circ \varphi^{-1}(z) = \mathrm{Re}\ (\frac{\alpha}{z^n}) + \mathrm{Re}(h(z) - \frac{d}{c} g(z)),$$

where

$$\alpha = \frac{b}{i} - \frac{da}{c} \neq 0.$$

QED

Also in the next section we shall require the existence of a Green's function on a parabolic Riemann surface.

DEFINITION 7. Let S be a connected Riemann surface and $p \in S$. A function g defined on $S - \{p\}$ is a *Green's function* if

1. g is positive and harmonic in $S-\{p\}$;
2. if φ is an analtyic chart near p with $\varphi(p) = 0$, then $g + \log|\varphi|$ is harmonic in a neighborhood of p;
3. if h has properties 1 and 2, then $g \leq h$.

We first remark that condition 2 is independent of the particular chart φ, since any other analtyic chart ψ can be expressed as

$$\psi = a\varphi(1 + \sum_1^\infty \alpha_k \varphi^k),$$

and so

$$\log|\psi| = \log|\varphi| + \log|a| + \log|1 + \sum_1^\infty \alpha_k \varphi^k|$$

and we see that $\log|\psi| - \log|\varphi|$ is harmonic in a neighborhood of p.

PROPOSITION 16. *Let S be a connected hyperbolic Riemann surface and $p \in S$. Then there exists a unique Green's function on $S - \{p\}$.*

Proof: Uniqueness is clear by property 3 of a Green's function. The proof of existence is like the proof of Theorem 4 in the hyperbolic case. We set the problem in the framework of all the notation on the top of p. 209. Thus, v is a bounded continuous function on $S-D_1$, $v > 0$ and v is harmonic in $S-\bar{D}_1$, and $v \equiv 0$ on ∂D_1. As before, there exists a constant $C > 0$ such that

$$\log|\varphi| \leq Cv \text{ in } \bar{D}_2 - D_1.$$

As before, the function

$$w_1 = \begin{cases} -Cv \text{ in } S-D_1, \\ -\log|\varphi| \text{ in } D_1, \end{cases}$$

is superharmonic in $S-\{p\}$. Much more trivially, the function

$$w_2 = \begin{cases} 0 \text{ in } S-D_1, \\ -\log|\varphi| \text{ in } D_1, \end{cases}$$

is subharmonic in $S-\{p\}$. Let g be the <u>least harmonic majorant</u> of w_2. As in Definition 5, g is harmonic on $S - \{p\}$ and if $A \geq C \sup_{S-D_1} v$, then $w_2 \leq w_1 + A$, so that

$$w_2 \leq g \leq w_1 + A.$$

By the removable singularity theorem, $g + \log|\varphi|$ is harmonic in D_1, so property 2 follows. Also since $w_2 \geq 0$, also $g \geq 0$ and since g is not constant, the strong maximum principle implies property 1. To check property 3, suppose h has properties 1 and 2. Then $h + \log|\varphi|$ is harmonic in D_1 and is positive on ∂D_1, so the minimum principle for harmonic functions implies $h + \log|\varphi| > 0$ in D_1. Therefore, $h > w_2$ on $S - \{p\}$, and the definition of g therefore implies $g \leq h$.

QED

<u>Remark</u>. We can prove even more. Namely, if h is positive and superharmonic on $S - \{p\}$ and $h + \log|\varphi|$ is superharmonic near p, then $g \leq h$. It's exactly the same proof.

<u>Problem 9</u>. Find the Green's function for the unit disk $\{z: |z| < 1\}$.

Chapter VII

CLASSIFICATION OF SIMPLY CONNECTED RIEMANN SURFACES

As an application of the results of the previous chapter, we are going to prove that every simply connected Riemann surface is analytically equivalent to the sphere $\hat{\mathbb{C}}$, the complex plane $\hat{\mathbb{C}}$, or the unit disk $\{z\colon |z| < 1\} \subset \mathbb{C}$. These cases are exclusive, of course, since the compactness of the sphere shows it is not even homeomorphic to the plane or disk; and the plane and disk, though homeomorphic, are not analytically equivalent (Liouville's theorem) (see p. 42).

We shall require some slight generalizations of some of the basic results of Chapter III. Namely, we shall require a *permanence of functional relations* generalizing that of p. 66, and a *monodromy theorem* generalizing that of p. 64. In addition, we shall require a generalization of Lemma 2 on p. 117 which deals with unrestricted analytic continuation.

The framework for this discussion has just been mentioned - the analytic continuation of meromorphic functions defined on arbitrary Riemann surfaces, rather than $\mathbb{C}$. Given a Riemann surface S, we can form definitions as at the beginning of Chapter III and speak of M_S, the *sheaf of germs of meromorphic functions on* S. All the material of pp. 46-64 can be discussed with very little change. The applications we have in mind are given in the next two lemmas.

LEMMA 1. Let $p \in S$, a simply connnected Riemann surface. Let u be harmonic on $S - \{p\}$ such that if φ is an analytic chart near p with $\varphi(p) = 0$, then

$$u \circ \varphi^{-1}(z) = \mathrm{Re}\left(\sum_{k=N}^{\infty} a_k z^k\right), \ z \text{ near } 0,$$

where $a_N \neq 0$. Here $-\infty < N < \infty$. Then there exists a meromorphic function f on S such that

(1) $\qquad \mathrm{Re}(f) \equiv u.$

Proof: It is obvious that we may define f near p by setting

$$f \circ \varphi^{-1}(z) = \sum_{k=N}^{\infty} a_k z^k, \ z \text{ near } 0.$$

It is now a question of continuing f analytically to all of S. The generalized principle of the permanence of functional relations implies that the analytic continuation will always satisfy the identity (1). Briefly, the reason is that if f is meromorphic in an analytic disk D and $\mathrm{Re}(f) = u$ in a neighborhood of some point of D, then $\mathrm{Re}(f) = u$ holds throughout D (see Proposition 3 of p. 167).

The second point is that analytic continuation is possible along every path in S with initial point p. The reason is that Proposition 1.4 of p. 165 shows that (1) holds locally; this and the permanence of functional relations combine as in the proof of Lemma 2 on p. 117 to show that the process of analytic continuation never "stops."

Now we have the hypothesis needed to apply the monodromy theorem, and the lemma is proved.

QED

LEMMA 2. *Let* $p \in S$, *a simply connected Riemann surface. Let* u *be harmonic on* $S - \{p\}$ *such that if* φ *is an analytic chart near* p *with* $\varphi(p) = 0$, *then*

$$u \circ \varphi^{-1}(z) = \log |z| + \operatorname{Re}\left(\sum_{k=0}^{\infty} a_k z^k\right),$$

$$z \text{ near } 0.$$

Then there exists a holomorphic function f *on* S *such that*

$$|f| \equiv e^u . \tag{2}$$

Proof: The outline of the proof is the same as in the previous lemma. First, we prove that f exists near p. Using $e^{\log|z|} = |z|$, we naturally choose

$$f \circ \varphi^{-1}(z) = z \exp\left(\sum_{k=0}^{\infty} a_k z^k\right), \ z \text{ near } 0.$$

Then (2) obviously holds near p. Second, we apply the permanence of functional relations to show that (2) remains valid under analytic continuation of f. The point to be checked is that if f is holomorphic in an analytic disk and (2) holds in a neighborhood of some point, then (2) holds throughout the disk. This follows as before since log $|f|$ is harmonic. One might think there is trouble here at zeros of f; by (2), however, if f has a zero along some path of analytic continuation, then (2) will have been violated before the zero is reached.

Third, (2) holds locally at least. For locally we can write u = ReF, F holomorphic (we are not now treating neighborhoods of the exceptional point p). Then we set $f = e^F$, implying (2). Therefore, as before, analytic continuation is possible along every path from p. We are also using in this step the fact that (2) determines f locally essentially uniquely. That is, any other choice of f is just f multiplied by a constant of modulus 1, since holomorphic functions with constant modulus must be constant. (A similar fact about (1) was used implicitly in the proof of Lemma 1; in that case functions satisfying (1) have constant differences.)

Fourth, the monodromy theorem finishes the proof.

QED

Next, a technicality.

LEMMA 3. *Let E be any bounded nonempty set in C. Then there exist complex numbers α and β, $\alpha \neq 0$, such that if*

$$\tilde{E} = \{\alpha z + \beta : z \in E\},$$

then

$$\sup\{|w| : w \in \tilde{E}\} = 1,$$
$$\inf\{|w| : w \in \tilde{E}\} = \tfrac{1}{2}.$$

Proof: Let $a = \inf\{\operatorname{Re} z : z \in E\}$ and choose b such that $a + ib \in E^-$ (using the boundedness of E). Let $E_1 = \{z-a-ib : z \in E\}$, so that $\inf\{\operatorname{Re} z : z \in E_1\} = 0$, $0 \in E_1^-$. Define for $t \geq 0$

$$m(t) = \inf\{|z+t| : z \in E_1\},$$

$$M(t) = \sup\{|z+t| : z \in E_1\}.$$

Then m and M are continuous increasing functions, $m(0) = 0$, and the boundedness of E_1 implies $\frac{m}{M} \to 1$ as $t \to \infty$. Choose t such that $\frac{m(t)}{M(t)} = \frac{1}{2}$. Let $c = M(t)$ and

$$\tilde{E} = \{\tfrac{z+t}{c} : z \in E_1\}.$$

Then $\tilde{E}$ satisfies the conditions of the lemma, and

$$\alpha = \frac{1}{c}, \quad \beta = \frac{t-a-ib}{c}.$$

QED

Classification Theorem. Any connected, simply connected Riemann surface is analytically equivalent to the Riemann sphere, the complex plane, or the unit disk.

Proof: Let S be the connected, simply connected Riemann surface. We have three cases to consider.

S is compact: By the corollary to Theorem 4 on p. 216, if $p \in S$, then there exists a harmonic function u on $S - \{p\}$ such that in terms of a given analytic chart φ near p with $\varphi(p) = 0$,

$$u \circ \varphi^{-1}(z) = \operatorname{Re}\left(\frac{\alpha}{z}\right) + \operatorname{Re}F(z), \quad z \text{ near } 0, \quad \alpha \neq 0,$$

where F is holomorphic near 0. By Lemma 1 there exists a meromorphic function f on S such that

$$\operatorname{Re}(f) \equiv u.$$

Now the only pole of f is the point p, and this is a pole of order 1. Thus, f takes the value ∞ exactly one time. By Proposition 9.1 of p. 44, f takes every value in $\hat{\mathbb{C}}$ exactly one time. That is, $f: S \to \hat{\mathbb{C}}$ is an analytic equivalence between S and $\hat{\mathbb{C}}$, proving the result in this case.

S is parabolic: If $p \in S$, and φ is an analytic chart in a neighborhood of p, then by the corollary to Theorem 4 on p. 216, there exists a harmonic function u on $S - \{p\}$ such that

$$u \circ \varphi^{-1}(z) = \operatorname{Re}\left(\frac{\alpha}{z}\right) + \operatorname{Re}F(z), \quad z \text{ near } 0, \quad \alpha \neq 0,$$

where F is holomorphic near 0. We are assuming $\varphi(p) = 0$. The construction of u shows that u is <u>bounded</u> outside any neighborhood of p. For this see p. 212, where $0 \leq v \leq 1$ in $S-D_1$; p. 214 showing that $u = v$ in $S-D_1$; and p. 218, showing that our function u is a linear combination of two functions bounded in $S-D_1$. Note that we are tacitly assuming that φ is rescaled if necessary to guarantee the existence of D_2, the analytic disk given by $|\varphi| < 2$.

By Lemma 1 there exists a meromorphic function f on S such that

$$\mathrm{Re}(f) \equiv u.$$

Note that f has a pole only at p, that p is a simple pole, and that Ref is bounded outside any neighborhood of p. We wish to obtain another function with the stronger property that $|f|$ is bounded outside any neighborhood of p.

To do this let $a_1 < a_2 < a_3 < \dots$ be a sequence of positive integers. Since these are real numbers tending to ∞ and the real part of f is bounded outside any neighborhood of p, and since f is one-to-one in a neighborhood of p, it follows that for sufficiently large n there exists a unique $p_n \in S$ such that $f(p_n) = a_n$. By eliminating the first few terms in the sequence, we can assume this holds for all n. Also, it is clear that $p_n \to p$. Since Ref is bounded outside any neighborhood U of p, we obtain for sufficiently large n

$$|f-a_n| \geq a_n - \mathrm{Re} f \geq \tfrac{1}{2} a_n \text{ outside } U,$$

and therefore since we can assume f is one-to-one in U, $\frac{1}{f-a_n}$ has a simple pole exactly at p_n and is bounded outside any neighborhood of p_n. By Lemma 3 there exist constants α_n and β_n such that if

$$f_n = \frac{\alpha_n}{f-a_n} + \beta_n,$$

then

$$\sup_{\bar{D}_2 - D_1} |f_n| = 1,$$

$$\inf_{\bar{D}_2 - D_1} |f_n| = \tfrac{1}{2} .$$

Since S is parabolic, Proposition 15.4 on p. 205 implies

$$\sup_{S-D_1} |f_n| = 1$$

($|f_n|$ is subharmonic).

