THE STRUCTURE OF OPEN CONTINUOUS MAPPINGS
HAVING TWO VALENCES

A. K. LYZZAIK AND KENNETH STEPHENSON

ABSTRACT. The authors study open continuous functions which map the unit disc to compact Riemann surfaces and which assume each value in the range space (with a finite number of exceptions) either p or q times for some positive integers p, q. Although the questions here originated in efforts to understand mapping properties of locally univalent analytic functions, the authors remove analyticity assumptions and show that the underlying issues are topological and combinatoric in nature. The mappings are studied by embedding their image surfaces in compact covering spaces, a setting which allows the consideration of fairly general ranges and which accommodates branch and exceptional points. Known results are generalized and extended; several open questions are posed, particularly regarding the higher dimensional analogues of the results.

INTRODUCTION

The work presented here is part of the effort to understand the valences associated with functions analytic on the open unit disc D. The class of functions we consider, those having essentially two valences, seems to have had its beginning with Brannan and Kirwan [3], who proved that there are no locally one-to-one analytic functions on the disc which assume every value in the plane either p or $p - 1$ times.

Some of the open questions left by Brannan and Kirwan were resolved by Lyzzaik and Styer [8], who introduced the idea of embedding the images of these functions in compact surfaces. Meromorphic functions were treated by Srebro [12] and functions between Riemann surfaces by Srebro and Wajnryb [13, 14]. In [10], Ortel and Smith considered functions with valences p and 1 rather than p and $p - 1$. Meanwhile, Brannan and Lyzzaik [4] started the study of functions with valences p and q and began to allow for exceptional points with other valences. In that paper and (independently) the paper of Carne, Ortel, and Smith [5], the conjecture in [10] that the Brannan-Kirwan result was true with $p - 1$ replaced by any q, $0 < q < p$, was proven. With the exception of a part of the work in [8], all these results concerned locally univalent functions.

In trying to understand the accumulated results, we were led to a broader class of mappings, the class of "(p, q)-maps" which we describe below. We allow branch points and exceptional points, and our functions take values in arbitrary compact surfaces rather than just the Riemann sphere. Working in this setting we not only establish the earlier results and others in a unified way, but in fact reduce the questions to purely combinatoric considerations. To a large extent, this is what one wants to do with all valence questions, since, despite the pathologies of the individual functions, valence questions have a definite "combinatoric" flavor. We feel, then, that this work provides the correct framework in which to address essentially all questions on two valent functions, and, at the same time, it provides some insights and techniques which may be useful in further valence studies.

Now let us define the (p, q)-maps and some of the associated terminology and then outline the organization of the paper.

Definition. *Let p and q be integers, $0 < q < p$ and let $\mathscr{X}$ be a compact Riemann surface without boundary. A continuous open mapping f of the open unit disc into $\mathscr{X}$ which has a finite number of branch points is termed a (p, q)-map if every point in $\mathscr{X}$, with the possible exception of a finite subset, is assumed either p or q times, counting multiplicities.*

Being discrete, open, and continuous, these maps have local behavior which is topologically equivalent to that of analytic functions. In particular, the notion of branch point applies in the usual way, so we can define the *valence function* of f, counting multiplicities at branch points, by

$$\nu_f(x) = \operatorname{card}(f^{-1}\{x\}), \qquad x \in \mathscr{X}.$$

In accordance with our definition, $\mathscr{X}$ is the disjoint union

$$\mathscr{X} = \mathscr{P} \cup \mathscr{Q} \cup \mathscr{E},$$

where the points of $\mathscr{P}$ are taken p times, those of $\mathscr{Q}$ are taken q times, and $\mathscr{E}$ is the remaining *exceptional set*. The usual branch order is associated with each branch point of f; the *total branch order* of f, to be denoted β_f, is the sum of these. A quantity termed the *deficiency* of f is associated with each exceptional point e of f,

$$\delta(e) = q - \nu_f(e);$$

the *total deficiency* of f is the sum of these and will be denoted δ_f. We will show, incidently, that $\delta(e) \geq 1$.

The main lines of our work are reflected in three results. The first is an embedding theorem, which puts us in the geometric setting so important to our work.

Theorem. *Given a (p, q)-map $f: D \to \mathscr{X}$, there exists a p-fold compact covering space $(\widetilde{\mathscr{X}}, \pi)$ of $\mathscr{X}$ and an embedding $\phi: D \to \widetilde{\mathscr{X}}$ so that $f \equiv \pi \circ \phi$.*

This gives us a quite explicit picture of the image surface of f as a disc embedded in $\widetilde{\mathscr{X}}$ which we can begin to manipulate. In particular, we introduce certain restricted homotopies within the class of (p, q)-maps, termed (p, q)-*homotopies*, which allow us to modify this embedded disc without changing the underlying valence structure. The term *valence structure* will be defined later, but in particular, these homotopies preserve total branch order and total deficiency. The main result of the paper is that we can replace a general (p, q)-map by one with the same valence structure but having the greatest possible regularity.

Main Theorem. *Let* $f\colon D \to \mathscr{X}$ *be a* (p, q)-*map. Then* f *is* (p, q)-*homotopic to a simplicial map* h *between a triangulation of* $\overline{D}$ *and a triangulation of* $\mathscr{X}$.

We exploit the combinatorics of simplicial maps—Euler characteristics, the Riemann-Hurwitz relation, etc.—to obtain information on the valence structures of general (p, q)-maps. For instance, we obtain this inequality.

Theorem. *Let* $f\colon D \to \mathscr{X}$ *be a* (p, q)-*map. Then*

$$\beta_f + \delta_f \geq p + q - 1 - 2qg(\mathscr{X}),$$

where $g(\mathscr{X})$ *is the genus of* $\mathscr{X}$. *The inequality is sharp in the case of genus* 0 *or* 1.

To illustrate its use, consider an analytic function f which assumes each value in the finite complex plane either p or q times, as in [3, 4, 5], or [10]. Here $g(\mathscr{X}) = 0$ since $\mathscr{X}$ is the Riemann sphere, ∞ is the only exceptional point, and $\delta_f = \delta(\infty) = q$. We conclude that $\beta_f + q \geq p + q - 1$, hence $\beta_f \geq p - 1$; in particular, there exist no locally univalent functions with these properties.

Now concerning the organization of the paper: Our outlook is largely geometric, so we begin with some concrete examples which may aid the reader's intuition during subsequent proofs. In §2, we present necessary preliminaries concerning topology and covering spaces. The main work begins in §3, where we prove the embedding theorem. Though the embedding technique goes back to Lyzzaik-Styer [8] and Brannan-Lyzzaik [4], our treatment is essentially self-contained and stresses the underlying topological features; particularly helpful is a certain "depth" function whose use simplifies earlier arguments. In §4 we define our first type of (p, q)-homotopy, with considerable effort being required to establish the requisite regularity of $\mathscr{C}$. In §5 we introduce triangulations, we apply another type of (p, q)-homotopy to reach the simplicial maps, and we establish the inequality given above. §6 contains applications, along with several examples which illustrate both the uses and the limitations of our techniques. For example, Srebro-Wajnryb [13] use embeddings for functions defined on multiply connected domains, but only in the case $q = p - 1$; we will point out the special nature of this situation. We conclude with some open questions and directions for further work in §7. Among the new notions we introduce in

this paper, we feel that those involving the depth function and the homotopies which preserve valence structures have the most potential for use in other valence studies.

Both authors were visitors at the Open University, Milton Keynes, England, when their collaboration started; we thank the mathematics faculty for their generous hospitality.

1. EXAMPLES

It is helpful to have some geometric images to aid the intuition in the work to follow, so we begin by constructing four examples. Example 1 will be familiar to all analysts and exemplifies a class of functions which will need to be handled separately throughout much of our work—the class BQ defined in the next section.

Example 1. Let b be a meromorphic function mapping the unit disc onto the Riemann sphere which is given as a quotient of a p-fold finite Blaschke product and a q-fold finite Blaschke product. Explicitly, suppose

$$b(z) = \frac{\prod\limits_{j=1}^{p} \frac{z-\alpha_j}{1-\overline{\alpha}_j z}}{\prod\limits_{k=1}^{q} \frac{z-\beta_k}{1-\overline{\beta}_k z}} \ , \qquad z \in D \ ,$$

where $\{\alpha_1, \alpha_2, \ldots, \alpha_p\}$ and $\{\beta_1, \beta_2, \ldots, \beta_q\}$ are disjoint sets of points in D. This function belongs to a well-known family which maps the unit circle to itself. We make the additional assumption that none of the branch values of b lie on the unit circle. In this case (and only in this case) every value $w \in D$ is assumed p times, and every value $w \in C^\infty \backslash D$ (including $w = \infty$) is assumed q times by b in D, counting multiplicities. That is, $\mathscr{X} = C^\infty$, $\mathscr{P} = D$, $\mathscr{Q} = C^\infty \backslash D$, and $\mathscr{E} = \varnothing$. Of course, b is a rational function of degree $p + q$, so it has total branch order $2p + 2q - 2$. By symmetry about the unit circle, half this branching occurs in D, so $\beta_b = p + q - 1$.

Typically, the domain of b will look something like Figure 1, where the curves mapping to ∂D are indicated and the points mapping to $\mathscr{P}$ are shaded. One can verify that b maps ∂D in a positive direction $p - q$ times about the unit circle and maps the curves inside D a total of q times about the unit circle. □

Example 2. Suppose $\mathscr{X}$ is a torus and $\pi: C \to \mathscr{X}$ is a universal covering map with the square $\{z: 0 \le \Re z < 1, \ 0 \le \Im z < 1\}$ as fundamental domain. For any integer $n \ge 2$, let ϕ be a conformal map of D onto $\{z: 0 < \Re z < n, \ 0 < \Im z < n\}$ and define the function

$$f \equiv \pi \circ \phi : D \to \mathscr{X}.$$

We claim that f is a (p, q)-map: It is analytic, hence open and continuous. Let γ_1 and γ_2 be the simple closed curves

$$\pi(\{x : 0 \le x \le 1\}) \quad \text{and} \quad \pi(\{iy : 0 \le y \le 1\}),$$
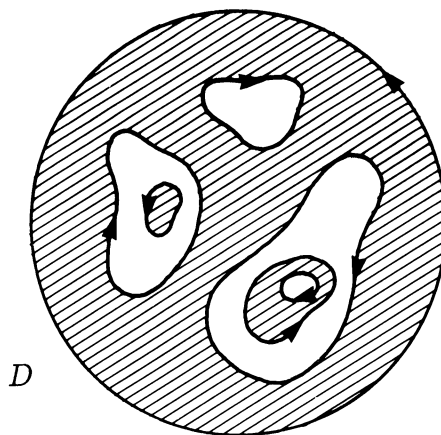
FIGURE 1

respectively, and $e = \gamma_1 \cap \gamma_2 = \pi(0)$. Every point in $\mathscr{P} = \mathscr{X} \backslash (\gamma_1 \cup \gamma_2)$ is assumed n^2 times, every point in $\mathscr{Q} = (\gamma_1 \cup \gamma_2) \backslash \{e\}$ is assumed $n^2 - n$ times, and $\mathscr{E} = \{e\}$ is assumed $n^2 - 2n + 1$ times. $\mathscr{X} = \mathscr{P} \cup \mathscr{Q} \cup \mathscr{E}$, so f is a (p, q)-map with $p = n^2$, $q = n^2 - n$. Note that $\delta_f = n - 1$ and $\beta_f = 0$, giving

$$\beta_f + \delta_f = n - 1 = p + q - 1 - 2qg(\mathscr{X}). \quad \square$$

Example 3. Here, we illustrate a technique for obtaining (p, q)-maps for arbitrary p, q, and $g(\mathscr{X}) \geq 1$. For ease of exposition, we have chosen $p = 3$, $q = 1$, $g(\mathscr{X}) = 1$.