By Proposition 8 on p. 178, there exists a subsequence $n_1 < n_2 < \ldots$ such that $\lim_{k\to\infty} f_{n_k}$ exists uniformly on compact subsets of $D_2 - \bar{D}_1$, and then Proposition 15.4 on p. 205 again implies $\lim_{k\to\infty} f_{n_k}$ exists uniformly on $S - D_{3/2}$. Let

$$h = \lim_{k\to\infty} f_{n_k} \text{ in } S - \bar{D}_1 .$$

Then h is holomorphic on $S-\bar{D}_1$ and $|h| \leq 1$. By renaming all the sequences, we can assume $n_k \equiv k$.

Now we consider f_n in D_2, where f_n has a simple pole at p_n and no other pole. Thus, we may write

$$f_n = \frac{g_n}{\varphi-\varphi(p_n)}, \text{ in } D_2,$$

where g_n is <u>holomorphic</u> in D_2. Note that in $D_2-\bar{D}_1$

$$\begin{aligned} |g_n-g_m| &= |[\varphi-\varphi(p_n)]f_n - [\varphi-\varphi(p_m)]f_m| \\ &\leq |\varphi-\varphi(p_n)||f_n-f_m| + |\varphi(p_n) - \varphi(p_m)||f_m| \\ &\leq 4|f_n-f_m| + |\varphi(p_n) - \varphi(p_m)|, \end{aligned}$$

and therefore the sequence g_n converges uniformly on, say, $\partial D_{3/2}$ (since $p_n \to p$, $\varphi(p_n) \to 0$). By the maximum principle, g_n converges uniformly in $D_{3/2}$, say

$$\lim_{n\to\infty} g_n = g, \text{ holomorphic in } D_{3/2}.$$

Now define

$$f_p = \begin{cases} h \text{ in } S-\bar{D}_1 , \\ \frac{g}{\varphi} \text{ in } D_{3/2} . \end{cases}$$

Then we see that f_p is well defined, is meromorphic on S, and has at most a pole of first order at p and no other poles. Since $g_n \to g$ uniformly in $D_{3/2}$ and $f_n \to h$ uniformly in $S-D_{3/2}$, we obtain the result that $f_n \to f_p$ uniformly in $\bar{D}_2-D_1$. Therefore,

$$\sup_{\bar{D}_2 - D_1} |f_p| = 1,$$

$$\inf_{\bar{D}_2 - D_1} |f_p| = \tfrac{1}{2},$$

proving that f_p is not constant. Since S is parabolic, the nonconstant function $|f_p|$ cannot be bounded, and since $|f_p| \le 1$ in $S-\bar{D}_1$, it follows that f_p really does have a pole at p.

Summarizing the construction thus far, we have shown that for every $p \in S$ there exists a meromorphic function f_p on S such that f_p has a pole of order 1 at p and f_p is bounded outside every neighborhood of p.

These conditions essentially uniquely determine f_p. For if $\tilde{f}_p$ has the same properties, then there is a unique complex number $\alpha \neq 0$ such that $\tilde{f}_p - \alpha f_p$ has no pole at p, and is therefore a bounded meromorphic function on all of S. Since S is parabolic, $\tilde{f}_p - \alpha f_p$ is constant, and thus

$$\tilde{f}_p = \alpha f_p + \beta.$$

Conversely, for any constants α and β, $\alpha \neq 0$, the function $\alpha f_p + \beta$ has properties similar to those of f_p.

Also, as we have previously discussed at the top of p. 228 , for a given fixed p, the function $\frac{1}{f_p - f_p(q)}$ has a simple pole at q if q is in a sufficiently small neighborhood of p, and $\frac{1}{f_p - f_p(q)}$ has no other poles. Futhermore, this function is bounded outside any neighborhood of q, if q is sufficiently near p. Thus, by the remark above,

$$\frac{1}{f_p - f_p(q)} = \alpha f_q + \beta,$$

where α and β are constants depending only on q. Thus, for q sufficiently near p there exists a Möbius transformation T_q such that

$$f_q = T_q \circ f_p .$$

Now let p_o be a fixed point in S and let

$$A = \{p \in S: \exists \text{ Möbius transformation } T \text{ such that } f_p = T \circ f_{p_o}\} .$$

Then $p_o \in A$, and the argument just given shows that A is <u>open</u>. The same argument shows that A is <u>closed</u>. In both cases we rely on the fact that the Möbius transformations form a <u>group</u> under composition. Since S is connected, A = S.

Now we prove that f_{p_o} <u>is one-to-one</u>. Suppose

$$f_{p_o}(p) = f_{p_o}(q).$$

Then there exists a Möbius transformation T such that

$$f_p = T \circ f_{p_o} .$$

Therefore,

$$\infty = f_p(p) = T(f_{p_o}(p)) = T(f_{p_o}(q)) = f_p(q).$$

Since f_p has pole at p only, $q = p$.

Now we prove that $\hat{C} - f_{p_o}(S)$ cannot have more than one point. Otherwise, there are two complex numbers $\alpha, \beta \notin f_{p_o}(S)$ - note that definitely $\infty \in f_{p_o}(S)$. Since $f_{p_o}(S)$ is simply connected, the monodromy theorem implies there exists a holomorphic determination of

$$\sqrt{\frac{w-\alpha}{w-\beta}}$$

for $w \in f_{p_o}(S)$; choose that determination which is 1 at $w = \infty$. Define

$$F = \sqrt{\frac{f_{p_o}-\alpha}{f_{p_o}-\beta}}\ .$$

Then F is holomorphic on S and $F(p_o) = 1$. Furthermore, F never takes the value zero and it is impossible for $F(p) = -F(p')$. For if this holds, then $F(p)^2 = F(p')^2$, which implies $f_{p_o}(p) = f_{p_o}(p')$ and $p = p'$, since f_{p_o} is one-to-one. Since F is not constant and takes the value 1, F takes every value z for $|z-1| < \epsilon$, some $\epsilon > 0$. Therefore,

$$|F(p) + 1| \geq \epsilon \quad \text{for all } p \in S.$$

Thus, $\frac{1}{F+1}$ is a bounded holomorphic function on S and is therefore constant since S is parabolic. This is a contradiction.

Now we cannot have $f_{p_o}(S) = \hat{\mathbb{C}}$ by Proposition 9.2 on p. 44, since S is <u>not</u> compact. Therefore, there is a unique complex γ such that $f_{p_o}(S) = \hat{\mathbb{C}} - \{\gamma\}$. Therefore,

$$\frac{1}{f_{p_o} - \gamma}$$

is a one-to-one analytic mapping of S onto $\mathbb{C}$, and thus forms the desired analytic equivalence between S and $\mathbb{C}$.

<u>S is hyperbolic</u>: Here we use Proposition 16 and let g_p be the unique Green's function on $S - \{p\}$. By Lemma 2 there exists a holomorphic function f_p on S such that

$$|f_p| \equiv e^{-g_p}.$$

Then

1. $f_p(p) = 0$,
2. $|f_p| < 1$,
3. f_p is holomorphic on S,
4. f_p does not vanish on $S - \{p\}$,
5. if h is a function on S satisfying 1,2,3 then $|h| \leq |f_p|$.

We have to prove the last statement. By 1, if φ is an analytic chart with $\varphi(p) = 0$, then near p

$$h = \alpha\varphi^n(1 + \beta\varphi + \ldots),$$

where we can assume $\alpha \neq 0$ and $n \geq 1$. Thus, near p we have

$$\log|h| = n \log|\varphi| + \log|\alpha| + \log|1+\beta\varphi+ \ldots|,$$

showing that

$$\frac{-\log|h|}{n} + \log|\varphi|$$

is harmonic near p. Also, $\frac{-\log|h|}{n} > 0$ by 2 and is harmonic away from zeros of h. Let

$$\tilde{h} = \min(g_p, \frac{-\log|h|}{n}).$$

Then $\tilde{h}$ is superharmonic on $S - \{p\}$, $\tilde{h} > 0$, and near p

$$\tilde{h} + \log|\varphi| = \min(g_p + \log|\varphi|, -\frac{\log|h|}{n} + \log|\varphi|)$$

is superharmonic, being the minimum of two harmonic functions near p. By the minimal property of the Green's function,

$$g_p \leq \tilde{h}.$$

Thus,

$$g_p \leq -\frac{\log|h|}{n},$$

so

$$|f_p| = e^{-g_p} \geq e^{\frac{1}{n}\log|h|} = |h|^{\frac{1}{n}},$$

and thus

$$|h| \leq |f_p|^n \leq |f_p|.$$

This proves property 5.

Now let $p,q \in S$ and set

$$h = \frac{f_p - f_p(q)}{1-\overline{f_p(q)}f_p} \quad .$$

Then since all numbers involved have modulus less than 1, we see that $h(q) = 0$, $|h| < 1$, h is holomorphic on S. Therefore, property 5 above implies

$$|h| \leq |f_q| \quad .$$

Since $h(p) = -f_p(q)$, we obtain in particular

$$|f_p(q)| \leq |f_q(p)|.$$

By symmetry we conclude

$$|f_p(q)| = |f_q(p)| \quad \text{for all } p,q \in S.$$

(In terms of the Green's functions, this relation states

$$g_p(q) = g_q(p). \qquad)$$

Thus, we conclude that the holomorphic function $\frac{h}{f_q}$ satisfies

$$\left|\frac{h}{f_q}\right| \leq 1 \text{ on } S,$$

$$\left|\frac{h(p)}{f_q(p)}\right| = 1.$$

By the strong maximum principle, it follows that $\frac{h}{f_q}$ is constant, and in particular

(3) $$\left|\frac{h}{f_q}\right| \equiv 1 \text{ on } S.$$

Now we prove that f_p is one-to-one. Suppose $f_p(q) = f_p(q')$. Then $h(q') = 0$ and by (3) $f_q(q') = 0$. By property 4 of f_q, we conclude that $q' = q$.

Therefore, for any $p \in S$, f_p is a one-to-one analytic mapping of S into the unit disk $\Delta = \{z: |z|<1\}$. We now prove that $f_p(S) = \Delta$. If this is not the case, then a simple topological argument shows that there exists

$$\alpha \in \partial f_p(S), \ |\alpha| < 1.$$

Since $f_p(S)$ is open, $\alpha \notin f_p(S)$. Choose $p_1, p_2, p_3, \ldots$ in S such that

$$f_p(p_n) \to \alpha.$$

Since $f_p(S)$ is simply connected and $\alpha \notin f_p(S)$, the monodromy theorem implies there exists an analytic determination of $\log(w-\alpha)$ for $w \in f_p(S)$. Note that

$$\operatorname{Re} \log(f_p - \alpha) = \log |f_p - \alpha| < \log 2.$$

Let T be a Möbius transformation mapping

$$\{z: \operatorname{Re} z < \log 2\}$$

onto Δ and such that $T(\log(-\alpha)) = 0$. Consider the function

$$F = T \circ \log(f_p - \alpha).$$

Then F is holomorphic on S and $|F| < 1$, $F(p) = T(\log(-\alpha)) = 0$. By property 5 of f_p,

$$|F| \leq |f_p|.$$

We therefore conclude successively that

$$f_p(p_n) - \alpha \to 0,$$

$$\log(f_p(p_n) - \alpha) \to \infty,$$

$$T \circ \log(f_p(p_n) - \alpha) \to \partial\Delta,$$

i.e.,

$$|F(p_n)| \to 1,$$

and thus

$$|f_p(p_n)| \to 1.$$

Thus, $|\alpha| = 1$, a contradiction.

QED

COROLLARY. "THE RIEMANN MAPPING THEOREM" Let S be a connected, simply connected open subset of $\mathbb{C}$ with $S \neq \mathbb{C}$.

<u>Then there exists an analytic equivalence of</u> S <u>and the unit disk</u>.

<u>Proof</u>: We have only to show that the Riemann surface S is hyperbolic. We proceed as on p. 232. If $\alpha \in \mathbb{C} - S$, then there exists a holomorphic determination of $\sqrt{w-\alpha}$ for $w \in S$. Define $F(w) = \sqrt{w-\alpha}$. Then one shows that $F(w) = -F(w')$ implies $w = w'$ by squaring both sides, so that it is impossible that $F(w) = -F(w')$. Suppose $w_0 \in S$. Since F is an open mapping, there exists $\varepsilon > 0$ such that F(S) includes the set $\{z: |z-F(w_0)|<\varepsilon\}$. Therefore, F(S) is disjoint form the set $\{z: |z + F(w_0)| < \varepsilon\}$, or, $|F(w) + F(w_0)| \geq \varepsilon$ for all $w \in S$. Thus $\frac{1}{F+F(w_0)}$ is a bounded, nonconstant holomorphic function on S, proving that S is hyperbolic.

<u>QED</u>

Now we want to indicate some applications of the classification theorem. The first of these is a trivial application, but answers the question of which Riemann surfaces are homeomorphic to a sphere. Cf. p. 42.

<u>THEOREM 1</u>. <u>Let</u> S <u>be a connected compact Riemann surface</u>. <u>Then the following conditions are equivalent</u>.

1. S <u>is analytically equivalent to</u> $\hat{\mathbb{C}}$.
2. S <u>is homeomorphic to</u> $\hat{\mathbb{C}}$.
3. S <u>is simply connected</u>.

4. *There exists a meromorphic function* f *on* S *such that every meromorphic function on* S *is a rational function of* f.

5. *There exists a meromorphic function* f *on* S *having a simple pole at some point and no other pole*.

Proof: *$1 \Rightarrow 2$*: Trivial.

$2 \Rightarrow 3$: Trivial, since a sphere is simply connected.

$3 \Rightarrow 1$: Follows from the classification theorem.

$1 \Rightarrow 4$: We can assume $S = \hat{\mathbb{C}}$ and we then take $f(z) = z$. The result is immediate.

$4 \Rightarrow 5$: We prove that the function f of 4 must be one-to-one. Suppose $p, q \in S$, $p \neq q$. By Theorem 3 of Chapter VI, there exists a meromorphic function g on S such that $g(p) \neq g(q)$. By condition 4, there exists a rational function A such that $g = A \circ f$. Thus, $A(f(p)) \neq A(f(q))$, which implies $f(p) \neq f(q)$. By Proposition 9.2 of Chapter II, f takes every value the same number (one) of times, so f takes the value ∞ one time.

$5 \Rightarrow 1$: By Proposition 9.2 of Chapter II, f maps S onto $\hat{\mathbb{C}}$ in a one-to-one fashion. Thus, f is an analytic equivalence of S onto $\hat{\mathbb{C}}$.

QED

We are now going to discuss the next easiest case. Theorem 1 is concerned with a compact surface of genus 0. We shall next discuss the compact surfaces of genus 1. This case is already so involved that we shall devote a separate chapter to it.

Chapter VIII

THE TORUS

Our use of the classification theorem in proving Theorem 1 of the previous chapter is rather disappointing. For we have applied the classification theorem in the compact case only, and the proof of this case occupies only half of p. 226, whereas the proof of the other two cases requires eleven more pages. Essentially all that has been used is Theorem 4 of Chapter VI and its corollary. In this chapter we shall get to use the full force of the classification theorem in discovering what all the "analytic" tori are. I.e., we shall "classify" the analytic tori.

At first glance, it perhaps seems that the classification theorem, which is addressed to simply connected surfaces, could not be used on tori, which are manifestly not simply connected. In any case, the utility of the classification theorem would be minute if it had no application to anything but simply connected surfaces. Indeed, the theorem states essentially that simply connected Riemann surfaces are trivial in a certain sense.

One of the primary applications of the classification theorem is to the <u>universal</u> <u>covering</u> <u>surface</u> of an arbitrary connected Riemann surface. The universal covering surface is a connected Hausdorff space, and can be made into a Riemann surface in a natural way, as we shall see in Lemma 1. Also, it is simply connected, so the classification theorem applies. Once we know that the universal covering surface is analytically equivalent to the sphere, plane, or disk, then standard topological methods can be invoked to obtain analytic information about the original surface. Actually, in the case of a torus the universal covering surface is obviously the plane, topologically; the "covering map" is also rather obvious; and as a result in this chapter not even the definitions of the concepts mentioned in this paragraph will be given. But the topologically alert reader will know the general setting of what follows.

DEFINITION 1. If T and S are topological spaces and $f: T \to S$, then f is a *local homeomorphism* if for every point $p \in T$ there exist a neighborhood U of p and a neighborhood V of $f(p)$ such that f is a homeomorphism of U onto V.

LEMMA 1. *Let T be a Hausdorff topological space and S a Riemann surface. Let $f: T \to S$ be a local homeomorphism. Then there exists a unique complete analytic atlas on T such that f is an analytic function from the Riemann surface T to S.*

<u>Proof</u>: First we prove uniqueness, so we suppose first that T is a Riemann surface. If $p \in T$, there exist a neighborhood U of p and a neighborhood V of f(p) such that $f\colon U \to V$ is a homeomorphism and V is the domain of an analytic chart $\varphi\colon V \to \varphi(V)$ on S. Let f_1 be the restriction of f to U. Then since f is analytic, f_1 is an analytic equivalence of U onto V, so $\varphi\circ f_1$ must be an analytic chart on U. Knowing an analytic chart in a neighborhood of each point of T implies that we know the complete analytic atlas for T, so the uniqueness follows.

Conversely, we use the above procedure to <u>define</u> charts $\varphi\circ f_1$ on T. We now show these charts form an analytic atlas. If we have another choice, $\tilde{U}$, $\tilde{V}$, $\tilde{\varphi}$, and $\tilde{f}_1$ (the restriction of f to $\tilde{U}$), then where the composition is defined we have

$$\tilde{\varphi}\circ\tilde{f}_1\circ(\varphi\circ f_1)^{-1} = \tilde{\varphi}\circ\tilde{f}_1\, f_1^{-1}\circ\varphi^{-1} = \tilde{\varphi}\circ\varphi^{-1}$$

since $\tilde{f}_1\circ f_1^{-1}$ = identity (we might have to decrease the sizes of everything to achieve this). Since S is a Riemann surface, $\tilde{\varphi}\circ\varphi^{-1}$ is holomorphic, and thus we have an analytic atlas for T. We have to show that f is now analytic, but this is clear. For, <u>on U</u> we have

$$f = f_1 = \varphi^{-1}\circ(\varphi\circ f_1) ,$$

and this is a composition of two analytic functions. Thus, f is analytic in a neighborhood of any point of T.

QED

LEMMA 2. *Let T and S be Riemann surfaces and $f: T \to S$ an analytic local homeomorphism. Let T_1 be a Riemann surface and $g: T_1 \to T$ a continuous function such that $f \circ g$ is analytic. Then g is analytic.*

$$\begin{tikzcd} T_1 \arrow[r, "g"] \arrow[dr, "f\circ g"'] & T \arrow[d, "f"] \\ & S \end{tikzcd}$$

Proof: Given $p \in T_1$, there exist neighborhoods U_1 of p, U of $g(p)$, and V of $f(g(p))$ such that $g: U_1 \to U$ and $f: U \to V$ is an analytic equivalence. Then on U_1 we have

$$g = f_1^{-1} \circ (f \circ g) ,$$

where f_1 is the restriction of f to U. Thus, g is analytic.

QED

Now that the preliminaries are finished, we are ready to discuss tori. The situation is this: S is a Riemann surface which is homeomorphic to a torus.

The problem is to discover what kind of analytic atlas S can have. What we shall do is prove that S is analytically equivalent to one of the C/Ω discussed in Problem 1 of Chapter II, p.24. The problem is essentially to find the complex numbers ω_1 and ω_2 such that $\Omega = \{n_1\omega_1+n_2\omega_2: n_1,n_2 \text{ integers}\}$.

To start with it is convenient to choose a topological representation of S as C/Ω for some Ω which we can pick arbitrarily. Thus, choose arbitrary complex ρ_1 and ρ_2 whose ratio is not real. Then we suppose that S is the set of all cosets $[z] = \{z+n_1\rho_1+n_2\rho_2: n_1,n_2$ integers $\}$, and S is made into a Hausdorff space in the way described in Problem 1. We thus have a concrete representation of S as a topological space, but the analytic atlas for S is unknown. In particular, it is probably not the analytic atlas described in Problem 1, unless we happened to choose ρ_1 and ρ_2 correctly. We reiterate that we are going to prove it is such an analytic atlas with the proper choice of ρ_1 and ρ_2.

Now we are ready to apply Lemmas 1 and 2. First, let C be the complex plane as a topological space without the usual complete analytic atlas and let

$$\pi: C \to S$$

be the natural mapping defined by $\pi(z) = [z]$. Clearly, π is a local homeomorphism, so Lemma 1 shows there is a unique way to make C a Riemann surface such that π is analytic. Let C^* denote this Riemann surface; C^* is homeomorphic but not necessarily analytically equivalent to C as we usually consider it as a Riemann surface.

There are obvious translations on C^*. For example, let

$$t_1 : C^* \to C^*$$

be defined by

$$t_1(z) = z+\rho_1 \ .$$

Then since $\pi(z+\rho_1) = [z+\rho_1] = [z] = \pi(z)$, we have

$$\pi \circ t_1 = \pi \ .$$

Or, we have a commutative diagram.

$$\begin{tikzcd} C^* \arrow[r, "t_1"] \arrow[dr, "\pi"'] & C^* \arrow[d, "\pi"] \\ & S \end{tikzcd}$$

By Lemma 2, t_1 is analytic. Likewise, t_2 is analytic, where $t_2(z) = z+\rho_2$. Two obvious facts about these translations are that t_1 and t_2 commute:

$$t_1 \circ t_2 = t_2 \circ t_1 \ ,$$

and that t_1 and t_2 generate an Abelian group: if n_1 and

n_2 are any integers, the mapping

$$t_1^{n_1} \circ t_2^{n_2} ,$$

which stands for n_1-fold composition of t_1 composed with n_2-fold composition of t_2, is just the mapping

$$z \to z+n_1\rho_1+n_2\rho_2 .$$

Now we apply the classification theorem to C^*, which is simply connected, connected and <u>not</u> compact. Thus, C^* is analytically equivalent to C or the disk $\Delta = \{z: |z|<1\}$. In spite of appearances, it does not seem obvious that C^* is equivalent to C and not Δ, which is indeed the case. It is clear that this question must be faced; cf. p. 42. Let ? = C or Δ as the case may be, and let f be the analytic equivalence:

$$f: ? \to C^* .$$

Using t_1 and t_2, we now define corresponding mappings of ? to itself:

$$A_1 = f^{-1} \, t_1 \circ f ,$$
$$A_2 = f^{-1} \, t_2 \circ f .$$

Then the properties of t_1 and t_2 are obviously reflected in A_1 and A_2: A_1 and A_2 are analytic maps of ? onto ?, they are both one-to-one, they commute, for

$$\begin{aligned} A_1 \circ A_2 &= (f^{-1} \circ t_1 \circ f) \circ (f^{-1} \circ t_2 \circ f) \\ &= f^{-1} \circ t_1 \circ t_2 \circ f \\ &= f^{-1} \circ t_2 \circ t_1 \circ f \\ &= A_2 \circ A_1 , \end{aligned}$$

and they generate an Abelian group, with the formula

$$A_1^{n_1} \circ A_2^{n_2} = f^{-1} \circ t_1^{n_1} \circ t_2^{n_2} \circ f .$$

Now we remark that the only analytic equivalences of Δ onto Δ or of $\mathbb{C}$ onto $\mathbb{C}$ are <u>Möbius</u> transformations. Thus, A_1 and A_2 are both Möbius transformations.

It turns out that the thing relevant to our discussion is the <u>fixed point</u> structure of A_1 and A_2. Suppose now that A is a Möbius transformation of the form

$$A(z) = \frac{az+b}{cz+d} , \quad ad-bc \neq 0 .$$

A point $z \in \hat{\mathbb{C}}$ is a fixed point of A if $A(z) = z$. That is,

$$\frac{az+b}{cz+d} = z .$$

Observe that ∞ is a fixed point if and only if $c = 0$. If $c \neq 0$, the above equation can be written

$$az+b = cz^2+dz ,$$

a quadratic equation for z, which has either two roots or one root. Thus, every Möbius A has one or two fixed points in $\hat{\mathbb{C}}$ (note that if $c=0$, we can write $A(z) = az+b$, and then A has two fixed points if and only if $a \neq 1$).

Now if A is Möbius and an equivalence of Δ onto Δ, and if z is a fixed point of A, then the <u>conjugate</u> of z with respect to $\partial\Delta$ is also a fixed point of A. For suppose w is the conjugate of z (that is, $w = 1/\bar{z}$). Then a property of Möbius transformation is that they preserve conjugacy - thus, $A(z)$ and $A(w)$ must be conjugate with respect to $A(\partial\Delta)$. But $A(z) = z$ and $A(\partial\Delta) = \partial\Delta$, so we see that z and $A(w)$ are conjugate with respect to $\partial\Delta$. Thus, $A(w) = w$.

It is obvious that t_1 has no fixed points in $\mathbb{C}^*$. Thus, <u>A_1 has no fixed points in ?.</u> If $? = \mathbb{C}$, we must have therefore

$$A_1(z) = z+\omega_1 \quad ,$$

and likewise

$$A_2(z) = z+\omega_2 \quad .$$

Here ω_1 and ω_2 are nonzero complex numbers.