Let $\mathscr{X}$ be the surface in Figure 2, with a distinguished point e. Let F_1 and F_2 be two families of simple closed curves with base point e but otherwise disjoint, each family forming a basis for the fundamental group of $\mathscr{X}$. In Figure 2, F_1 consists of the two loops formed by solid lines, while F_2, the two loops formed by dashed lines. Let $\mathscr{X}_1$, $\mathscr{X}_2$, and $\mathscr{X}_3$ be three new copies of $\mathscr{X}$, each identified with $\mathscr{X}$ in the natural way by the identity projection. We want to remove at least one curve family from each surface $\mathscr{X}_j$, while also ensuring that each curve family is removed from exactly two $(= p - q)$ of the $\mathscr{X}_j$'s. In Figure 2, we have chosen to remove the two loops of F_1 from $\mathscr{X}_1$, the two loops of F_2 from $\mathscr{X}_2$, and all four loops from $\mathscr{X}_3$.

After removing the curve families, the $\mathscr{X}_j$ consist of five simply connected pieces. We will denote them by S_k, $k = 1, 2, \ldots, 5$, though for purposes of illustration, only S_1, S_2, and S_3 are labelled in the figure. We finish the construction by successively attaching these separate pieces to one another to build up a single simply connected surface S. This is a fairly standard "cut and paste" operation, so we will work rather informally. Start by letting $S = S_1$. Identify a point $x_2 \in S_2$ and a simple curve σ_2 from it to e in S_2 which does not intersect any loops of our curve families. The same curve (i.e., the curve with the same projection in $\mathscr{X}$) can be found in S. Cut S and S_2 along σ_2 and identify opposite edges of the cuts on the two surfaces to form
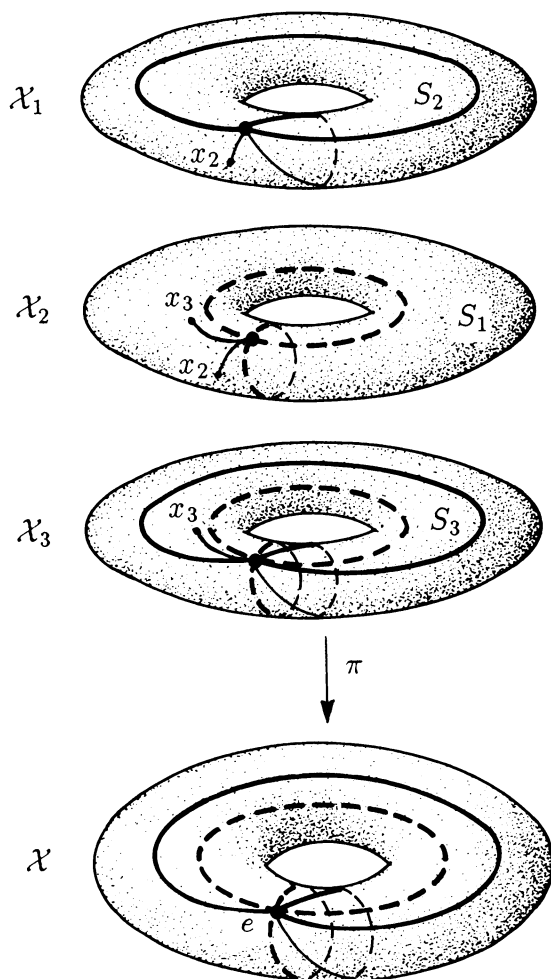
FIGURE 2

a single surface. The result becomes our new S: it is simply connected by van Kampen's theorem, since we have joined two simply-connected sets along a simply connected boundary (see [9, pp. 113–115]); it has a natural projection to $\mathscr{X}$ inherited from the two pieces; and our pasting results in a simple branch point at x_2 under this projection. We do the same with a point $x_3 \in S_3$ and σ_3: cut S and S_3, paste them together with a simple branch point at x_3, and obtain yet a new simply-connected surface S. Continuing in this way, we find that we have built a simply connected surface S which incorporates all 5 surface fragments. Under the natural projection, call it $\pi: S \to \mathscr{X}$, each point of $\mathscr{X} \backslash (F_1 \cup F_2)$ is covered 3 times; each point of $(F_1 \cup F_2) \backslash \{e\}$ is covered once; and e is omitted. Since S is simply connected, there is a homeomorphism $\phi: D \to S$. The map $f \equiv \pi \circ \phi: D \to \mathscr{X}$ has image surface S and is the desired $(3, 1)$-map. Observe that $\beta_f = 4$ and $\delta_f = 1$, giving $\beta_f + \delta_f = 5$. Incidentally, we will be referring back to this construction and adding features to the accompanying picture when we introduce fattened q-sets. $\square$
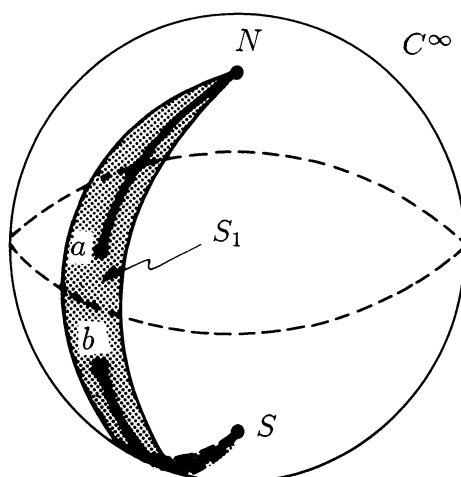
FIGURE 3

Example 4. In this example, as in the previous one, we obtain f by explicitly constructing its image surface. Here $\mathscr{X}$ is the Riemann sphere, $p = 3$, $q = 2$, $\mathscr{E}$ contains the north and south poles, and $\mathscr{B}$ has two points a and b. The image surface S is constructed from the open set $S_1 \subseteq C^\infty$ pictured in Figure 3 and two additional copies of C^∞, denoted S_2, S_3. Points in S_1 are the p-set, while the remaining points of C^∞, aside from the north and south poles, comprise the q-set. Slit S_1 and S_2 along the arc from a to the north pole; cross connect them along this arc, introducing a simple branch point above a. Do the same thing with S_1 and S_3 along the arc from b to the south pole. The resulting surface is simply connected. If $f: D \to \mathscr{X}$ is a function with image surface S, its total branch order is $\beta_f = 2$. The north and south poles are each covered only once, giving $\delta_f = 2$. □

2. NOTATION AND PRELIMINARIES

We will be working with open continuous maps of finite valence, $f: D \to \mathscr{X}$, where D is the open unit disc and $\mathscr{X}$ is a compact Riemann surface. Since orientation is not an issue, we make the assumption throughout the sequel that all maps are orientation preserving. They will be treated as ramified covering maps in the spirit of Ahlfors and Sario [1], where the notions one usually associates with analytic functions are still available: branch points, local inverses, the monodromy theorem, path liftings, etc. (Indeed, by a theorem of Stoïlow [15], f will actually be analytic if we put an appropriate conformal structure on D. However, with analyticity we would lose some useful flexibility.)

We assume familiarity with the concepts of covering surfaces as defined in Chapter I, §§1–3 of Ahlfors and Sario [1] and use that notation and terminology whenever possible. In particular, suppose $(\widetilde{F}, f)$ is a covering surface for F. If $\widetilde{\Omega}$ is an open subset of $\widetilde{F}$, then $\widetilde{\Omega}$ is a *complete covering* for $\Omega \subseteq F$ if each point $x \in \Omega$ has a compact neighborhood N, all of whose preimages

$f^{-1}(N) \cap \tilde{\Omega}$ are compact in $\tilde{\Omega}$. Such a complete covering has some number n of "sheets", meaning each point in Ω is covered the same number n of times by $\tilde{\Omega}$, counting multiplicities. Paths in Ω lift uniquely and completely to paths in $\tilde{\Omega}$ if one specifies the starting location and if the lift does not encounter any branch points.

Suppose $(\tilde{F}, f)$ is a covering surface of a Riemann surface F. The valence function associated with f is defined as follows, where multiplicities are counted,

$$\nu_f(x) = \operatorname{card}(f^{-1}\{x\}), \qquad x \in F,$$

and is known to be lower semicontinuous. (Note that the valence 0 is allowed.) We would like to introduce a companion notion: Let $\rho(\cdot, \cdot)$ be a Riemannian metric on F. For $x \in F$ and $r > 0$, denote by $\Delta(x, r)$ the open disc $\{y \in F : \rho(x, y) < r\}$. Since f is a covering map, one can show that any precompact component (i.e., a component having compact closure) of $f^{-1}(\Delta(x, r))$ is a complete covering of $\Delta(x, r)$.

Definition. *Suppose n is a positive integer. Define the n-depth function $r_n(x)$ at a point $x \in F$ to be the supremum of those $r > 0$ such that $f^{-1}(\Delta(x, r))$ contains complete coverings of $\Delta(x, r)$ having a total of at least n sheets.*

It is not difficult to see that $r_n(x)$ is nonnegative, Lipschitz in x, and monotone decreasing in n. Moreover, $r_n(x) > 0$ if and only if $f^{-1}\{x\}$ has cardinality at least n, counting multiplicities; indeed, $\nu_f(x) = \sup\{n : r_n(x) > 0\}$. We will use this depth function several times to obtain a certain metric uniformity in sets we consider, always assuming that the metric ρ is the spherical, Euclidean, or hyperbolic metric appropriate to F.

Proposition 1. *Let $(\tilde{F}, f)$ be a covering surface of a Riemann surface F, and let S be a subset of F consisting of points which are covered by f the same finite number of times n, counting multiplicities. Then S has an open neighborhood $\mathscr{V}$ in F such that for every component V of $\mathscr{V}$ the set $f^{-1}(V)$ contains complete coverings of V whose total number of sheets is n. If S is open, then $\mathscr{V}$ may be chosen as S itself.*

Proof. Define the open set $\mathscr{V}$ as the union of discs

$$\mathscr{V} = \bigcup_{y \in S} \Delta\left(y, \frac{r_n(y)}{3}\right).$$

Each disc in this union has precompact components of its preimage containing n sheets. Let $\widetilde{\mathscr{V}}$ be the union of these precompact open sets in $\tilde{F}$. Let V be a component of $\mathscr{V}$ and $\tilde{V}$ a component of $\widetilde{\mathscr{V}}$ for which $f(\tilde{V}) \cap V \neq \varnothing$. We will show that $(\tilde{V}, f)$ is a complete covering surface of V. Since values in S are taken only n times, the complete coverings of V in $f^{-1}(V)$ must have exactly n sheets.

Since $f(\tilde{V})$ is connected and contained in $\mathscr{V}$, it lies in V; so $(\tilde{V}, f)$ is a covering surface of V. To prove completeness, it is only necessary to show

that every point $y \in V$ has a neighborhood N with the property that each component of $f^{-1}(N) \cap \widetilde{V}$ is compact. For $y \in F$, let $\Delta_y = \Delta(y, r_n(y)/3)$ and $\Delta'_y = \Delta(y, r_n(y))$. Now fix a $y_0 \in V$. Let $N \subseteq \Delta_{y_0}$ be a closed neighborhood. If $f^{-1}(N) \cap \widetilde{V} = \varnothing$, there is nothing to prove; so assume C is a nonempty component of $f^{-1}(N) \cap \widetilde{V}$. By our construction, there is $y_1 \in V$ and a precompact preimage of Δ_{y_1} in $\widetilde{V}$, which we denote by $\widetilde{\Delta}_{y_1}$, intersecting C. Then $\Delta_{y_0} \cup \Delta_{y_1} \subseteq \Delta'_{y_j}$ for either $j = 0$ or $j = 1$. Now, a slightly smaller disc in Δ'_{y_j} has n precompact preimages in $\widetilde{F}$ which are complete covers of it; let $\widetilde{\Delta}'_{y_j}$ be the one which contains $\widetilde{\Delta}_{y_1}$. This contains only precompact preimages of Δ_{y_0}, hence only compact preimages of N. The precompact preimage of Δ_{y_0} which intersects C is one of the discs going into the definition of $\widetilde{\mathscr{V}}$ and hence will be in $\widetilde{V}$. We conclude that C is compact. Therefore, $(\widetilde{V}, f)$ is a complete covering surface of V, as desired.

In case S is open, then the n preimages of S in the set $\widetilde{\mathscr{V}}$ will be complete coverings of S under f, so we may replace $\mathscr{V}$ by S itself. This completes the proof. □

We will make use of the Riemann-Hurwitz relation, which on compact bordered surfaces takes the following form (see [9, pp. 33–34]). Here and in the sequel, $g(\cdot)$ denotes *genus*.

Lemma 1. *Let F and $\widetilde{F}$ be compact bordered orientable surfaces with m and n boundary components, respectively. Suppose that $(\widetilde{F}, f)$ is a k-sheeted complete covering of F which has branch points with total order $\beta_f < \infty$ in $\widetilde{F}$ and no branch points on $\partial\widetilde{F}$. Then*

$$\beta_f = 2[(g(\widetilde{F}) - 1) - k(g(F) - 1)] + n - km.$$

In particular, if $\widetilde{F}$ is planar, then F is planar and

$$\beta_f = (n - 2) - k(m - 2) \leq 2k - 2.$$

We will work frequently with path liftings. First, note that a simple closed curve γ in a Riemann surface $\mathscr{X}$ will bound a topological disc if and only if it is null homotopic in $\mathscr{X}$. For later reference, we have this lemma.

Lemma 2. *If γ is a null homotopic simple closed curve in a compact surface $\mathscr{X}$, then $\mathscr{X} \setminus \gamma$ has exactly two components, a topological disc and an open surface of genus $g(\mathscr{X})$, and γ is the boundary of each.*

Suppose now that (D, f) is a covering surface of a compact Riemann surface $\mathscr{X}$ and let $\gamma \subseteq \mathscr{X}$ be a simple closed curve which avoids the branch values of f. Standard lifting arguments establish the following dicotomy regarding components of $f^{-1}(\gamma)$: Either the component is compact, in which case it is (the trace of) a simple closed curve $\tilde{\gamma}$, and the restriction of f to $\tilde{\gamma}$ is a k-fold

smooth covering of γ for some integer k; or the component is not precompact, in which case it accumulates at points of the unit circle ∂D.

Let us establish some notation which we use in the next proposition and at several points in the sequel. With f, $\mathscr{X}$, and γ as above, assume $f^{-1}(\gamma)$ has $0 \leq s < \infty$ compact components in its preimage, to be denoted $\tilde{\gamma}_1, \tilde{\gamma}_2, \ldots, \tilde{\gamma}_s$. For each $j = 1, 2, \ldots, s$ let k_j be the number of times $\tilde{\gamma}_j$ covers γ; write $k = \sum_{j=1}^{s} k_j$. Being disjoint and closed, the $\tilde{\gamma}_j$'s separate D into $s+1$ components, $C_0, C_1, \ldots, C_s$. We will agree to denote by C_0 the one having the unit circle in its boundary.

Proposition 2. *Let f, $\mathscr{X}$, and s be as described above and assume $s > 0$. Then*

(a) *$\mathscr{X} \setminus \gamma$ has two components, Ω_1, Ω_2, of which at least one is a topological disc.*

(b) *Each (C_j, f), $j = 1, 2, \ldots, s$, is a complete covering surface of either Ω_1 or Ω_2.*

(c) *If $g(\mathscr{X}) > 0$, then only one, say Ω_1, is a topological disc; the components C_j, $j = 1, 2, \ldots, s$, are topological discs; and Ω_1 has the property that each of its points has exactly k preimages in $\bigcup_{j=1}^{s} C_j$, counting multiplicities.*

(d) *If $g(\mathscr{X}) = 0$, then both Ω_1 and Ω_2 are topological discs. With the added hypothesis that $f(C_0) \subseteq \Omega_2$, Ω_1 has the property that each of its points has exactly k preimages in $\bigcup_{j=1}^{s} C_j$, counting multiplicities.*

Proof. (a) If $s > 0$, then $f^{-1}(\gamma)$ has at least one compact component, $\tilde{\gamma}_1$. This is null homotopic in D, so γ^{k_1} is null homotopic in $\mathscr{X}$. There are no nontrivial nilpotents in the fundamental group of a surface, so γ is itself null homotopic. By Lemma 2, we see that γ separates $\mathscr{X}$ and that one of its complementary components, say Ω_1, is a disc, while the other has $g(\Omega_2) = g(\mathscr{X})$.

(b) All noncompact components of $f^{-1}(\gamma)$ lie in C_0, since they have accumulation points in ∂D. Thus each of the connected sets $f(C_j)$, $j = 1, 2, \ldots, s$, lies entirely in Ω_1 or Ω_2. Fix j and assume $f(C_j) \subseteq \Omega_n$. (C_j, f) is clearly a covering map of Ω_n; it extends continuously to ∂C_j; and it carries the boundary to the boundary of Ω_n. In particular, f is a proper mapping of C_j to Ω_n, so the preimages in C_j of compact subsets of Ω_n are compact. We conclude that (C_j, f) is a complete covering surface of Ω_n. Note that ∂C_j consists of certain of the curves $\tilde{\gamma}_j$, say $\tilde{\gamma}_{j_1}, \tilde{\gamma}_{j_2}, \ldots, \tilde{\gamma}_{j_m}$. This means f is an M-to-1 map of ∂C_j onto γ, where $M = \sum_{l=1}^{m} k_{j_l}$. We conclude that (C_j, f) is an M-sheeted covering of Ω_n.

(c) When $g(\mathscr{X}) > 0$, $g(\Omega_2) = g(\mathscr{X}) > 0$, so Ω_2 is not planar. This means it is not a disc, but it also implies by the last part of Lemma 1 that the images $f(C_j)$, $j = 1, 2, \ldots, s$, must lie in Ω_1. If two of these sets where contiguous, they would not map to the same component of $\mathscr{X} \setminus \gamma$; therefore the components

C_j are simply connected. By the note at the end of the previous paragraph, each point of Ω_1 has $k = \sum_{j=1}^{s} k_j$ preimages in $\bigcup_{j=1}^{s} C_j$.

(d) When $g(\mathscr{X}) = 0$, $\mathscr{X}$ is the Riemann sphere, so both Ω_1 and Ω_2 are discs by the Jordan curve theorem. Assume that $f(C_0) \subseteq \Omega_2$. Along with the observation that contiguous components C_j cannot be mapped to the same component of $\mathscr{X} \setminus \gamma$, this tells us that each curve $\tilde{\gamma}_m$, $m = 1, 2, \ldots, s$, is in the boundary of precisely one of the other components C_j, $1 \leq j \leq s$, which gets mapped to Ω_1. By the note at the end of (b) above, $\tilde{\gamma}_m$ contributes k_m to the count of times that f covers Ω_1 on C_j. Therefore, each point of Ω_1 is covered k times in total on $\bigcup_{j=1}^{s} C_j$, as claimed.   □

The term "image surface" for an open, continuous, discrete function f means the same, thanks to Stoïlow's theorem, as it does for analytic functions; namely, it is the Riemann surface for the inverse function f^{-1}; that is, the surface on which a local inverse of f, restricted to taking values in D, can be continued in a single-valued and well-defined way. If $f: D \to \mathscr{X}$ and $(\widetilde{\mathscr{X}}, \pi)$ is a covering surface of $\mathscr{X}$, then the image surface of f is embedded in $\widetilde{\mathscr{X}}$ if there is a one-to-one map $\tilde{f}: D \to \widetilde{\mathscr{X}}$ with $f \equiv \pi \circ \tilde{f}$. We will use the following lemma, whose proof is a standard application of the monodromy theorem to $h^{-1} \circ f$.

Lemma 3. *Suppose $(\widetilde{\mathscr{X}}, \pi)$ is a covering surface of $\mathscr{X}$ and that $f, h: D \to \mathscr{X}$ are open, continuous, discrete, orientation preserving maps having the same image surface embedded in $\widetilde{\mathscr{X}}$. Then there is an orientation preserving homeomorphism η of D with $h \circ \eta \equiv f$.*

3. THE EMBEDDING THEOREM

Henceforth, p and q will denote fixed integers, $0 < q < p$, and f will be a (p, q)-map of the unit disc D into the compact Riemann surface $\mathscr{X}$. Our aim is Theorem 1, which says that the image surface of f can be embedded in a compact p-sheeted covering surface $(\widetilde{\mathscr{X}}, \pi)$ of $\mathscr{X}$.