If $?=\Delta$, then if A_1 has only one fixed point, it must be on $\partial\Delta$. This follows since A_1 has <u>no</u> fixed points <u>in Δ</u> and since the conjugate with respect to $\partial\Delta$ of a fixed point of A_1 is also a fixed point of A_1. Likewise, if A_1 has two fixed points, they both lie on $\partial\Delta$.

<u>Proof that $?=\mathbb{C}$</u>. Suppose the contrary, that $?=\Delta$. There are two cases to consider. Suppose first that A_1 has two fixed points, α and β. Then

$$A_1(A_2(\alpha)) = A_2(A_1(\alpha)) = A_2(\alpha) \quad ,$$

so $A_2(\alpha)$ is a fixed point of A_1. Likewise, $A_2(\beta)$ is a fixed point of A_1. So either $A_2(\alpha) = \alpha$ and $A_2(\beta) = \beta$, or $A_2(\alpha) = \beta$ and $A_2(\beta) = \alpha$. Now define the Möbius transformation

$$m(z) = \frac{z-\alpha}{z-\beta} \quad .$$

Then the transformation

$$m \circ A_1 \circ m^{-1}$$

maps 0 to 0 and ∞ to ∞, and thus is multiplication by a complex number a_1. Since $m(\partial\Delta)$ is a straight line through 0, it follows that $m(\Delta)$ is a half plane bounded by a straight line through 0. And the mapping $z \to a_1 z$ maps this half plane onto itself. Thus, a_1 is a positive real number. That is,

$$m \circ A_1 \circ m^{-1}(z) = a_1 z, \qquad 0<a_1<\infty \quad .$$

If we have $A_2(\alpha) = \alpha$ and $A_2(\beta) = \beta$, then also

$$m \circ A_2 \circ m^{-1}(z) = a_2 z, \qquad 0<a_2<\infty \quad .$$

On the other hand, if $A_2(\alpha) = \beta$ and $A_2(\beta) = \alpha$, then

$m \circ A_2 \circ m^{-1}$ maps 0 to ∞ and ∞ to 0. Thus, for some nonzero complex b,

$$m \circ A_2 \circ m^{-1}(z) = \frac{b}{z} .$$

Then it follows that

$$(m \circ A_2 \circ m^{-1}) \circ (m \circ A_2 \circ m^{-1})(z) = \frac{b}{b/z} = z ,$$

so that also

$$A_2 \circ A_2 = \text{identity} .$$

But then also $t_2 \circ t_2 = \text{identity}$, a contradiction since $t_2 \circ t_2$ is translation by $2\rho_2$. Therefore, we conclude that

$$m \circ A_j \circ m^{-1}(z) = a_j z, \quad j=1,2 .$$

Now we need a lemma.

<u>LEMMA 3</u>. <u>Let</u> $x, y \in \mathbb{R}$. <u>Then there exist integers</u> m_k, n_k <u>such that for each</u> k, m_k <u>and</u> n_k <u>are not both zero</u>, <u>and</u>

$$\lim_{k\to\infty} (m_k x + n_k y) = 0 .$$

<u>Proof</u>: We can obviously assume x and y are not both zero and that $\frac{x}{y} = \xi$ is irrational. Let N be any positive integer. For $1 \le j \le N+1$ there exists a unique

integer ℓ_j such that

$$0 < j\xi - \ell_j < 1 \quad .$$

Among the N intervals $(0,\frac{1}{N})$, $(\frac{1}{N}, \frac{2}{N})$, ..., $(\frac{N-1}{N}, 1)$ there must be one which contains two of the numbers $j\xi - \ell_j$, say for j_1 and j_2. Then

$$|(j_1\xi-\ell_{j_1}) - (j_2\xi-\ell_{j_2})| < \frac{1}{N} \; .$$

<u>QED</u>

Now we apply this lemma to the real numbers $\log a_1$ and $\log a_2$ to obtain $m_k \log a_1 + n_k \log a_2 \to 0$. Exponentiating, $a_1^{m_k} a_2^{n_k} \to 1$. Thus, for each z we have

$$m \circ A_1^{m_k} \circ A_2^{n_k} \circ m^{-1}(z) \to z \quad .$$

Therefore, since m and m^{-1} are continuous,

$$A_1^{m_k} \circ A_2^{n_k}(z) \to z$$

for each z, and thus

$$t_1^{m_k} \circ t_2^{n_k}(z) \to z$$

for each z. This says $z + m_k\rho_1 + n_k\rho_2 \to z$, and thus $m_k\rho_1 + n_k\rho_2 \to 0$. This contradicts the fact that ρ_1 and ρ_2 have a nonreal ratio; cf. the discussion under Problem 1.

The only other case is the case in which A_1 and

A_2 each have only one fixed point. Suppose $A_1(\alpha) = \alpha$. As we saw on p. 250, $A_2(\alpha)$ is a fixed point of A_1, so also $A_2(\alpha) = \alpha$. Let m be the Möbius transformation

$$m(z) = \frac{e^{i\theta}}{z-\alpha} \quad .$$

The $m(\alpha) = \infty$ and $m(\partial\Delta)$ is a straight line. We choose θ to force this straight line to be parallel to the real axis. Then $m \circ A_1 \circ m^{-1}$ maps ∞ to ∞ and has no other fixed point, so

$$m \circ A_1 \circ m^{-1}(z) = z + a_1 \quad .$$

Since $m \circ A \circ m^{-1}$ maps the associated half plane onto itself, $a_1 \in \mathbb{R}$. Likewise,

$$m \circ A_2 \circ m^{-1}(z) = z + a_2 \quad , \quad a_2 \in \mathbb{R} \quad .$$

By Lemma 3, there exist integers m_k and n_k such that $m_k a_1 + n_k a_2 \to 0$. Therefore, as in the argument above we obtain $m_k \rho_1 + n_k \rho_2 \to 0$, a contradiction.

Thus, we have now completely contradicted the assumption that $? = \Delta$. The only other possibility must hold. Thus, $\underline{? = \mathbb{C}}$.

Now from p.249 we know that $A_j(z) = z + \omega_j$, $j=1,2$. Exactly as in the above discussion, it follows that ω_1 and ω_2 have a nonreal ratio. Thus, if we define $\Omega = \{n_1\omega_1 + n_2\omega_2 : n_1, n_2 \text{ integers}\}$, we have a Riemann surface $\mathbb{C}/\Omega$ as defined in Problem 1.

Consider the diagram

$$\begin{tikzcd} \mathbb{C}^* \arrow[d, "\pi"] & \mathbb{C} \arrow[l, "f"'] \arrow[d, "\pi_1"] \\ S & \mathbb{C}/\Omega \arrow[l, "F"] \end{tikzcd}$$

where the map π_1 is $z \to z + \Omega$. What we want to do is obtain an analytic function F from $\mathbb{C}/\Omega$ to S. First, we can <u>define</u> a function F by

$$F(z + \Omega) = \pi\circ f(z).$$

This makes sense, for if $z + \Omega = z' + \Omega$, then $z = z' + n_1\omega_1 + n_2\omega_2$ for some integers n_1 and n_2, and thus

$$\begin{aligned}\pi\circ f(z) &= \pi\circ f(z' + n_1\omega_1 + n_2\omega_2)\\ &= \pi\circ f(A_1^{n_1}\circ A_2^{n_2}(z'))\\ &= \pi\circ t_1^{n_1}\circ t_2^{n_2}(f(z'))\\ &= \pi\circ f(z').\end{aligned}$$

Thus, F is uniquely defined such that $F\circ\pi_1 = \pi\circ f$. Since π_1 is locally an analytic equivalence, we have $F = \pi\circ f\circ\pi_1^{-1}$ <u>locally</u> and thus F is analytic. Since π and f are surjections, so is $F\circ\pi_1$ and thus so is F. Finally, F is one-to-one. For, suppose $F(z+\Omega) = F(z'+\Omega)$. Then $\pi\circ f(z) = \pi\circ f(z')$, so that there exist integers n_1 and n_2 such that

$$\begin{aligned}f(z) &= f(z') + n_1\rho_1 + n_2\rho_2\\ &= t_1^{n_1}\circ t_2^{n_2}\circ f(z').\end{aligned}$$

Thus,

$$z = f^{-1} \circ t_1^{n_1} \circ t_2^{n_2} \circ f(z')$$

$$= A_1^{n_1} \circ A_2^{n_2}(z')$$

$$= z' + n_1\omega_1 + n_2\omega_2.$$

Thus, $z + \Omega = z' + \Omega$, proving F is one-to-one.

We shall now formally state a theorem which includes the above discussion. We need to recall Definition 2 of Chapter V.

THEOREM 1. *Let S be a compact, connected Riemann surface. Then the following conditions are equivalent.*

1. *S is analytically equivalent to the Riemann surface of a polynomial*
$$w^2 - 4(z-e_1)(z-e_2)(z-e_3),$$
where e_1, e_2, e_3 *are distinct complex numbers.*
2. *S is analytically equivalent to the Riemann surface of a polynomial*
$$w^2 - (z-a_1)(z-a_2)(z-a_3)(z-a_4),$$
where a_1, a_2, a_3, a_4 *are distinct complex numbers.*
3. *S is homeomorphic to a torus.*
4. *S is analytically equivalent to a torus of the form* $\mathbb{C}/\Omega$ *of Problem 1.*

Proof: 1 ⇒ 2: This is simple algebra. We can assume $S \subset \overline{M}$ is the Riemann surface of the polynomial 1. Let $\alpha \in \mathbb{C}$, $\alpha \neq e_1$ or e_2 or e_3. Define

$$f = \frac{1}{\pi-\alpha} \quad ,$$

where π and V are the functions on $\overline{M}$ discussed in Chapter IV, restricted to S. Thus,

$$V^2 = 4(\pi-e_1)(\pi-e_2)(\pi-e_3);$$

$$\begin{aligned} f^4V^2 &= 4f(f(\pi-\alpha)+(\alpha-e_1)f)(f(\pi-\alpha)+(\alpha-e_2)f)(f(\pi-\alpha) \\ &\quad + (\alpha-e_3)f) \\ &= 4f(1 + (\alpha-e_1)f)(1 + (\alpha-e_2)f)(1 + (\alpha-e_3)f) \\ &= 4(\alpha-e_1)(\alpha-e_2)(\alpha-e_3)f(f-\frac{1}{e_1-\alpha})(f-\frac{1}{e_2-\alpha})(f-\frac{1}{e_3-\alpha}) \end{aligned}$$

Choose a complex number $\beta = \sqrt{4(\alpha-e_1)(\alpha-e_2)(\alpha-e_3)}$ and let

$$g = \frac{f^2V}{\beta} \quad .$$

Thus,

$$g^2 = f(f-\frac{1}{e_1-\alpha})(f-\frac{1}{e_2-\alpha})(f-\frac{1}{e_3-\alpha}).$$

Now f and g are meromorphic on S and f takes every value 2 times, since the same is true of π. Now consider the diagram of p. 156

$$\begin{tikzcd} S \arrow[rr, dashed, "\Phi"] \arrow[dr, "{f,g}"'] & & T \arrow[dl, "{\pi,V}"] \\ & \hat{\mathbb{C}} \times \hat{\mathbb{C}} & \end{tikzcd}$$

The polynomial equation satisfied by f and g is easily seen to be irreducible and it is of degree 2 in g. Therefore, Theorem 2 of Chapter VI implies Φ is an analytic equivalence. This proves $1 \Rightarrow 2$.

$\underline{2 \Rightarrow 3}$: This follows trivially from the cutting and gluing process described on pp. 9-13 . Also, it follows from the Riemann-Hurwitz formula of p. 112. In this case $V = 4$ and $n = 2$, so that $g = 1$.

$\underline{3 \Rightarrow 4}$: This is the content of the discussion preceding this theorem.