Recall that ν_f is the *valence function* of f, and that the definition of (p, q)-maps implies that $\mathscr{X}$ is the disjoint union

$$\mathscr{X} = \mathscr{P} \cup \mathscr{Q} \cup \mathscr{E},$$

where $\mathscr{P} = \{x \in \mathscr{X} : \nu_f(x) = p\}$ is the p-*set*, $\mathscr{Q} = \{x \in \mathscr{X} : \nu_f(x) = q\}$ is the q-*set*, and $\mathscr{E}$ is the remaining *exceptional set*. We assume throughout this work that $\mathscr{P}$ and $\mathscr{Q}$ are nonempty and that $\mathscr{E}$ is finite, possibly empty. The sets $f^{-1}(\mathscr{P})$ and $f^{-1}(\mathscr{Q})$ in D are denoted by $\widetilde{\mathscr{P}}$ and $\widetilde{\mathscr{Q}}$, respectively. If P is a component of $\mathscr{P}$, then $\tilde{P}$ will denote a component of $f^{-1}(P) \subseteq \widetilde{\mathscr{P}}$; likewise for $\mathscr{Q}$.

The set $\mathscr{B} \subseteq \mathscr{X}$ consists of *branch values*, i.e., the images under f of its branch points, and is finite by hypothesis. A branch point for f is a point

$z \in D$ at which f fails to be locally one-to-one. By a theorem of Stoïlow [15], at such a point f behaves locally like the map $z \to z^m$ for some positive integer m; m is the *multiplicity* and $m - 1$ is the *order* of f at z. The *total order* of f is the sum over D of all its orders and will be denoted β_f.

The exceptional set $\mathscr{E}$ consists of points $e \in \mathscr{X}$ for which $\nu_f(e)$ is neither p nor q. Indeed, we will prove shortly that $\nu_f(e) < q$.

Definition. *The* deficiency *at* $e \in \mathscr{E}$, $\delta(e)$, *and the* total deficiency, δ_f, *are given by*

$$\delta(e) = q - \nu_f(e), \qquad \delta_f = \sum_{e \in \mathscr{E}} \delta(e).$$

In our later construction, branch points may be placed over certain points of the exceptional set.

A considerable percentage of our work in this paper will center on special neighborhoods of $\mathscr{Q}$ known as *fattened q-sets*.

Definition. *An open neighborhood $\mathscr{V}$ of $\mathscr{Q}$ is termed a* fattened *q-set if each of its components V meets $\mathscr{Q}$ and has the property that $f^{-1}(V)$ contains complete coverings of V whose total number of sheets is q.*

Applying Proposition 1 with $S = \mathscr{Q}$, there exists a fattened q-set $\mathscr{V}$. Our main task in this section, after some preliminaries, is to modify $\mathscr{V}$ to obtain a fattened q-set having additional regularity. (Later, even more regularity will be demanded.)

Preliminary results. First we establish important properties of $\mathscr{P}$, $\mathscr{Q}$, and $\mathscr{E}$. For our purposes, a *continuum* is a compact connected subset of a Riemann surface.

Lemma 4. *Let $f: D \to \mathscr{X}$ be a (p, q)-map, with $\mathscr{X} = \mathscr{P} \cup \mathscr{Q} \cup \mathscr{E}$, as above. Then*

    (a) *$\mathscr{P}$ is an open subset of $\mathscr{X}$ whose components P are simply connected in $\mathscr{X}$.*

    (b) *$\mathscr{Q} \cup \mathscr{E}$ is a continuum in $\mathscr{X}$.*

    (c) *If $e \in \mathscr{E}$, then $\nu_f(e) < q$.*

Proof. (a) Since f is an open map, ν_f is lower semicontinuous. In particular, $\mathscr{P}$ is open in $\mathscr{X}$ and $\widetilde{\mathscr{P}} = f^{-1}(\mathscr{P})$ is open in D. Because the points of $\mathscr{E}$ are isolated, $\nu_f(e) < p$, $e \in \mathscr{E}$, so the maximum valence of f is p. (Note that f would be described as "p valent" in the usual terminology.)

Apply Proposition 1 to the open set $S = \mathscr{P}$. Each of its components P has a preimage containing complete coverings of P whose total number of sheets is p. Since no point of $\mathscr{X}$ is taken more than p times by f, these coverings comprise all of $f^{-1}(P)$. Write

$$f^{-1}(P) = \widetilde{P}_1 \cup \widetilde{P}_2 \cup \cdots \cup \widetilde{P}_m,$$

where each $\widetilde{P}_j$ is a k_j-sheeted covering of P and $\sum_{j=1}^m k_j = p$.

Now apply Proposition 2. Let γ be an arbitrary simple closed curve in P. Its preimage, $f^{-1}(\gamma)$, has only compact components in the sets $\widetilde{P}_j$, and these cover it p times. Using the notation preceding Proposition 2, $f^{-1}(\gamma) = \bigcup_{j=1}^{s} \tilde{\gamma}_j$. Since $f^{-1}(\gamma) \cap C_0 = \varnothing$, the extra hypothesis in 2(d) even holds. Thus we can apply 2(c) or 2(d) to conclude that $\mathscr{X} \setminus \gamma$ has a simply connected component Ω which is bounded by γ and is covered p times by f. Thus $\Omega \subseteq \mathscr{P}$, so γ is null homotopic in P. Therefore, P is simply connected.

(b) There are countably many components of $\mathscr{P}$, denoted $\{P_n\}_{n=1}^{\infty}$. For each n, let $\{P_{n,k}\}_{k=1}^{\infty}$ be a nested sequence of precompact Jordan subdomains of P_n whose union is all of P_n. (Here we take $P_n = \varnothing = P_{n,k}$ if the number of components is finite and less than n.) Define

$$X_k = \mathscr{X} \setminus \left(\bigcup_{n=1}^{k} P_{n,k} \right), \qquad k = 1, 2, \dots .$$

The sequence $\{X_k\}_{k=1}^{\infty}$ is a decreasing sequence of continua, each a compact bordered subsurface of $\mathscr{X}$. Since

$$\mathscr{Q} \cup \mathscr{E} = \bigcap_{k=1}^{\infty} X_k,$$

and $\mathscr{Q} \cup \mathscr{E}$ is closed, it is a continuum. (Note: this proves that $\mathscr{Q} \cup \mathscr{E}$ contains the *cluster set* of f on ∂D. See [11].)

(c) As $\mathscr{E}$ is finite, its points are isolated. If $\nu_f(e) > q$, then a punctured neighborhood of e lies in $\mathscr{P}$, contradicting the fact that components of $\mathscr{P}$ are simply connected. Since $\nu_f(e) \neq q$, we must have $\nu_f(e) < q$. $\square$

There is a class of (p, q)-maps (actually, the most well behaved) which will need to be handled separately in much of the sequel.

Definition. *The class of (p, q)-maps with empty exceptional sets is termed BQ.*

The archetype of this class is the quotient of finite Blaschke products of Example 1. In fact, we will be showing that all BQ-maps are essentially such Blaschke quotients. In the rest of this subsection we establish necessary and sufficient conditions for BQ along with some properties of these maps.

Lemma 5. *Let P be a component of $\mathscr{P}$, γ a simple closed curve in P, and A the component of $P \setminus \gamma$ sharing a boundary with ∂P. If A lies in a fattened q-set $\mathscr{V}$, then $f \in BQ$.*

Proof. Without loss of generality we may assume that γ separates the branch values of P from $\mathscr{Q}$. Then A is a topological annulus, one boundary being $\partial P \subseteq (\mathscr{Q} \cup \mathscr{E})$ and the other γ. Since it is connected, A lies in a component V of $\mathscr{V}$. For convenience, we replace V by the component of $V \setminus \gamma$ which intersects $\mathscr{Q}$.

From the definition of fattened q-sets, V has complete coverings $(\widetilde{V}_j, f)$, $j = 1, 2, \dots, k$, of total multiplicity q, while A, being a subset of $\mathscr{P}$, has

complete coverings $(\tilde{A}_i, f)$, $i = 1, 2, \ldots, s$, of total multiplicity p. Because A contains no branch values, each $\tilde{A}_j$ is an annulus having one boundary component which is a simple closed curve covering γ. Denote this boundary component by $\tilde{\gamma}_j$ and observe that it covers γ the same number of times that $\tilde{A}_j$ covers A. The $\tilde{\gamma}_j$'s are mutually disjoint. Let $W = \bigcup_{j=1}^{k} \tilde{V}_j$. The coverings $\tilde{A}_j$ break into two classes, those which lie in W and cover A a total of q times and those which are disjoint from W and cover A the remaining $(p-q)$ times.

We will adhere to the notation established before Proposition 2. It might be helpful to refer to Figure 1, which shows the approximate situation in the domain D. Note, however, that because γ separates $P \cap \mathscr{B}$ and $\mathscr{Q}$, C_0 actually turns out to be an annulus in the current situation. Let $\tilde{\gamma}_1$, $\tilde{\gamma}_2$, ..., $\tilde{\gamma}_n$ be the curves which lie in ∂C_0, where $1 \le n \le s$. The curves $\tilde{\gamma}_{n+1}$, ..., $\tilde{\gamma}_s$ lie in $D \setminus \overline{C}_0$; suppose that they cover γ a total of r times. Let Ω_1, Ω_2 denote the components of $\mathscr{X} \setminus \gamma$ with $\Omega_1 \subseteq P$ and $\mathscr{Q} \subseteq \Omega_2$. By Proposition 2(b), each of the components C_1, C_2, ..., C_s is a complete covering of either Ω_1 or Ω_2, with the image alternating as one moves between contiguous C_j's. Since Ω_1 has compact closure in P, its preimage in D also has compact closure in D; so C_0 must be mapped by f into Ω_2. This has several consequences: Applying Proposition 2(c) and (d), every point of Ω_2 is covered r times in $\bigcup_{j=1}^{s} C_j$; and $\mathscr{Q} \subseteq \Omega_2$, so $r \le q$. Also, the sets $\tilde{V}_j$ which lie in $\bigcup_{j=1}^{s} C_j$ are not contiguous to C_0, so they have the curves $\tilde{\gamma}_{n+1}$, ..., $\tilde{\gamma}_m$ as boundary components, and $\tilde{A}_1$, ..., $\tilde{A}_n$ lie in C_0.