$\underline{4 \Rightarrow 1}$: Now we have to do some work. In fact, we need to introduce some rather classical and famous concepts of the theory of elliptic functions. Suppose that $\omega_1, \omega_2 \in \mathbb{C}$ have nonreal ratio and define as usual $\Omega = \{n_1\omega_1 + n_2\omega_2 : n_1, n_2 \text{ integers}\}$. Define

$$\wp(z) = \frac{1}{z^2} + \sum_{\substack{\zeta \in \Omega \\ \zeta \neq 0}} \left(\frac{1}{(z-\zeta)^2} - \frac{1}{\zeta^2} \right).$$

We must first prove this series converges. If K is a compact subset of $\mathbb{C}-\Omega$, then for $z \in K$

$$\left|\frac{1}{(z-\zeta)^2} - \frac{1}{\zeta^2}\right| = \frac{|-z^2+2z\zeta|}{|(z-\zeta)^2\zeta^2|} \le c|\zeta|^{-3} .$$

We now check that the series of constants

$$\sum_{\substack{\zeta\in\Omega \\ \zeta\neq 0}} |\zeta|^{-3}$$

converges. Since ω_1 and ω_2 have nonreal ratio, it follows that for any $\theta\in[0,2\pi]$, $\omega_1 \cos\theta + \omega_2 \sin\theta \neq 0$. Since this is a continuous function of θ, there exists a positive constant δ such that

$$|\omega_1 \cos\theta + \omega_2 \sin\theta| \ge \delta, \; 0\le\theta\le 2\pi .$$

Multiplying both sides of this inequality by a positive number, we obtain

$$|x\omega_1 + y\omega_2| \ge \delta\sqrt{x^2+y^2}, \text{ x and y real.}$$

Therefore, summing on squares in R^2, we obtain

$$\sum_{\substack{\zeta\in\Omega \\ \zeta\neq 0}} |\zeta|^{-3} \le \delta^{-3} \sum_{\substack{n_1,n_2 \\ \text{not both} \\ \text{zero}}} (n_1^2+n_2^2)^{-3/2}$$

$$= \delta^{-3} \sum_{k=1}^{\infty} \sum_{\max(|n_1|,|n_2|)=k} (n_1^2+n_2^2)^{-3/2}$$

$$\le \delta^{-3} \sum_{k=1}^{\infty} 4(2k+1)k^{-3} < \infty .$$

Therefore, the series defining $\wp$ converges uniformly on any compact subset of $\mathbb{C}-\Omega$, and one sees likewise that for any lattice point ζ_o, the series for $\wp(z) - \frac{1}{(z-\zeta_o)^2}$ converges uniformly on any compact subset of $\mathbb{C}-(\Omega-\{\zeta_o\})$. Therefore, $\wp$ is meromorphic on $\mathbb{C}$ and has poles of order 2 at each $\zeta\in\Omega$. The function $\wp$ is called the Weierstrass pe-function.

The first remark to be made is that $\wp$ is an even function. For,

$$\wp(-z) = \frac{1}{z^2} + \sum_{\substack{\zeta\in\Omega \\ \zeta\neq 0}} \left(\frac{1}{(-z-\zeta)^2} - \frac{1}{\zeta^2}\right) .$$

Replacing the "dummy" ζ by $-\zeta$, we therefore obtain

$$\wp(-z) = \frac{1}{z^2} + \sum_{\substack{\zeta\in\Omega \\ \zeta\neq 0}} \left(\frac{1}{(z-\zeta)^2} - \frac{1}{\zeta^2}\right)$$

$$= \wp(z) \quad .$$

Next, we compute $\wp'$; since the series for $\wp$ converges uniformly locally, this can be done formally:

$$\wp'(z) = \frac{-2}{z^3} + \sum_{\substack{\zeta\in\Omega \\ \zeta\neq 0}} \frac{-2}{(z-\zeta)^3}$$

$$= -2 \sum_{\zeta\in\Omega} \frac{1}{(z-\zeta)^3} \quad .$$

Thus, if $\zeta_o\in\Omega$,

$$\wp'(z+\zeta_o) = -2 \sum_{\zeta\in\Omega} \frac{1}{(z+\zeta_o-\zeta)^3}$$

$$= -2 \sum_{\zeta\in\Omega} \frac{1}{(z-\zeta)^3}$$

$$= \wp'(z) ,$$

where we have replaced the "dummy" ζ by $\zeta+\zeta_0$. This implies that

$$\wp(z+\omega_1) - \wp(z) \equiv \text{constant}.$$

We evaluate this constant by setting $z = -\frac{\omega_1}{2}$. Since $\wp$ is even,

$$\wp(\tfrac{\omega_1}{2}) - \wp(-\tfrac{\omega_1}{2}) = 0 \quad ,$$

and thus the constant is zero. Therefore, using the same argument for ω_2,

$$\wp(z+\omega_1) = \wp(z),$$

$$\wp(z+\omega_2) = \wp(z).$$

These relations imply that we can regard $\wp$ as a meromorphic function on $\mathbb{C}/\Omega$, whose value at $z+\Omega$ is just $\wp(z)$. To keep track of the notation, let $\wp_0$ be this function:

$$\wp_0(z+\Omega) = \wp(z) \quad .$$

Then since $\wp$ has double poles at the points in Ω and no other points, we see that $\wp_0$ has a double pole at $0+\Omega$ and no other pole. Since $\mathbb{C}/\Omega$ is a compact Riemann surface, $\wp_0$ takes every value 2 times.

Likewise, if we define

$$\wp_0'(z+\Omega) = \wp'(z),$$

then $\wp_0'$ takes every value 3 times, since $\wp'$ has triple poles at the points of Ω. We shall be interested in particular in the zeros of $\wp'$. If $\varsigma_0 \in \Omega$ but $\frac{1}{2}\varsigma_0 \notin \Omega$, then since $\wp'$ is odd, being the derivative of an even function,

$$\wp'(\tfrac{1}{2}\varsigma_0) = \wp'(\tfrac{1}{2}\varsigma_0 - \varsigma_0) = \wp'(-\tfrac{1}{2}\varsigma_0) = -\wp'(\tfrac{1}{2}\varsigma_0) .$$

Since $\frac{1}{2}\varsigma_0 \notin \Omega$, $\wp'(\frac{1}{2}\varsigma_0) \neq \infty$, and thus $\wp'(\frac{1}{2}\varsigma_0) = 0$. Thus,

$$\wp_0'(\tfrac{1}{2}\omega_1+\Omega) = 0 \quad ,$$

$$\wp_0'(\tfrac{1}{2}\omega_2+\Omega) = 0 \quad ,$$

$$\wp_0'(\tfrac{1}{2}\omega_1+\tfrac{1}{2}\omega_2+\Omega) = 0 \quad .$$

Now define

$$\wp_0(\tfrac{1}{2}\omega_1+\Omega) = e_1 \quad ,$$

$$\wp_0(\tfrac{1}{2}\omega_2+\Omega) = e_2 \quad ,$$

$$\wp_0(\tfrac{1}{2}\omega_1+\tfrac{1}{2}\omega_2+\Omega) = e_3 \quad .$$

Since $\wp_0'$ takes every value 3 times, we have found all of its zeros. And these must therefore also be simple zeros of $\wp_0'$. Also, we see that $\wp_0$ takes the value e_1 2 times at $\frac{1}{2}\omega_1+\Omega$ (a zero of $\wp_0'$) and likewise for e_2

and e_3. In particular, e_1, e_2, e_3 are distinct.

Now consider the meromorphic function

$$\frac{\wp'^2}{(\wp - e_1)(\wp - e_2)(\wp - e_3)}$$

on $\mathbb{C}/\Omega$. The numerator and denominator have poles only at $0+\Omega$, and near there we have the Laurent development

$$\frac{(-\frac{2}{z^3}+\dots)^2}{(\frac{1}{z^2}+\dots)^3} = \frac{\frac{4}{z^6}+\dots}{\frac{1}{z^6}+\dots} = 4+\dots \quad .$$

Thus, the function has no pole at $0+\Omega$, and in fact is equal to 4 at $0+\Omega$. The only other possibilities for poles are at $\frac{\omega_1}{2} + \Omega$, $\frac{\omega_2}{2} + \Omega$, and $\frac{\omega_1}{2} + \frac{\omega_2}{2} + \Omega$. But at these points the numerator has zeros of order 2 and the denominator has zeros of order 2. Thus, the function has no pole at all, and is thus constant. Therefore,

$$\wp'^2 \equiv 4(\wp - e_1)(\wp - e_2)(\wp - e_3) \quad .$$

This is the classical differential equation for $\wp$.

As on p. 257, consider the diagram

$$\begin{tikzcd}
\mathbb{C}/\Omega \arrow[rr, dashed, "\Phi"] \arrow[dr, "{\wp, \wp'}"'] & & T \arrow[dl, "{\pi, V}"] \\
& \hat{\mathbb{C}} \times \hat{\mathbb{C}} &
\end{tikzcd}$$

Since $\wp_0$ takes every value 2 times and the algebraic equation relating $\wp_0$ and $\wp_0'$ has degree 2 in $\wp_0'$ and is irreducible, Theorem 2 of Chapter VI implies Φ is an analytic equivalence. And T is the Riemann surface of $w^2-4(z-e_1)(z-e_2)(z-e_3)$.

QED

Even among the tori $\mathbb{C}/\Omega$ there are lots of equivalences. We now treat this problem.

THEOREM 2. Two tori $\mathbb{C}/\Omega$ and $\mathbb{C}/\tilde{\Omega}$ are analytically equivalent if and only if there exists a nonzero complex a such that

$$\tilde{\Omega} = a\Omega \quad .$$

Proof: If $\tilde{\Omega} = a\Omega$, then we can define a mapping $\mathbb{C}/\Omega \to \mathbb{C}/\tilde{\Omega}$ by the formula $z+\Omega \to az+\tilde{\Omega}$, and this is easily seen to be an analytic equivalence.

Conversely, suppose $F:\mathbb{C}/\Omega \to \mathbb{C}/\tilde{\Omega}$ is an analytic equivalence. If $F(0+\Omega) = \alpha+\tilde{\Omega}$, then let

$$t: \mathbb{C}/\tilde{\Omega} \to \mathbb{C}/\tilde{\Omega}$$

be defined by

$$t(z+\tilde{\Omega}) = z-\alpha+\tilde{\Omega} \quad .$$

Then t is an analytic equivalence and

$$t\circ F(0+\Omega) = t(\alpha+\tilde{\Omega}) = 0+\tilde{\Omega} \quad .$$

By considering $t\circ F$ instead of F, we see that there is no loss of generality in assuming

$$F(0+\Omega) = 0+\tilde{\Omega} \quad .$$

Given $z_o \in \mathbb{C}$ choose arbitrarily $w_o \in \mathbb{C}$ such that

$$F(z_o+\Omega) = w_o+\tilde{\Omega} \quad .$$

Let $\pi: \mathbb{C} \to \mathbb{C}/\Omega$ and $\tilde{\pi}: \mathbb{C} \to \mathbb{C}/\tilde{\Omega}$ be the canonical mappings and choose a neighborhood $\tilde{U}$ of w_o such that $\tilde{\pi}$ is one-to-one on $\tilde{U}$. Then for z near z_o consider the mapping

$$g_{w_o}(z) = \tilde{\pi}^{-1}(F(z+\Omega)) \quad .$$

This is a holomorphic function in a neighborhood of z, and if we choose a different w_o' such that $F(z_o+\Omega) = w_o'+\tilde{\Omega}$, then $w_o' = w_o + \tilde{\zeta}_o$, where $\tilde{\zeta}_o \in \tilde{\Omega}$, and the associated $\tilde{\pi}^{-1}$ is thus equal to the original $\tilde{\pi}^{-1} + \tilde{\zeta}_o$. Thus,

$$g_{w_o'}(z) = g_{w_o}(z) + \zeta_o \quad .$$

It follows that

$$\frac{d}{dz}\, g_{w_o}(z)$$

is well defined near z_o in the sense that it is independent of the choice w_o. Thus, we can define a function h on $\mathbb{C}$ by the formula

$$h(z) = \frac{d}{dz} g_{w_o}(z) \quad , \quad z \text{ near } z_o \quad .$$

Then h is holomorphic on $\mathbb{C}$ and since for $z_o' = z_o + \zeta_o$, $\zeta_o \in \Omega$, we can take w_o to be the same and thus for z near z_o'

$$\begin{aligned} g_{w_o}(z) &= \tilde{\pi}^{-1}(F(z+\Omega)) \\ &= \tilde{\pi}^{-1}(F(z-\zeta_o+\Omega)) \\ &= g_{w_o}(z-\zeta_o) \quad , \end{aligned}$$

it follows that $h(z) = h(z-\zeta_o)$ for z near z_o'. Hence, $h(z_o') = h(z_o'-\zeta_o) = h(z_o)$, so h represents a meromorphic function on $\mathbb{C}/\Omega$. But h is holomorphic and thus constant, say $h \equiv a$.

Therefore, following the above notation we have

$$g_{w_o}(z) = az + b$$

for z near z_o. Applying $\tilde{\pi}$ to both sides we obtain

$$F(z+\Omega) = az+b+\tilde{\Omega} \quad \text{for } z \text{ near } z_o.$$

Here b is a constant which can depend on z_o. By a connectivity argument it is easy to see that the constants b which can appear here must differ from each other only by elements of $\tilde{\Omega}$. Since $F(0+\Omega) = 0+\tilde{\Omega}$, the b associated with $z_o = 0$ must itself belong to $\tilde{\Omega}$. Therefore, we have proved

$$F(z+\Omega) = az+\tilde{\Omega} \quad .$$

This much has been done assuming only that F is analytic and not necessarily one-to-one.