Now we show that $r = q$. Suppose $r < q$. On the one hand, this forces at least one of the sets $\tilde{V}_j$ to lie in C_0. On the other hand, at least one of the annuli $\tilde{A}_1$, ..., $\tilde{A}_n$ is not in W; otherwise the points of $(\partial A \cap \partial P) \subseteq (\mathscr{Q} \cup \mathscr{E})$ would be covered at least $p - r$ times in C_0 and r times in $\bigcup_{j=1}^{s} C_j$. Suppose for convenience that $\tilde{V}_1 \subseteq C_0$ and that $\tilde{A}_1 \cap W = \varnothing$. Since $\tilde{V}_1$, $\tilde{A}_1 \subseteq C_0$ are disjoint and C_0 is connected, $\partial \tilde{A}_1 \cap C_0$ contains infinitely many points. Choose $x_0 \in \partial \tilde{A}_1 \cap C_0$ with $y_0 = f(x_0)$ not an exceptional point. Now $\partial \tilde{A}_1 \cap C_0$ is mapped by f to ∂P, so $y_0 \in (\partial P \setminus \mathscr{E}) \subseteq \mathscr{Q}$. Also, $y_0 \notin \partial V$ since $\partial V \subseteq (\mathscr{P} \cup \mathscr{E})$. Since $y_0 \in \partial A \setminus \partial V$ and $A \subseteq V$, we conclude that $y_0 \in V$. Thus y_0 is covered by f a total of q times in W plus at least once by $x_0 \notin W$. This contradicts the fact that $y_0 \in \mathscr{Q}$. Therefore, $r = q$.

With $r = q$, we see that $\bigcup_{j=1}^{s} C_j$ covers every point of Ω_1 exactly p times and every point of Ω_2 exactly q times. We conclude that $\mathscr{E} = \varnothing$, so $f \in BQ$, completing the proof. $\square$

The converse of the previous result also holds. In particular, if P is a component of $\mathscr{P}$, then ∂P is a compact subset of $\mathscr{Q} \cup \mathscr{E} = \mathscr{Q}$; so any fattened q-set, being open, will have a curve γ and annulus A as in the hypothesis of Lemma 5. Additional properties of these BQ-maps are gathered in this lemma.

Lemma 6. *Let $f : D \to \mathscr{X}$ be in BQ. Then:*

   (a) *$\mathscr{P}$ has a single component,*

   (b) *$\mathscr{X}$ is the Riemann sphere,*

   (c) *There exists a fattened q-set $\mathscr{U}$ having a single simply connected component, and*

   (d) *$\beta_f = p + q - 1$.*

Proof. As observed above, the hypotheses of the previous lemma hold, so we will continue with the notation from that proof.

(a) If P' were another component of $\mathscr{P}$, then $P' \subseteq \Omega_2$ so P' is covered exactly q times by $\bigcup_{j=1}^{s} C_j$. Its preimage would then necessarily intersect C_0. But it is easy to see that $f(C_0) \subseteq A \subseteq P$. Otherwise C_0 would contain a point mapping to $\partial A \cap \mathscr{Q}$, and these points are already covered q times in the complement of C_0.

(b) Consider any component C_j not contiguous to C_0: It is a complete covering of Ω_2, so by Lemma 1, Ω_2 is planar and by Lemma 2, $g(\mathscr{X}) = g(\Omega_2) = 0$. Therefore, $\mathscr{X}$ is the Riemann sphere.

(c) The open set Ω_2 provides such a fattened q-set. Note also, that by our construction, Ω_2 contains no branch values from $\mathscr{P}$.

(d) We use the Riemann-Hurwitz relations. Since C_0 is connected, $f(C_0) \subseteq A$, and A contains none of the branch values of f, we conclude that $n = 1$ and $C_0 = \tilde{A}_1$. Break the other components $C_1, C_2, \ldots, C_s$ into two classes and renumber them so $\{C_1, \ldots, C_u\}$ are the ones covering Ω_1 and $\{C_{u+1}, \ldots, C_s\}$ are the ones covering Ω_2. For $j = 1, \ldots, s$ let c_j and t_j denote, respectively, the number of boundary components of C_j and the number of times C_j covers the corresponding Ω_i. The following equalities hold:

$$s = \sum_{j=1}^{u} c_j = 1 + \sum_{j=u+1}^{s} c_j \; , \quad \sum_{j=1}^{u} t_j = p \; , \quad \text{and} \quad \sum_{j=u+1}^{s} t_j = q.$$

Ω_1 and Ω_2 are simply connected and all the branch points of f are in $\bigcup_{j=1}^{s} C_j$, so using the above equalities and Lemma 1, we conclude

$$\beta_f = \sum_{j=1}^{u} (c_j + t_j - 2) + \sum_{j=u+1}^{s} (c_j + t_j - 2) = p + q - 1.$$

This completes the proof of (d) and of the lemma. $\quad\square$

Regular fattened q-sets. For our embedding theorem, we need to prove existence of a fattened q-set with the special properties of this proposition. We say that an open set in $\mathscr{X} \setminus \mathscr{E}$ is *simply connected relative* to $\mathscr{X} \setminus \mathscr{E}$ if every component of its complement in $\mathscr{X}$ contains a point of $\mathscr{E}$.

Proposition 3. *There exists a fattened q-set $\mathscr{U}$ whose components are simply connected relative to $\mathscr{X} \setminus \mathscr{E}$ and which does not have any branch values of f which lie in $\mathscr{P}$.*

Proof. If $f \in BQ$, then the result follows from Lemma 6(c). Therefore, we may assume that $f \notin BQ$.

Let $\mathscr{V}$ be a fattened q-set. We will fix a component V of $\mathscr{V}$ and show that by adjoining components of the complement $\mathscr{X} \setminus V$ we can obtain a larger open set U which has q complete lifts to D, yet remains disjoint from the other components of $\mathscr{V}$. At each stage, then, U along with these other components forms a new fattened q-set. Handling each component of $\mathscr{V}$ in turn, we finally obtain a new fattened q-set $\mathscr{U}$, all of whose components are simply connected relative to $\mathscr{X} \setminus \mathscr{E}$. As the last step, we will remove at most a finite number of simple arcs to excise branch values in $\mathscr{P} \cap \mathscr{U}$.

Let K be the union of components of $\mathscr{X} \setminus V$ which contain points of $\mathscr{E}$ and define $U = \mathscr{X} \setminus K$. We need to show that U has complete lifts to D totaling q sheets. Since there are only q complete lifts of V, it suffices to prove that if $\widetilde{U}$ is a component of the preimage of U under f which intersects one of the complete lifts of V, then $(\widetilde{U}, f)$ is a complete covering of U.

First, we consider a few facts about our situation: We have $U \setminus V \subseteq \mathscr{P}$. Suppose $x \in (U \setminus V) \cap \mathscr{Q}$; the component Q of $\mathscr{Q}$ containing x is disjoint from V since $x \notin V$, hence Q is separated from $\mathscr{E} \subseteq (\mathscr{X} \setminus U)$ by V. This contradicts Lemma 4(b). Also, each component $E \subseteq U \setminus V$ is planar because components of $\mathscr{P}$ are planar and simply connected because V is connected. Being compact, E lies in some component $P \subseteq \mathscr{P}$, and since P is simply connected we may separate E from $\mathscr{E}$ by a simple closed curve $\gamma \subseteq (P \cap U)$. Applying Proposition 2, the component Ω_1 of $\mathscr{X} \setminus \gamma$ containing E will be covered p times by precompact topological discs $C_j \subseteq D$. Since $\partial C_j \subseteq f^{-1}(V)$ and $C_j \subseteq \widetilde{\mathscr{P}}$, it is clear that $C_j \cap f^{-1}(V)$ is connected. The upshot is that any curve $\tau \subseteq C_j$ running, say, between points a and b of $C_j \cap f^{-1}(V)$ is fixed endpoint homotopic in C_j to a curve τ' running from a to b in $f^{-1}(V)$.

Now, suppose that x is a point of $U \cap \mathscr{P}$ and choose $r > 0$ so that $\Delta = \Delta(x, \frac{r}{2}) \subseteq \Delta(x, r)$ and that $\Delta(x, r) \subseteq U \cap \mathscr{P}$ is precompact. Consider a component $\Omega \neq \varnothing$ of $f^{-1}(\Delta)$ which intersects $\widetilde{U}$. Then Ω lies in a complete lift of $\Delta(x, r)$ which, being connected, lies entirely in $\widetilde{U}$. In particular, $\Omega \subseteq \widetilde{U}$ is precompact, as desired.

Next, suppose $x \in U \cap \mathscr{Q}$, which by the comments above lies in $V \cap \mathscr{Q}$, and choose r and Δ as above, only with $\Delta(x, r) \subseteq V$ rather than $\Delta(x, r) \subseteq U \cap \mathscr{P}$. Again, let $\Omega \neq \varnothing$ be a component of $f^{-1}(\overline{\Delta})$ which intersects $\widetilde{U}$. If $\Omega \cap \widetilde{U}$ contains a point of a complete lift $\widetilde{V}$ of V, then as before, $\Omega \subseteq \widetilde{V} \subseteq \widetilde{U}$ is precompact. The only difficulty, then, is in proving that $\Omega \cap \widetilde{U}$ contains a point of $\widetilde{V}$. Since V has q complete lifts to D, it is enough to show that Ω lies in one of these; and for this it is enough to find a curve σ from some point $a \in \Omega \cap \widetilde{U}$ to some point b in one of the complete lifts of V for which the image $f(\sigma)$ lies in V. Recall that $\widetilde{U}$ intersects one of these lifts, call it $\widetilde{V}$, so there is a curve μ in $\widetilde{U}$ from a to $b \in \widetilde{V} \cap \widetilde{U}$. We may assume, without loss of

generality, that $f(\mu)$ lies in $\mathscr{P}$, since if it hits a point over $\mathscr{Q}$, that point will automatically be in one of the complete lifts and we could stop the path there; also, we may assume $b \in \widetilde{V} \setminus \mathscr{Q}$, since $\widetilde{V}$ is an open neighborhood of $\mathscr{Q} \cap U$. If μ cuts through one of the components of $f^{-1}(U \setminus V)$, then as described above, that part of the path can be moved by a homotopy in $f^{-1}(U)$ to lie in $\widetilde{V}$. By compactness, we need make only a finite number of such adjustments to obtain a path σ which is fixed endpoint homotopic to μ, but which lies in $\widetilde{V}$. Therefore $a \in \widetilde{V} \cap \Omega$ and this completes the proof that $(\widetilde{U}, f)$ is a complete covering of U. Now let V' be a component of $\mathscr{V}$ other than V and let U' be its corresponding U. Obviously $V' \cap \mathscr{Q} \neq \varnothing$, and by Lemma 4(b) $\partial V \cap \partial V' \cap \mathscr{E} \neq \varnothing$. It follows that V' is contained in some component of K, namely one whose complement contains U; hence $U \cap U' \neq \varnothing$. This completes the construction of a new fattened q-set having components which are simply connected relative to $\mathscr{X} \setminus \mathscr{E}$; we will denote this by $\mathscr{U}$.