If we assume that $F: \mathbb{C}/\Omega \to \mathbb{C}/\tilde{\Omega}$ is one-to-one and onto, then since $F(0+\Omega) = 0+\tilde{\Omega}$, we have

$$z+\Omega = 0+\Omega \Leftrightarrow F(z+\Omega) = F(0+\Omega)$$
$$\Leftrightarrow az+\tilde{\Omega} = 0+\tilde{\Omega} \quad .$$

That is,

$$z\in\Omega \Leftrightarrow az\in\tilde{\Omega} \quad .$$

But this means that $\tilde{\Omega} = a\Omega$.

QED

Of course, we can describe the relation $\tilde{\Omega} = a\Omega$ algebraically rather than geometrically. If

$$\Omega = \{n_1\omega_1+n_2\omega_2 : n_1, n_2 \text{ integers}\} \quad ,$$
$$\tilde{\Omega} = \{n_1\tilde{\omega}_1+n_2\tilde{\omega}_2 : n_1, n_2 \text{ integers}\} \quad ,$$

then we have

$$a\omega_1 = n_{11}\tilde{\omega}_1+n_{12}\tilde{\omega}_2 \quad ,$$
$$a\omega_2 = n_{21}\tilde{\omega}_1+n_{22}\tilde{\omega}_2 \quad ,$$

for certain integers n_{jk}. Also,

$$\tilde{\omega}_1 = m_{11}a\omega_1+m_{12}a\omega_2 \quad ,$$

$$\tilde{\omega}_2 = m_{21}a\omega_1+m_{22}a\omega_2 \quad ,$$

for integers m_{jk}. Therefore, we have the product of matrices

$$\begin{pmatrix} n_{11} & n_{12} \\ n_{21} & n_{22} \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

Therefore, the product of the determinants is 1, or

$$n_{11}n_{22} - n_{21}n_{12} = \pm 1 .$$

Conversely, if this equation holds, then the relations expressing $a\omega_1$ and $a\omega_2$ in terms of $\tilde{\omega}_1$ and $\tilde{\omega}_2$ can be inverted, and thus $\tilde{\Omega} = a\Omega$.

Almost as an afterthought we mention that if S is a compact, connected Riemann surface, then a necessary and sufficient condition that there exist a meromorphic function on S which takes every value 2 times is that S be analytically equivalent to the Riemann surface of a polynomial

$$w^2 - (z-a_1)(z-a_2)\dots(z-a_\ell) ,$$

where $a_1, a_2, \dots, a_\ell$ are distinct.

The proof is almost trivial. If S is the Riemann surface of the above polynomial, then the function π takes every value 2 times since the polynomial has degree 2 in w. Conversely, suppose f is a meromorphic function on S which takes every value 2 times. By the proof of the corollary on p. 160, there exists a meromorphic function g on S such that f and g satisfy a polynomial equation which is irreducible and has degree 2 in g. Thus, for certain rational functions a and b,

$$g^2 - 2a(f)g + b(f) = 0 \quad .$$

Completing the square,

$$(g-a(f))^2 = a(f)^2 - b(f) \quad .$$

Let $g_1 = g-a(f)$ and $f_1 = a(f)^2 - b(f)$. Then

$$g_1^2 = f_1 \quad .$$

Since f_1 is a rational function of f, we can write

$$g_1^2 = \alpha^2 \frac{\prod_{k=1}^{m}(f-\alpha_k)^2}{\prod_{k=1}^{m'}(f-\alpha_k')^2} \times \frac{\prod_{j=1}^{n}(f-\beta_j)}{\prod_{j=1}^{n'}(f-\beta_j')} \quad ,$$

where the α's and β's are complex constants and the numbers β_j and β_j' are distinct. Let

$$g_2 = \frac{1}{\alpha} \frac{\prod_{k=1}^{m'}(f-\alpha_k')}{\prod_{k=1}^{m}(f-\alpha_k)} \prod_{j=1}^{n'}(f-\beta_j')\, g_1 \quad .$$

Then

$$g_2^2 = \prod_{j=1}^{n}(f-\beta_j) \prod_{j=1}^{n'}(f-\beta_j') \quad .$$

Thus, we have produced distinct complex numbers $\alpha_1,\ldots,\alpha_\ell$ $(\ell=n+n')$ and a meromorphic function g_2 such that

$$g_2^2 = \prod_{k=1}^{\ell} (f-\alpha_k) \ .$$

This type of Riemann surface is called hyperelliptic.

Appendix

FINAL EXAMINATION

1. Let a and b be relatively prime positive integers. Analyze the Riemann surface of the polynomial

$$A(z,w) = w^{2a} - 2z^b w^a + 1 \quad .$$

Do the same for the polynomial

$$B(z,w) = z^{2a} - 2w^b z^a + 1 \quad .$$

Be sure to compute the genus in each case and check that they are equal.

2. Let $A(z,w)$ be an irreducible polynomial of degree at least 2 in w. Prove that there does not exist a rational function f such that

$$A(z,f(z)) \equiv 0 \quad .$$

3. If $A(z,w)$ is an irreducible polynomial and S is the Riemann surface of A, prove that S cannot have exactly one branch point (of possibly high order).

4. The Riemann surface of the polynomial $w^3 + z^3 - 1$ is easily seen to have genus 1. Thus, it is homeomorphic to a torus and by our general theorem is analytically equivalent to the Riemann surface of a polynomial of the form

$$w^2 - 4(z-e_1)(z-e_2)(z-e_3) \quad .$$

Find such a polynomial explicitly.

Hint. Use algebra only.

5. Prove that the sum of two algebraic functions is algebraic. Compute explicitly a polynomial $A(z,w)$ such that

$$A(z,z^{1/2}+z^{1/3}) = 0 \quad .$$

6. Are the following (noncompact) Riemann surfaces parabolic or hyperbolic?

   a. A compact Riemann surface minus a point.

   b. A Riemann surface minus the closure of an analytic disk.

   c. A Riemann surface on which a Green's function exists.

SOLUTIONS TO PROBLEMS 6 AND 7

Problem 6 (p.139). Analysis of $A(z,w) = w^3 - 3zw + z^3$.

Irreducible: If not, A must have a linear factor, so

$$A = (w+\alpha)(w^2+\beta w+\gamma) \quad ,$$

where α,β,γ are polynomials in z which must satisfy

$$\alpha + \beta = 0 \quad ,$$
$$\alpha\beta + \gamma = -3z,$$
$$\alpha\gamma = z^3$$

The third relation shows that $\alpha = cz^k$, where $c \neq 0$ and $k \in \{0,1,2,3\}$; solving for α and β and using the second relation shows that

$$-c^2z^{2k} + c^{-1}z^{3-k} \equiv -3z \quad ,$$

an impossible identity.

Critical points: By definition, $z=\infty$ is critical. Since $A(0,w) = w^3$, $z=0$ is critical. For other z, we look for solutions of the pair of equations $A(z,w) = 0$ and

$$\frac{\partial A}{\partial w} = 3w^2 - 3z = 0 \quad .$$

That is,

$$\begin{cases} w^3 - 3zw + z^3 = 0 \quad , \\ \qquad\qquad w^2 = z \quad . \end{cases}$$

Thus, $w^3 - 3w^3 + w^6 = 0$, so $w^6 = 2w^3$ and since $w \neq 0$ we have $w = 2^{1/3}\omega^k$, where $2^{1/3} > 0$ and $\omega = e^{2\pi i/3}$, $k = 0,1,2$. Thus,

$$\underline{z = 2^{2/3}\omega^{2k}}, \quad k=0,1,2.$$

Puiseaux expansions:

$z=0$. First, we argue heuristically. If w_1, w_2, w_3 are the zeros of A, then

$$(*) \qquad \begin{cases} w_1 + w_2 + w_3 = 0 \quad , \\ w_1w_2 + w_2w_3 + w_3w_1 = -3z \quad , \\ w_1w_2w_3 = -z^3 \quad . \end{cases}$$

If there is no branching, these are all holomorphic near $z=0$ and $|w_k| \le C|z|$, contradicting the second line of (*). If the branching is of order 2, then each $|w_k|$ is asymptotic to const $|z|^{\ell/3}$ for some integer ℓ. The third line of (*) shows $\ell=3$, again contradicting the second line. The only other possibility is a branch point of order 1 and a holomorphic solution $e(t,tQ)$. For this solution we have

$$t^3Q^3 - 3t^2Q + t^3 = 0 \quad .$$

Thus,

$$tQ^3 - 3Q + t = 0 \quad .$$

Thus, $Q(0) = 0$, so we let $Q = tQ_1$ and find

$$t^4Q_1^3 - 3tQ_1 + t = 0 \quad .$$

Thus,

$$t^3Q_1^3 - 3Q_1 + 1 = 0 \quad .$$

The derivative of this polynomial with respect to Q_1 equals -3 at t=0, so the implicit function theorem implies Q_1 exists with $Q_1(0) = 1/3$. Thus, the Riemann surface has an element

$$e(t,t^2/3 + \ldots) \quad .$$

The branched element we represent as $e(t^2,tQ)$ and find

$$t^3Q^3 - 3t^3Q + t^6 = 0 \quad .$$

Thus,

$$Q^3 - 3Q + t^3 = 0 \quad .$$

At t=0 there is a solution $Q(0) = \sqrt{3}$ and the implicit function theorem again can be applied to provide an element

$$e(t^2, \sqrt{3}t + \ldots) \quad .$$

<u>$z=\infty$</u>. The heuristic analysis is similar. Now we try $e(\frac{1}{t}, \frac{Q}{t})$ and obtain

$$t^{-3}Q^3 - 3t^{-2}Q + t^{-3} = 0;$$

$$Q^3 - 3tQ + 1 = 0.$$

At t=0 we obtain 3 <u>distinct</u> solutions $Q(0) = -\omega^j$, so we find 3 unbranched solutions

$$e(\frac{1}{t}, \frac{-1}{t} + \ldots) \quad ,$$

$$e(\frac{1}{t}, \frac{-\omega}{t} + \ldots) \quad ,$$

$$e(\frac{1}{t}, \frac{-\omega^2}{t} + \ldots) \quad .$$

<u>$z=2^{2/3}\omega^{2k}$</u>. (k=0,1,2) Again we omit the heuristics, except to note that $w=2^{1/3}\omega^k$ is <u>exactly</u> a double root, so there is at least one unbranched solution, and the corresponding root is $-2\cdot 2^{1/3}\omega^k$. Thus, there is an element

$$e(2^{2/3}\omega^{2k} + t, -2\cdot 2^{1/3}\omega^k + \ldots) \quad .$$

Now we see whether the other two solutions are branched. If not, then there is an element

$$e(2^{2/3}\omega^{2k} + t,\ 2^{1/3}\omega^{k} + ct^{\ell} + \dots) \quad ,$$

where $c \neq 0$ and $\ell \geq 1$. Then

$$(2^{1/3}\omega^{k} + ct^{\ell} + \dots)^3 - 3(2^{2/3}\omega^{2k} + t)(2^{1/3}\omega^{k} + ct^{\ell} + \dots) + (2^{2/3}\omega^{2k} + t)^3 \equiv 0 \ .$$

Expanding and simplifying, the coefficient of t on the left side is

$$-3\cdot 2^{1/3}\omega^{k} + 3\cdot 2^{4/3}\omega^{4k} = 3\cdot 2^{1/3}\omega^{k} \neq 0$$

(this holds even if $\ell=1$). Thus, the other two solutions are branched and we obtain the element

$$e(2^{2/3}\omega^{2k} + t^2,\ 2^{1/3}\omega^{k} + \dots) \quad .$$

<u>Observation</u>: The total branching order is $V=4$ (first order branch points at 0, $2^{2/3}\omega^{2k}$) and $n=3$, so the genus is 0 (recall $V = 2(n+g-1)$).

<u>Another example of an algebraic function</u>.

Let

$$A(z,w) = 2zw^5 - 5w^2 + 3z^2 \quad .$$

Then

$$\frac{\partial A}{\partial w} = 10zw^4 - 10w \quad .$$

Now the critical points are $z=0$, $z=\infty$, and for the others we obtain

$$\frac{\partial A}{\partial w} = 0 = 10w(zw^3-1) \quad .$$

If $w=0$, then $A=0 \Rightarrow z=0$, which we are not now considering. Thus,

$$zw^3=1 \text{ and } A = 0 = 2w^2 - 5w^2 + 3z^2 \ ,$$

so

$$z^2=w^2. \quad \text{Therefore, } z^2w^6=1=w^8 \quad .$$

Thus, if $\omega = e^{2\pi i/8}$, $w = \omega^k$, $0 \le k \le 7$, and $z = \omega^{-3k}$. We still must check $z^2 = w^2 : \omega^{-6k} = \omega^{2k}$, which is valid. So we have found all the critical points, and we now analyze them.