The final step in our proof is to excise branch values of f which lie in $\mathscr{P} \cap U$. Observe that any component Σ of the intersection of components $P \subseteq \mathscr{P}$ and $U \subseteq \mathscr{U}$ will be simply connected relative to $\mathscr{X} \setminus \mathscr{E}$. Let b be a point of $\mathscr{B} \cap \Sigma$. If $P \cap \partial \Sigma = \varnothing$, then $P \subseteq \Sigma \subseteq U$, implying by Lemma 5 that $f \in BQ$, contradicting our assumption at the beginning of the proof. Therefore $P \cap \partial \Sigma$ contains some point a. Run a simple curve τ from b to a in P and replace U by $U \setminus \tau$. This gives us a new fattened q-set whose components remain simply connected relative to $\mathscr{X} \setminus \mathscr{E}$. Doing this for each of the finite number of points $b \in \mathscr{B}$ completes our construction. $\square$

Definition. *A fattened q-set $\mathscr{U}$ whose components are simply connected relative to $\mathscr{X} \setminus \mathscr{E}$ and do not contain branch values lying in $\mathscr{P}$ will be called a* regular fattened q-set.

We remark that in the case that $g(\mathscr{X}) > 0$, one can show using Proposition 2 that each component U of $\mathscr{U}$ is a topological disc; that is, all points of $\mathscr{E}$ lie in a single component of the complement of U. This is not necessarily the case when $\mathscr{X}$ is the Riemann sphere. For instance, the fattened q-set for the function in Example 4 is the sphere with arcs from a and b to the north and south poles removed; it is not simply connected, but each component of its complement contains an exceptional point.

Construction of $\widetilde{\mathscr{X}}$. It can be difficult to follow a construction of this type without some geometric picture to fall back upon. Therefore, the reader may prefer to look ahead to the next subsection, where we have provided a concrete illustration based on Example 3.

Our fundamental operation, loosely known as "pasting", is a well-known folk technique to those who work frequently with surfaces. For the sake of completness, we give a formal statement.

Lemma 7. *Let (R, g) be a covering surface of the Riemann surface $\mathscr{X}$, R not necessarily connected. Suppose $\rho: \Omega_1 \to \Omega_2$ is a homeomorphism between*

disjoint open subsets Ω_1 *and* Ω_2 *of* R *with the property that* $g \circ \rho \equiv g$ *on* Ω_1. *Furthermore, suppose that for any sequence* $\{\omega_j\} \subseteq \Omega_1$ *converging to the ideal boundary of* Ω_1, *either* $\omega_j \to \partial R$ *or* $\rho(\omega_j) \to \partial R$. *Then there exists a covering surface* (S, h) *of* $\mathcal{X}$, *unique up to homeomorphisms, and a covering map* $\psi : R \to S$ *so that* $g \equiv h \circ \psi$ *and* $\psi \circ \rho \equiv \psi$, *with* $\psi(x) = \psi(y)$ *for* $x, y \in R$ *if and only if* $x = y$, $x = \rho(y)$, *or* $\rho(x) = y$.

In other words, one can factor the covering (R, g) through an intermediate covering (S, h) in which the sets Ω_1 and Ω_2 have been identified. The proof follows from standard arguments regarding complex manifolds: One establishes an equivalence relation on R by saying that $x_1 \sim x_2$ if $x_1 \in \Omega_1$, $x_2 \in \Omega_2$, and $x_2 = \rho(x_1)$. Note that since ρ is one-to-one and since Ω_1 and Ω_2 are disjoint, the equivalence class $[x]$ is a singleton set if $x \notin \Omega_1 \cup \Omega_2$ and a two element set if $x \in \Omega_1 \cup \Omega_2$. The topological space S is the set of equivalence classes with the identification topology and ψ is the identification map. Perhaps surprisingly, the only difficulty is in showing that S is Hausdorff; this is where the hypothesis on sequences is needed. From there, straightforward arguments show that S is a complex manifold, that ψ is a covering map, and that the other conclusions of the lemma hold. The details are left to the interested reader.

Observe the following feature, which we need shortly: If $K \subseteq S$ is a connected subset of S and if a component of $\psi^{-1}(K) \subseteq R$ is compact, then K is compact.

Theorem 1. *Let* $f : D \to \mathcal{X}$ *be a* (p, q)-*map. Then there exists a* p-*fold compact covering surface* $(\widetilde{\mathcal{X}}, \pi)$ *of* $\mathcal{X}$ *and a one-to-one open continuous map* $\phi : D \to \widetilde{\mathcal{X}}$ *so that* $f \equiv \pi \circ \phi$.

Proof. Throughout the proof, we denote by U a component of $\mathcal{U}$, a regular fattened q-set, P a component of $\mathcal{P}$, and C a component of their intersection. We will be applying the previous lemma; however, as earlier, we will handle BQ separately. Assuming, then, that $f \notin BQ$, $\mathcal{E}$ is nonempty and U and P are simply connected relative to $\mathcal{X} \setminus \mathcal{E}$, so C is a topological disc.

In applying Lemma 7, the manifold R is the disjoint union of D and $p - q$ copies of each of the components U of $\mathcal{U}$. The covering map g is defined to be f on D and to be the identity map on each copy of a component U, giving the covering surface (R, g) of $\mathcal{X}$, as desired.

Now to define the map ρ of the lemma: First we consider a component C of $\mathcal{U} \cap \mathcal{P}$. As indicated, C is a disc. Since it lies in some P, there are p complete lifts of C to D, but only q of these lie in complete lifts of U. Since $U \cap P$ has no branch values of f and since C is simply connected, the monodromy theorem implies that the remaining $p - q$ lifts of C are disjoint open sets in D, each homeomorphic to C under f (i.e., under the restriction of g to D). Label the lifts $C_1, C_2, \ldots, C_{p-q}$. Now R contains $p - q$ disjoint copies of U, let us denote them $U_1, U_2, \ldots, U_{p-q}$, each of which contains an

open disc $\widehat{C}_j$ homeomorphic to C under the identity (i.e., under the restriction of g to U_j). Define $\rho : C_j \to \widehat{C}_j$ in the obvious way, namely, $\rho \equiv g^{-1} \circ g$, for $j = 1, \ldots, p - q$. Continue in this way for all components of $\mathcal{U} \cap \mathcal{P}$ to complete the definition of ρ. It is evident that ρ is one-to-one and that $g \circ \rho \equiv g$. The domain Ω_1 of ρ is the open set in D whose components are the $p - q$ lifts of components such as C. To verify the hypothesis of the lemma regarding sequences, consider any sequence $\{\omega_k\} \subseteq \Omega_1$ which converges to an ideal boundary point of Ω_1. Then from some index on, all elements of the sequence lie in a single set C_j and converge to a point $z_0 \in \partial C_j$. If $z_0 \in \partial D$, then we are done; therefore, assume $z_0 \in D$. Note that $\{f(\omega_k)\} \to f(z_0)$. If $f(z_0) \in U$, then $f(z_0)$ has q preimages in the interiors of the q complete lifts of U and, necessarily distinct from those, the preimage z_0 itself. This implies $f(z_0) \in \mathcal{P}$. But then $f(z_0) \in U \cap P$, contradicting the fact that $f(z_0)$ is a boundary point of the component C of $U \cap P$. Therefore, $f(z_0) \in \partial U$. But then we see that $\rho(\omega_k)$ converges to ∂U_j. Thus, our sequence hypothesis is verified. Let (S, h) be the covering space of $\mathcal{X}$ guaranteed by Lemma 7, $\psi : R \to S$ the identification map. Note that S is connected, since every copy U_j of a component U which was included in our definition of R will have points in the range of ρ, hence points identified with the disc $D \subseteq R$. Furthermore, note that the unit disc D is (identified with) a subset of R with $\Omega_1 \subseteq D$ while $\Omega_2 = \rho(\Omega_1)$ is disjoint from D; so the restriction of ψ to D is one-to-one.

Now, we consider the covering properties of h. Let $X = \mathcal{X} \setminus (\mathcal{E} \cup \mathcal{B})$. Consider an arbitrary point $y \in X$. Suppose $y \in U$, a component of $\mathcal{U}$, and let N be a compact connected neighborhood, $y \in N \subseteq (X \cap U)$. The fact that $h \circ \psi \equiv g$ implies that $h^{-1}(N) = (\psi \circ g^{-1})(N)$. The q complete lifts of U to D contain q compact connected preimages of N in $g^{-1}(N)$; there are a further $p - q$ compact connected preimages in the copies of U which are part of R. All points of $g^{-1}(N)$ are identified by ψ with points in one of these compact sets; therefore, the observation made after Lemma 7 implies that the preimage of N under h is compact in S. Now consider a point $y \in X \setminus \mathcal{U}$. Let N be a compact connected neighborhood, $y \in N \subseteq (\mathcal{P} \setminus \mathcal{Q})$. Then N has p disjoint compact preimages in D under g, and again all points of $g^{-1}(N)$ are identified by ψ with points in one of these compact sets. So the preimage of N under h is compact in S. The upshot is the following: Let $W = S \setminus h^{-1}(\mathcal{E} \cup \mathcal{B})$. Then the restriction of h to W is a complete smooth covering map of X. Moreover, each point of X is covered by h exactly p times. As the sets $h^{-1}(\mathcal{E} \cup \mathcal{B})$ and $\mathcal{E} \cup \mathcal{B}$ consist of isolated points, there is a compactification of the covering surface (W, h) which forms a p-fold ramified covering surface of $\mathcal{X}$ (see [1, Chapter I, 20E]). In particular, there is a compact covering surface $(\widetilde{\mathcal{X}}, \pi)$ of $\mathcal{X}$ and an imbedding $\tau : S \to \widetilde{\mathcal{X}}$ so that $h \circ \tau \equiv \pi$. Let $\phi : D \to \widetilde{\mathcal{X}}$ be the inclusion map of D in S followed by τ. This is one-to-one and satisfies the conclusions of the theorem, so we

are done in this case. Note that our final compactification may result in new branch points, but only over points of $\mathscr{E}$; for the restriction of W to each of the added copies of a component U is one-to-one.

For $f \in BQ$, things are only slightly different. First, we will use the set Ω_2 from the proof of Lemma 6(c) as the fattened q-set. Recall that $\mathscr{P}$ has only a single component P which intersects U in a topological annulus A. Designate a point $z_0 \in \mathscr{E}$ and let (U_0, g_0) be a complete branched covering of U which is smooth except for a branch point of order $p - q - 1$ at a point over z_0. In particular, note that U_0 is simply connected and that each point of U is covered $p - q$ times by U_0. Let R be the disjoint union of D and U_0 and g the map to $\mathscr{X}$ defined to be f on D and g_0 on U_0. Now, there are q complete lifts of U to D containing, of course, q complete lifts of A, counting multiplicities. The remaining lift must be a topological annulus, since A contains no branch values. Indeed, in the notation of the proof of Lemma 5, one can verify that this lift is precisely the set denoted C_0 and that it covers A the remaining $p - q$ times. Define $\rho\colon C_0 \to U_0$ by $\rho(a) = (g_0^{-1} \circ f)(a)$. This is a homeomorphism of C_0 onto a topological annulus A_0 in U_0 and clearly has the property that $g \circ \rho \equiv g$ on C_0. Recall that A has one of its boundary components in P and the other in U. The verification of the sequence condition in Lemma 7 is essentially the same as the previous case, so we omit it. Lemma 7 now gives a covering surface (S, h) of $\mathscr{X}$, with S again connected. The restriction of h to $S \setminus (h^{-1}(\mathscr{E} \cup \mathscr{B} \cup \{z_0\})$ is a complete smooth p-fold covering map of $X = \mathscr{X} \setminus (\mathscr{E} \cup \mathscr{B} \cup \{z_0\})$. We continue exactly as before to obtain the result. Unlike the previous case, here we do add branch points over U. Also note that the surface $\widetilde{\mathscr{X}}$ is the Riemann sphere.

This completes the proof of the embedding theorem. $\square$

Concrete illustration. We go back to the construction in Example 3 to illustrate the geometry involved in the embedding theorem. There, of course, we started with a compact surface and cut out the curves and points corresponding to $\mathscr{C}$ and $\mathscr{E}$ to obtain an image surface. Here we want to do exactly the opposite, starting with an image surface and replacing material to obtain a compact surface. Figure 4 is the same as Figure 2, only we have drawn in one component U of the standard fattened q-set (shaded). It surrounds one of the components of $\mathscr{C}$ which was lifted to two of the three copies of $\mathscr{X}$; in particular, the lift of U to the image surface consists of one component which is a complete covering, labelled $\widetilde{U}_1$, along with four pieces of the partial coverings, labelled $\widetilde{U}_2, \ldots, \widetilde{U}_5$. To pull these pieces together during compactification, we attach two homeomorphic copies of U to our image surface. For instance, we may attach one of these to $\widetilde{U}_2$ and $\widetilde{U}_3$ and the other to $\widetilde{U}_4$ and $\widetilde{U}_5$. (This seems natural only because of the way our image surface is drawn; one could equally well attach one copy to $\widetilde{U}_2$ and $\widetilde{U}_5$ and the other to $\widetilde{U}_4$ and $\widetilde{U}_3$.)

Pastings of this type for all the components of $\mathscr{U}$ ensure that every point of $\mathscr{X}$ is covered p times, with the exception of e. The final compactification
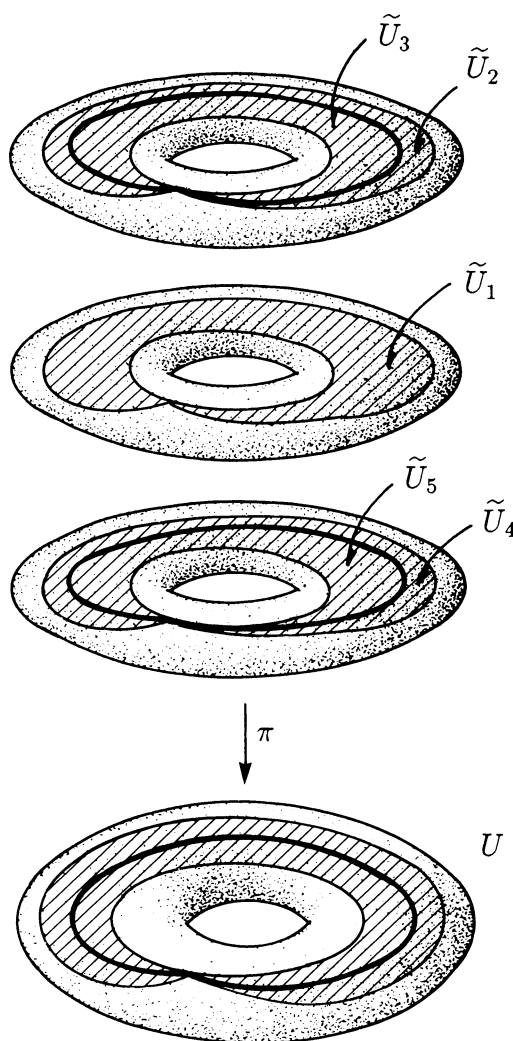
FIGURE 4

step is to fill in this isolated singularity in the canonical way, using the fact that it has a punctured neighborhood with a p-fold complete covering.

4. $\mathscr{U}$-HOMOTOPIES

With the embedding theorem in hand, the image surfaces of (p, q)-maps are available in a rather tame geometric setting, as discs embedded in compact Riemann surfaces. Needless to say, these embedded discs can be pathological; yet, one has the feeling that the complications are not essential, that by careful manipulation of such an embedded disc—pushing in here, pulling out there— one should be able to get a more regular embedding with essentially the same covering properties. This is exactly what we do by means of (p, q)-homotopies. For convenience, we describe the properties we want to preserve as the "valence structure":

Definition. *Two (p, q)-maps, f and g, will be said to have the same* valence structure *if there is a one-to-one correspondence between their branch points, corresponding points having the same branch orders, as well as a one-to-one correspondence between their exceptional points, corresponding points having the same deficiencies.*

Note in particular that $\beta_f = \beta_g$ and $\delta_f = \delta_g$ and that f and g have the same valence sequences.

Definition. *A (p, q)-homotopy is a homotopy $\Lambda: D \times [0, 1] \to \mathscr{X}$ with the property that all maps $\Lambda(\cdot, t): D \to \mathscr{X}$, $t \in [0, 1]$, are (p, q)-maps with the same valence structure.*

In this section we develop the first and most difficult of our (p, q)-homotopies, termed $\mathscr{U}$-homotopies; others will be introduced in the next section. We emphasize that the homotopies we use form only a very restrictive class suited to our needs. They clearly define an equivalence relation within the collection of (p, q)-maps; however, we do not develop a general homotopy theory. It does seem that this idea might well be pursued for the study of functions with other valences or for its own interest.

Throughout this section, $f: D \to \mathscr{X}$ is a (p, q)-map and $\mathscr{U}$ is a regular fattened q-set. The general situation is slightly different than that for BQ, so we assume $f \notin BQ$ until the last subsection.

Observe that $\partial \mathscr{U} \subseteq (\mathscr{P} \cup \mathscr{E})$. Define $\mathscr{W} = \overline{\mathscr{U}} \backslash \mathscr{E}$. Suppose $\lambda: \mathscr{W} \times [0, 1] \to \overline{\mathscr{U}}$ is continuous and has the following properties:

1. $\lambda(\cdot, 0)$ is the identity function on $\mathscr{W}$.
2. $\lambda(\cdot, t)$ is open and one-to-one on $\mathscr{U}$, $t \in [0, 1]$.
3. $\lambda(u, t) = u$, $u \in \partial \mathscr{U}$, $t \in [0, 1]$.

This λ is an isotopy of embeddings of $\mathscr{W}$ in $\overline{\mathscr{U}}$ and we will use it to define a homotopy of (p, q)-maps into $\mathscr{X}$. (See Keldyš [7, p. 3] for more on the subject of isotopies.) Recall that each component of $\mathscr{U}$ has q complete lifts to D under f. The remaining components of $f^{-1}(\mathscr{U})$ comprise a set which we will refer to as $\widehat{\mathscr{U}}$. This set $\widehat{\mathscr{U}}$ contains $p - q$ simply connected components over each component of $\mathscr{P} \cap \mathscr{U}$. During our compactification in the previous section, the pieces of these "partial" lifts were gathered together by pasting them to $p - q$ extra copies of the complete components of $\mathscr{U}$. Our homotopy modifies these partial lifts in a coherent way while leaving fixed the complete lifts and the rest of D.

Definition. *Let f, λ, and $\widehat{\mathscr{U}}$ be as described above. A $\mathscr{U}$-homotopy is a continuous map $\Lambda: D \times [0, 1] \to \mathscr{X}$ with the following properties:*

(1) $\Lambda(\cdot, 0) \equiv f(\cdot)$.
(2) $\Lambda(z, \cdot) \equiv f(z)$, $z \in D \backslash \widehat{\mathscr{U}}$.
(3) $\Lambda(z, \cdot) \equiv \lambda(f(z), \cdot)$, $z \in \widehat{\mathscr{U}}$.

Define $h(z) = \Lambda(z, 1)$, $z \in D$. We say h is $\mathscr{U}$-homotopic to f.

Because $\lambda(\cdot, t)$ is an isotopy, it is not difficult to verify that $\Lambda(\cdot, t)$ is a (p, q)-map for each $t \in [0, 1]$. For the same reason, the only branch points of $\Lambda(\cdot, t)$ are those of f; all these occur in $D \setminus \widehat{\mathcal{U}}$ because $\mathcal{U}$ is the standard fattened q-set, so the branch structure of $\Lambda(\cdot, t)$ is constant in t. Furthermore, Λ does not change the valence of any points outside of $\mathcal{U}$; in particular, the number and the deficiencies of the exceptional points are not changed (indeed, the exceptional set itself is unchanged). We conclude that $\mathcal{U}$-homotopies are (p, q)-homotopies.

Our objective in the remainder of this section is to develop the $\mathcal{U}$-homotopy necessary to replace a general, perhaps pathological, function f by a more manageable and regular function h. There is considerable work involved, since the q-set of f can be quite horrendous. Before beginning that, however, it may be helpful to the geometric intuition of the reader to formulate this $\mathcal{U}$-homotopy in terms of the embedded image surface of f. Recall that the image surface of a map $g: D \to \mathcal{X}$ being embedded in the p-fold covering surface $(\widetilde{\mathcal{X}}, \pi)$ of $\mathcal{X}$ means that $g \equiv \pi \circ \tilde{g}$ for some homeomorphism $\tilde{g}$ of D onto an open subset of $\widetilde{\mathcal{X}}$. Using the homotopy λ discussed earlier, one can establish the following result.