$\underline{z = \omega^{-3k}}$. The only possible multiple value for w is ω^k, and at these points

$$\frac{\partial^2 A}{\partial w^2} = 40zw^3 - 10 = 30 \ne 0.$$

We guess a branch point occurs, so we try for an element $e(\omega^{-3k}+t^2, \omega^k+tQ)$. Then

$$2(\omega^{-3k}+t^2)(\omega^k+tQ)^5 - 5(\omega^k+tQ)^2 + 3(\omega^{-3k}+t^2)^2 \equiv 0 \ .$$

Expanding,

$$\begin{aligned} &2(\omega^{-3k}+t^2)(\omega^{5k}+5\omega^{4k}tQ+10\omega^{3k}t^2Q^2+\ldots) \\ &\qquad -5(\omega^{2k}+2\omega^k tQ+t^2Q^2) + 3(\omega^{-6k}+2\omega^{-3k}t^2+t^4) \equiv 0; \\ &10\omega^k tQ+20t^2Q^2+2\omega^{5k}t^2+10\omega^{4k}t^3Q+\ldots \\ &\qquad -10\omega^k tQ-5t^2Q^2+6\omega^{-3k}t^2+3t^4 \equiv 0; \end{aligned}$$

dividing by t^2,

$$15Q^2 + 8\omega^{5k} + 10\omega^{4k}tQ + \ldots + 3t^2 \equiv 0,$$

where the omitted terms vanish at t=0. At t=0 we can let

$$Q(0) = \sqrt{-\frac{8\omega^{5k}}{15}} \quad \text{(either determination)}$$

and note that at t=0 and for this value of Q(0) the above expression has its derivative with respect to Q equal to

$$30Q \neq 0 \quad .$$

Thus, the implicit function theorem is in force and we obtain branch points

$$e(\omega^{-3k}+t^2,\ \omega^k+\sqrt{-\frac{8\omega^{5k}}{15}}\ t+\ldots),\ 0\le k\le 7.$$

In order to treat the critical points 0 and ∞ we look for meromorphic elements of the form

$$e(t^m, t^{\ell}Q) \quad ,$$

where m and ℓ are integers ($m\neq 0$) and Q is holomorphic near 0, $Q(0)\neq 0$. Then

$$2t^{m+5\ell}Q^5 - 5t^{2\ell}Q^2 + 3t^{2m} \equiv 0 \quad .$$

We now try to juggle m and ℓ to obtain some definite information as $t\to 0$. Thus, we would like to have at least two exponents of t in this equation coincide and to correspond to the dominant terms near t=0. Obviously it is impossible to have all three exponents coincide. Thus, the various possibilities in this case are

$$\text{(a)}\ m + 5\ell = 2\ell < 2m\ ,$$

$$\text{(b)}\ 2\ell = 2m < m + 5\ell\ ,$$

$$\text{(c)}\ m + 5\ell = 2m < 2\ell\ .$$

In case (a) we have $m = -3\ell > \ell$, so $\ell < 0$. Thus, we must have $\ell = -1$, $m = 3$, and the equation for Q becomes

$$2Q^5 - 5Q^2 + 3t^8 \equiv 0\ .$$

Thus, $2Q(0)^3 = 5$ and the derivative with respect to Q is $10Q^4 - 10Q = 15Q \neq 0$ for $Q(0)$. Thus, the implicit function theorem shows we obtain the branch point

$$e(t^3,\ (\tfrac{5}{2})^{1/3}\ \tfrac{1}{t} + \ldots)\ .$$

In case (b) we have $m = \ell < 3\ell$ so $\ell > 0$. Choosing $m = \ell = 1$ gives

$$2t^4Q^5 - 5Q^2 + 3 \equiv 0\ .$$

Again we obtain solutions corresponding to either choice of $Q(0)$ and we get two regular elements

$$e(t,(\tfrac{3}{5})^{1/2}\ t+\ldots),\quad e(t,-(\tfrac{3}{5})^{1/2}\ t+\ldots)\ .$$

In case (c) we have $m = 5\ell < \ell$ so $\ell \leq 0$. Thus, we must $\ell=-1$, $m=-5$, and we obtain

$$2Q^5 - 5t^8Q^2 + 3 \equiv 0\ .$$

Again we obtain the branch point

$$e(t^{-5},\ -(\tfrac{3}{2})^{1/5}\ t^{-1} + \ldots)\ .$$

This completes the analysis of this example except for the observation that the branch point corresponding to $z=\infty$ is of order 4 and thus all five "sheets" of the Riemann surface are branched at ∞ in a single cycle. This proves that A is irreducible.

Notice the total branching order here is $V = 8 + 2 + 4 = 14$, so the genus g satisfies

$$7 = n + g - 1 = 4 + g,$$

or $g=3$.

Problem 7 (p. 139). Analysis of $A(z,w) = zw^3 - 3w + 2z^a$, a any integer. Now $\frac{\partial A}{\partial w} = 3zw^2 - 3$, so critical points other than $z=0,\infty$, come from solving

$$\begin{cases} zw^2 = 1, \\ -2w + 2z^a = 0 \, . \end{cases}$$

So $w=z^a$ and thus $z^{2a+1} = 1$. Let $b=2a+1$ and

$$\omega = e^{2\pi i/b} \, .$$

Then we have the critical points $z = \omega^k$, $0 \le k \le |2a+1| - 1$; the corresponding double value of w is ω^{ak}. Here is a good place to present a criterion for branching: suppose $z_o \neq \infty$ is a critical point for a polynomial A and that $A(z_o,w_o) = \frac{\partial A}{\partial w}(z_o,w_o) = 0$ but $\frac{\partial A}{\partial z}(z_o,w_o) \neq 0$. Then any element $e(z_o+t^m, Q(t))$ in the Riemann surface for A must be a branch point if $Q(0) = w_o$. That is, $m \ge 2$. For suppose $m=1$. Then for t near 0

$$A(z_o+t, Q(t)) \equiv 0.$$

Differentiate this identity with respect to t and set t=0 to obtain

$$0 = \frac{\partial A}{\partial z}(z_o,w_o) + \frac{\partial A}{\partial w}(z_o,w_o)Q'(0) = \frac{\partial A}{\partial z}(z_o,w_o) \neq 0,$$

a contradiction.

In the present case we have $z_o = \omega^k$, $w_o = \omega^{ak}$, and

$$\frac{\partial A}{\partial z}(z_o,w_o) = w_o^3 + 2az_o^{a-1} = \omega^{3ak} + 2a\omega^{(a-1)k} = 0.$$

Here we also have more information. Namely, $\frac{\partial^2 A}{\partial w^2}(z_o,w_o) = 6z_o w_o \neq 0$, so w_o is <u>exactly</u> a double root. Thus, the meromorphic element in this case has m=2, and we can express it as

$$e(\omega^k + t^2, \omega^{ak} + \ldots), \quad 0 \le k \le |2a+1| - 1.$$

To examine the critical points z=0 and ∞ consider elements

$$e(t^m, t^{\ell}Q),\ Q(0) \neq 0.$$

Then

$$t^{m+3\ell}Q^3 - 3t^{\ell}Q + 2t^{am} \equiv 0 .$$

For these exponents to be equal we require $m = -2\ell = -2am$, so bm=0, and thus m=0, which is not allowed.

<u>Case (a)</u>. <u>$m+3\ell = \ell < am$</u>

Here $m=-2\ell$ so we must have $\ell=1$, $m=-2$, and $1+2a<0$, or $\ell=-1$, $m=2$, and $1+2a>0$. We obtain in either case

$$Q^3 - 3Q + 2t^{|b|} = 0 \quad ,$$

and we have the branch point

$$e(t^{\pm 2}, \sqrt{3}\, t^{\mp 1} + \ldots) \quad \begin{cases} \text{top signs if } a \geq 0 \\ \text{bottom signs if } a < 0 \end{cases} .$$

Case (b). $m+3\ell = am < \ell$

Here $3\ell = (a-1)m$ and $m+2\ell<0$. The equation is

$$Q^3 - 3t^{-m-2\ell}Q + 2 = 0 \ ,$$

so

$$Q(0)^3 = -2 \ .$$

We note that $0>3m+6\ell = 3m+(2a-2)m = bm$. We can take $m=\mp 1$ if and only if $a\equiv 1 \pmod 3$, and we thus obtain smooth solutions only:

$$e(t^{\mp 1}, -2^{1/3}t^{\mp\frac{a-1}{3}} + \ldots) \quad \begin{cases} \text{top signs if } a \geq 0 \\ \text{bottom signs if } a < 0 \end{cases} ,$$

where we use all three determinations of $2^{1/3}$.

If $a \not\equiv 1 \pmod 3$, we must choose $m=\mp 3$ and we have a branch point of order 2.

Case (c). $\ell = am < m+3\ell$

Thus, $am<m+3am$, or $bm>0$. We can take $m=\pm 1$, $\ell=\pm a$ and obtain the smooth solution

$$e(t^{\pm 1}, \tfrac{2}{3}t^{\pm a} + \ldots) \ .$$

Summary.

If $a \geq 0$, $z = 0$ corresponds to a first order branch point,
$z = \infty$ corresponds to a second order branch point if $a \not\equiv 1 \pmod 3$, to no branch point if $a \equiv 1 \pmod 3$.

If $a < 0$, $z = \infty$ corresponds to a first order branch point,
$z = 0$ corresponds to a second order branch point if $a \not\equiv 1 \pmod 3$, to no branch point if $a \equiv 1 \pmod 3$.

Thus, $V = |2a+1| + 1 + \begin{cases} 2 \text{ if } a \not\equiv 1 \pmod 3 \\ 0 \text{ if } a \equiv 1 \pmod 3, \end{cases}$

so

$$g = \frac{V}{2} - 2 = \begin{cases} a \text{ if } a \not\equiv 1 \pmod 3,\ a \geq 0, \\ a - 1 \text{ if } a \equiv 1 \pmod 3,\ a \geq 0, \\ |a| - 1 \text{ if } a \not\equiv 1 \pmod 3,\ a < 0, \\ |a| - 2 \text{ if } a \equiv 1 \pmod 3,\ a < 0. \end{cases}$$

SOLUTIONS TO FINAL EXAM PROBLEMS

1. $A(z,w) = w^{2a} - 2z^b w^a + 1$

$$\begin{cases} \dfrac{\partial A}{\partial w} = 2aw^{2a-1} - 2az^b w^{a-1} \\[2ex] \dfrac{\partial^2 A}{\partial w^2} = 2a(2a-1)w^{2a-2} - 2a(a-1)z^b w^{a-2} \\[2ex] \dfrac{\partial A}{\partial z} = -\ 2bz^{b-1}w^a \end{cases}$$

Critical points: $z = \infty$ by definition is critical.

Suppose $\frac{\partial A}{\partial w} = 0$ and $A = 0$. Then $w^a = z^b$ so

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

Let $\omega = e^{\frac{\pi i}{b}}$. Then $z = \omega^k$, $0 \le k \le 2b - 1$, and

$$w^a = z^b = \omega^{bk} = (-1)^k.$$

Thus, for each $z = \omega^k$ there are a distinct solutions of $A(z,w) = 0$. Since $\frac{\partial^2 A}{\partial w^2} = 2a^2 w^{2a-2} \neq 0$, we have no more than double roots. And since $\frac{\partial A}{\partial z} = -\ 2bz^{2b-1} \neq 0$, we have a branch point of order 1 (cf. p. 279) associated with each solution w of $A(\omega^k, w) = 0$. Thus, there are a branch points of order 1 lying over each $z = \omega^k$, so the branching associated with these critical points is 2ab.

Now consider $z = \infty$. Look for elements of $\overline{M}$ of the form

$$e(t^{-m}, t^{\ell} Q),$$

where $m > 0$ and $Q(0) \neq 0$. Substituting,

$$t^{2a\ell}Q^{2a} - 2t^{a\ell-bm}Q^2 + 1 \equiv 0.$$

As on p. 280, we have 3 cases:

Case (a). $2a\ell = a\ell - bm < 0$

Here $\ell < 0$ and $a\ell = -bm$. If we choose $m = a$ and $\ell = -b$, we obtain

$$Q^{2a} - 2Q^a + t^{2ab} \equiv 0$$

and thus

$$Q(0)^a = 2.$$

Then we obtain

$$e(t^{-a}, 2^{1/a}t^{-b} + \dots).$$

Since a and b are relatively prime, this is an element of $\overline{M}$.

Case (b). $a\ell - bm = 0 < 2a\ell$

Here $\ell > 0$ and $a\ell = bm$. Choose $m = a$ and $\ell = b$, obtaining

$$t^{2ab}Q^{2a} - 2Q^a + 1 \equiv 0,$$

so that $Q(0)^a = \frac{1}{2}$. Then we obtain

$$e(t^{-a}, 2^{-\frac{1}{a}}t^b + \dots).$$

Case (c). $2a\ell = 0 < a\ell - bm$

Here $\ell = 0$ and $bm < 0$, which is impossible.

Summarizing, at $z = \infty$ we have two branch points, each of order $a - 1$. Thus, $V = 2ab + 2(a-1)$, so

$$ab + a - 1 = 2a + g - 1,$$

or

$$g = ab - a.$$

Second part. $B = z^{2a} - 2w^b z^a + 1.$

The equation $B = 0$ is

$$w^b = \frac{z^{2a}+1}{2z^a} = \frac{z^a+z^{-a}}{2} .$$

Thus, we simply obtain branch points at $z = 0$, $z = \infty$, and where $w = 0$, which is $z^{2a} + 1 = 0$. Since

$$w = \left(\frac{z^a+z^{-a}}{2}\right)^{1/b},$$

the branch points at finite z are of order $b - 1$ and at $z = 0$ or ∞ are of order $b - 1$ as well, since a and b are relatively prime. Thus, $V = 2a(b-1) + 2(b-1)$, so

$$a(b-1) + b - 1 = b + g - 1,$$

$$g = a(b-1) = ab - a.$$

Alternate solution: Solve for w^a:

$$w^a = z^b + \sqrt{z^{2b}-1} \quad \text{(either determination)}.$$

By inspection there are branch points over the roots of $z^{2b} = 1$, each of order 1. This gives 2ba to the total branching. Then near $z = \infty$ we have $w^a \simeq z^b \pm z^b$, or $w^a \simeq 2z^b$ on half the sheets and

$$w^a = z^b - z^b(1 - \tfrac{1}{2}z^{-2b} + \ldots) = \frac{1}{2z^b} + \ldots$$

on the other half. Thus, $w \simeq 2^{1/a}z^{b/a}$ gives a branch point of order $a - 1$ and $w \simeq 2^{-1/a}z^{-b/a}$ of order $a - 1$. So

$$V = 2ba + 2(a - 1),$$

and

$$ab + a - 1 = 2a + g - 1, \quad \text{or } g = ab - a.$$

2. Let n be the degree of A with respect to w. Let S_A be the Riemann surface of A. Let S be the component of $\overline{M}$ which contains all the germs $[f]_a = e(a+t, f(a+t))$, assuming f is rational. By hypothesis, $S \subset S_A$. Clearly, $\pi: S \to \hat{\mathbb{C}}$ is an analytic equivalence, so S is compact. As S_A is connected, $S = S_A$. Thus, $\pi: S_A \to \hat{\mathbb{C}}$ is an analytic equivalence. But π restricted to S_A takes every value n times. Thus, $n = 1$.

Alternate solution: Suppose f is rational: $f(z) = \frac{P(z)}{Q(z)}$ in lowest terms. Let

$$B(z,w) = Q(z)w - P(z).$$

Then $A(z,f(z)) = B(z,f(z)) = 0 \ \forall\, z$. Thus, Lemma 3 on p. 121 implies A and B have a common factor. Since A is irreducible and B is linear in w, this implies that $A = \text{const}\ B$, which shows A has degree 1.

3. We prove something more general. We have $\pi: S \to \hat{\mathbb{C}}$,

taking every value n times. Here $n > 1$, since otherwise π is an analytic equivalence and there are <u>no</u> branch points. Suppose S has branch points $e_1,\ldots,e_m$ and that $\pi(e_1) = \ldots = \pi(e_m) = z_o$. Since $\hat{\mathbb{C}} - \{z_o\}$ is simply connected, the corollary on p. 119 implies there exists a meromorphic function f on $\hat{\mathbb{C}} - \{z_o\}$ such that $A(z,f(z)) = 0$. (Actually, the corollary is stated for regions in $\mathbb{C}$ and holomorphic f, but the generalization to this case is easy.) By familiar estimates, f grows at most like a power of $z - z_o$ as $z \to z_o$. Thus, f is rational and the previous problem implies $n = 1$, a contradiction.

4. Let S be the Riemann surface of $w^3 + z^3 - 1$. Then π and V restricted to S satisfy $V^3 + \pi^3 \equiv 1$. Let

$$V = \frac{1+g}{f} ,$$

$$\pi = \frac{1-g}{f} ,$$

so that $f = \frac{2}{V+\pi}$ and $g = \frac{V-\pi}{V+\pi}$ are meromorphic on S. Then

$$1 = \left(\frac{1+g}{f}\right)^3 + \left(\frac{1-g}{f}\right)^3 = \frac{2+6g^2}{f^3} ,$$

so that

$$g^2 = \frac{1}{6}(f^3-2) .$$

Now let $g_1 = 2\sqrt{6}g$, obtaining

$$g_1^2 = 4(f^3-2).$$

Now apply the argument which appears on p. 256 and p. 262, concluding that S is analytically equivalent to the Riemann surface of the polynomial $w^2 - 4(z^3-2)$. A little care seems to be needed at this step. Namely, we need to know that f takes every value 2 times in order to be able to apply Theorem 2.4 on p. 158. We do this by checking that f takes the value 0 2 times, or that $V + \pi$ takes the value ∞ 2 times. This is easy. The surface S has 3 smooth sheets over ∞, and if $\omega = e^{\pi i/3}$ (cube root of -1), then on these three sheets we have respectively

$$V = \omega\pi(1 - \pi^{-3})^{1/3},$$

$$V = \omega^2\pi(1 - \pi^{-3})^{1/3},$$

$$V = \omega^3\pi(1 - \pi^{-3})^{1/3},$$

where $(1 - \pi^{-3})^{1/3}$ is the principal determination near $\pi = \infty$. Thus, in the first two cases we have

$$V + \pi = \pi[1 + \omega + \ldots]$$

and

$$V + \pi = \pi[1 + \omega^2 + \ldots]$$

and thus $V + \pi$ takes the value ∞ one time on each sheet. On the third sheet

$$V + \pi = \pi[1 - (1-\pi^{-3})^{1/3}]$$

$$= \pi[1 - (1 - \tfrac{1}{3}\pi^{-3} + \ldots)]$$
$$= \frac{1}{3\pi^2} + \ldots ,$$

and thus $V + \pi$ takes the value 0 on this sheet (at the point lying over $z = \infty$). Thus, $V + \pi$ takes the value ∞ exactly 2 times.

Alternate solution: Define

$$F = a\,\frac{1+\pi}{1-\pi} \quad \text{on } S,$$

so also F takes every value 3 times. Then

$$F - F\pi = a + a\pi,$$
$$\pi = \frac{F-a}{F+a} \quad .$$

Thus,

$$V^3 + \frac{(F-a)^3}{(F+a)^3} \equiv 1 \ ;$$

$$((F+a)V)^3 + (F-a)^3 \equiv (F+a)^3 \ ;$$
$$((F+a)V)^3 \equiv 2(3F^2a + a^3) \ .$$

Let

$$G = \frac{(F+a)V}{(24a)^{1/3}} \ .$$

Then

$$24aG^3 \equiv 6aF^2 + 2a^3;$$
$$F^2 = 4G^3 - \frac{a^2}{3} \quad .$$

5. We write the hypothesis in the following way. On a disk $\Delta \subset \mathbb{C}$ are given meromorphic functions f and g such that for certain polynomials $A(z,w)$ and $B(z,w)$,

$$A(z,f(z)) = B(z,g(z)) = 0, \ z\in\Delta.$$

We can assume A and B are irreducible. Let $z_1,\ldots,z_N$ be the critical points of either A or B. Then define for $z \neq z_j$

$$C(z,w) = \prod_{\substack{A(z,\alpha)=0\\ B(z,\beta)=0}} (w-\alpha-\beta).$$

By the usual symmetry argument, C is a polynomial in w with coefficients which are holomorphic functions of $z\in\hat{\mathbb{C}} - \{z_1,\ldots z_N\}$. By the usual estimates, these coefficients have polynomial growth at these exceptional points, and thus are rational functions of z. Obviously, $C(z,f(z)+g(z)) \equiv 0$ for $z\in\Delta$.

To do the second part we use the above formula. The required polynomial is therefore

$$C(z,w) = (w-z^{1/2}-z^{1/3})(w+z^{1/2}-z^{1/3})(w-z^{1/2}-\omega z^{1/3})$$
$$\times (w+z^{1/2}-\omega z^{1/3})(w-z^{1/2}-\omega^2 z^{1/3})(w+z^{1/2}-\omega^2 z^{1/3}),$$

where $\omega = e^{2\pi i/3}$ and $z^{1/2}$ and $z^{1/3}$ are any values of the roots. After multiplying all these terms together we are bound to get a polynomial. Here is the arithmetic: take the terms #1,3,5 together and likewise #2,4,6 to obtain

$$C(z,w) = [(w-z^{1/2})^3-z][(w+z^{1/2})^3-z]$$
$$= [w^3-3z^{1/2}w^2+3zw-z^{3/2}-z]$$
$$\times [w^3+3z^{1/2}w^2+3zw+z^{3/2}-z]$$

$$= (w^3+3zw-z)^2 - (3z^{1/2}w^2+z^{3/2})^2$$

$$= w^6 + (6z-9z)w^4 + (-2z)w^3 + (9z^2-6z^2)w^2 - 6z^2w + z^2 - z^3$$

$$= w^6 - 3zw^4 - 2zw^3 + 3z^2w^2 - 6z^2w + z^2 - z^3 .$$

.a. Parabolic. We check Proposition 15.2 of p. 204. If S is the compact Riemann surface and $p \in S$, let D be an analytic disk in $S - \{p\}$, and let u be a bounded continuous nonnegative function in $S - \{p\} - D$ which is harmonic in $S - \{p\} - D^-$ and $\equiv 0$ on ∂D. Since u is bounded near p, u has a unique extension to a harmonic function in $S - D^-$. As $S - D$ is compact, the maximum principle holds and implies that $\sup_{S-D} u = \sup_{\partial D} u = 0$, so $u \le 0$. Thus, $u \equiv 0$.

b. Hyperbolic. Choose a nonconstant function f which is continuous and real-valued on the boundary of the analytic disk in question. By Proposition 13 on p. 197, there exists a harmonic function u in the Riemann surface minus the closed disk, continuous up to the boundary, where it equals f. Moreover, u is bounded. Since f is not constant, u is not constant. Thus, u is a bounded nonconstant subharmonic function, and we apply Proposition 15.1 of p. 204.

c. Hyperbolic. By definition (p. 218) there is a point $p \in S$ and a function g on $S - \{p\}$ satisfying the conditions of Definition 7 of Chapter VI. Since

$g \to \infty$ as one approaches p, if A is a sufficiently large constant the function $u = \min(g, A)$ is superharmonic and not constant. In fact, u is superharmonic on S since $u \equiv A$ near p. As $0 < u \leq A$, u is also bounded. Apply Proposition 15.1 of Chapter VI.

REFERENCES

Ahlfors, L., and Sario, L., Riemann Surfaces, Princeton Univ. Press, Princeton, N. J., 1960.

Behnke, H., and Sommer, F., Theorie der analytischen Funktionen einer komplexen Veranderlichen, 2nd ed., Springer, Berlin, 1962.

Bers, L., Riemann Surfaces, Courant Institute of Mathematical Sciences, New York University, N. Y., 1958.

Chevalley, C., Introduction to the Theory of Algebraic Functions of One Variable, Amer. Math. Soc., Providence, R. I., 1951.

Cohn, H., Conformal Mapping on Riemann Surfaces, McGraw-Hill, 1967.

Favard, J., Cours d'analyse de l'École Polytechnique, Vol. 3, Gauthier-Villars, Paris, 1962.

Gunning, R., Lectures on Riemann Surfaces, Princeton Univ. Press, Princeton, N. J., 1966.

Heins, M., Complex Function Theory, Academic Press, N. Y., 1968.

Hurwitz, A., and Courant, R., Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen. Geometrische Funktionentheorie, 4th ed., Springer, Berlin, 1964.

Knopp, K., Theory of Functions, Vol. 2, Dover, N. Y., 1947.

Mackey, G., Lectures on the Theory of Functions of a Complex Variable, Van Nostrand, Princeton, N. J., 1967.

Markushevich, A., Theory of Functions of a Complex Variable, Vol. 3, Prentice-Hall, Englewood Cliffs, N. J., 1967.

Narasimhan, M., et al, Riemann Surfaces, Tata Inst. of Fundamental Research, Bombay, 1963.

Osgood, W. F., Lehrbuch der Funktionentheorie, Vol. 1, 5th ed., Teubner, Leipzig, 1928.

Pfluger, A., Theorie der Riemannschen Flächen, Springer, Berlin, 1957.

Riemann, B., Collected Works, Dover, N. Y., 1953.

Springer, G., Introduction to Riemann Surfaces, Addison-Wesley, Reading, Mass., 1957.

Stoïlow, S., Leçons sur les Principes Topologiques de la Théorie des Fonctions Analytiques, Gauthier-Villars, Paris, 1938.

Veech, W., A Second Course in Complex Analysis, Benjamin, N. Y., 1967.

Weyl, H., The Concept of a Riemann Surface, 3rd ed., Addison-Wesley, Reading, Mass., 1964.