Proposition 4. *Let $f: D \to \mathcal{X}$ be a (p, q)-map with standard fattened q-set $\mathcal{U}$, $\Lambda: D \times [0, 1] \to \mathcal{X}$ a $\mathcal{U}$-homotopy. If the image surface of f is embedded as a disc Δ in $\widetilde{\mathcal{X}}$ and if all branch points of π over points of $\mathcal{U}$ lie in Δ, then there is an isotopy $\Gamma: D \times [0, 1] \to \widetilde{\mathcal{X}}$ for which $\Lambda \equiv \pi \circ \Gamma$.*

Thus the homotopy Λ actually results from a continuous distortion of the embedded disc $\tilde{f}(D)$. For instance, if h is $\mathcal{U}$-homotopic to f, then the image surface of h may also be embedded in $(\widetilde{\mathcal{X}}, \pi)$. However, a note of caution is in order: the isotopy Γ does not necessarily result from composition of $\tilde{f}$ with an ambient isotopy of $\widetilde{\mathcal{X}}$, since, e.g., it may be that the image surface of f is dense while that of h is not.

Preliminary lemma.

Lemma 8. *Let Ω be a nonempty open simply-connected subset of $\mathcal{X}$. Assume that*

(i) $\partial\Omega \subseteq \mathcal{C} \cup \mathcal{E}$,
(ii) $\mathcal{X} \setminus (\overline{\Omega} \cup \mathcal{E}) \neq \varnothing$,

and that there is a component U of $\mathcal{U}$ so that

(iii) $U \cap \Omega$ *is connected, meets* $\mathcal{P}$, *and*
(iv) $\partial\Omega \subseteq \overline{U}$.

Then $\Omega \cap (\mathcal{E} \cup \mathcal{B})$ is nonempty.

Proof. Suppose that $\Omega \cap (\mathcal{E} \cup \mathcal{B}) = \varnothing$. Fix $\omega \in (\Omega \cap U \cap \mathcal{P})$ and let $\{a_1, a_2, \ldots, a_p\}$ be the (necessarily distinct) points of $f^{-1}(\omega)$. Assume that $\{a_1, a_2, \ldots, a_q\}$ are in the q complete lifts of U and consider a_p. Let x_0 be

a point of $\mathscr{X} \backslash (\overline{\Omega} \cup \mathscr{E})$; let γ be a path in D from a_p to any point $b \in f^{-1}\{x_0\}$ which misses $f^{-1}(\mathscr{E} \cup \mathscr{B})$; and let τ be the image $f \circ \gamma$ in $\mathscr{X}$. The path τ starts at ω and ends in the complement of $\overline{\Omega}$. In particular, by (i) and since τ misses $\mathscr{E}$ and $\mathscr{B}$, there is a first point $x_1 \in \mathscr{Q} \backslash \mathscr{B}$ which τ hits. Let μ denote the part of τ starting at ω and ending at x_1. We will find a path ν which lies in $(\Omega \cap U) \cup \{x_1\}$, going from ω to x_1; moreover, we show a fixed-endpoint (f.e.p.) homotopy in $\Omega \cup \{x_1\}$ from μ to ν which lifts under f^{-1} to a homotopy starting at a_p. Since μ lifts to part of the path γ in D, we see that ν will also lift to a path from a_p to some point of $f^{-1}\{x_1\}$. However, since ν lies in U and $x_1 \in \overline{U}$ by (iv), we know that ν has the q lifts starting at $\{a_1, a_2, \ldots, a_q\}$. Having avoided branch points, this implies there are at least $q + 1$ points in $f^{-1}\{x_1\}$, contradicting $x_1 \in \mathscr{Q}$. The rest of the proof, then, lies in finding ν and proving the existence of the lifting homotopy.

We need to study the continuation of f^{-1} along paths in Ω starting at ω with the "branch" of f^{-1} specified by $f^{-1}\{\omega\} = a_p$. We say that a path from ω *lifts* if this branch has a continuation along the path. We show that the points of Ω which can be reached by paths which lift comprise a simply-connected set $\mathscr{S}$. Note that Ω contains no branch values, thus continuation always exists in $\mathscr{P} \cap \Omega$. Furthermore, by (i), any component of $\mathscr{P}$ intersecting Ω lies entirely in Ω. Let P_1 be the component containing ω. By (iv), any component of $\mathscr{U}$ other than U which intersects Ω lies entirely in Ω. As for U itself, $U \cap \Omega$ is simply connected since U is simply connected relative to $\mathscr{X} \backslash \mathscr{E}$ while Ω is simply connected and contains no points of $\mathscr{E}$.

Consider the collection Φ of all sets S satisfying:

(a) $S \subseteq \Omega$ is open and simply connected,
(b) $\omega \in P_1 \subseteq S$,
(c) f^{-1} lifts on all paths from ω which lie in S,
(d) S is a union only of sets of these three types:
   $\{1\}$ components of $\mathscr{P}$,
   $\{2\}$ components of $\mathscr{U}$, and/or
   $\{3\}$ $U \cap \Omega$.

Partial order Φ by set inclusion. Every totally ordered subcollection has a least upper bound; namely, the union. We define $\mathscr{S}_0$ to be a maximal element in Φ. $\mathscr{S}_0$ is simply connected; to show it meets our description of $\mathscr{S}$ above, it suffices to show that any path λ starting at ω which lies in Ω will fail to lift if it contains a point of $\partial \mathscr{S}_0$. Suppose the first such point which λ encounters is $t \in \partial \mathscr{S}_0 \cap \Omega$.

Case 1, $t \in P \subseteq \mathscr{P}$: We show that $P \cup \mathscr{S}_0$ is simply connected. Otherwise, there is some $x \in \partial P \cap (\Omega \backslash \mathscr{S}_0)$ which is separated from $\partial \Omega$, and hence from $\mathscr{E}$, by $P \cup \mathscr{S}_0$. But $x \in \partial P$ implies $x \in \mathscr{Q}$, so x lies in a component Q_1 of $\mathscr{Q}$ which necessarily has a limit point in $\mathscr{E}$. We see then that $Q_1 \cap (P \cup \mathscr{S}_0) \neq \varnothing$, hence $Q_1 \cap \mathscr{S}_0 \neq \varnothing$. By (d) of the definition of sets in Φ, the component of

$\mathscr{U}$ containing Q_1 (or $U \cap \Omega$) must lie in $\mathscr{S}_0$. This would give $x \in \mathscr{S}_0$ and contradict $x \in \Omega \setminus \mathscr{S}_0$. So $P \cup \mathscr{S}_0$ is simply connected. If f^{-1} continues along λ to $t \in P$, then it continues on paths to all points of P. Therefore $P \cup \mathscr{S}_0$ meets the conditions for sets in Φ, contradicting the maximality of $\mathscr{S}_0$.

Case 2, $t \in Q \subseteq \mathscr{Q}$: Let U' denote the component of $\mathscr{U}$ containing Q (or $U \cap \Omega$). The argument above shows $U' \cup \mathscr{S}_0$ is simply connected. If f^{-1} continues along λ to $t \in Q$, then the value of $f^{-1}(t)$ lies in a complete lift of U'. In particular, f^{-1} can be continued on paths to all points of U'. Therefore $U' \cup \mathscr{S}_0$ meets the conditions for sets in Φ, again contradicting the maximality of $\mathscr{S}_0$.

We now conclude that $\mathscr{S}_0$ is the set $\mathscr{S}$ we were looking for: namely, a simply connected set consisting of all the points and only the points of Ω which can be reached by paths in Ω starting at ω along which f^{-1} continues. We have continuation along μ which runs from ω to x_1. Now $x_1 \in \mathscr{Q}$ and by (iv) $x_1 \in \overline{U}$; but $\partial U \subseteq \mathscr{P} \cup \mathscr{E}$, so $x_1 \in U$. Thus, the last bit of μ lies in U. By (iii), $U \cap \Omega$ is connected, so we can find a curve ν from ω to x_1 which lies in $U \cap \Omega$ and agrees with μ on this last bit. From ω up to the point where they rejoin, μ and ν lie in $\mathscr{S}$. As $\mathscr{S}$ is simply connected, these parts of μ and ν are f.e.p.-homotopic in $\mathscr{S}$ and this homotopy lifts under f^{-1}. This is what was needed to complete the proof of the lemma. □

A standard $\mathscr{U}$. We want to study in considerable depth the structure of that part of $\mathscr{Q}$ lying in a single component of $\mathscr{U}$. Our first task is to verify that a well-chosen $\mathscr{U}$ will have components with even more regularity than we have so far used. In what follows, the fattened q-set $\mathscr{U}$ will be the one obtained by the method used in the proof of Proposition 1 and modified as in the proof of Proposition 3. Briefly recall that we used the depth function r_q to define a fattened q-set $\mathscr{V}$; that we adjoined to each component of $\mathscr{V}$ those components of its complement which did not contain points of $\mathscr{E}$; and finally that we excised at most a finite number of simple curves to remove any branch values in $\mathscr{P}$. In this last step, the curves we chose were not uniquely determined; nonetheless, we will refer to the resulting fattened q-set $\mathscr{U}$ as *the standard* fattened q-set. We make use of the conformal structure that components of $\mathscr{U}$ inherit from the Riemann surface $\mathscr{X}$; however, we want to emphasize that this structure is not the essential ingredient here, but simply a convenient way to obtain topological maps with desirable properties.

Fix a component U of $\mathscr{U}$. This is open and simply connected relative to $\mathscr{X} \setminus \mathscr{E}$. Since $\mathscr{E}$ is finite, U is a finitely connected and planar Riemann surface, and it is a classical result that U is conformally equivalent to a "circle domain" $R = D \setminus (\bigcup_{j=1}^{N} D_j)$ where the D_j are closed discs or points lying in D. If $\mathscr{X}$ has positive genus, U is simply connected so $R = D$.

Lemma 9. *If U is a component of the standard fattened q-set $\mathscr{U}$ and ψ is a one-to-one conformal map of R onto U, then ψ extends continuously to ∂R.*

Proof. It is well known that this conclusion is independent of the conformal mapping ψ. Also, it suffices to prove the result for the set W defined below, since U is obtained from W by removal of a finite number of simple arcs.

Define V by

$$V = \bigcup_{x \in \mathscr{C} \cap U} \Delta\left(x, \frac{r_q(x)}{3}\right),$$

and let W be obtained from V by adjoining those components of $\mathscr{X} \backslash V$ which are disjoint from $\mathscr{E}$. Note that $\mathscr{C} \cap W = \mathscr{C} \cap U$.

Assume ψ is a one-to-one conformal map of R onto W. To prove ψ extends continuously to ∂R, it suffices to prove that ψ has singleton cluster sets at all points of ∂R. Therefore, fix $\xi \in \partial R$ and suppose γ is an arc in R ending at ξ with a cluster set containing two, hence infinitely many points. Choose one such cluster value $x_0 \notin \mathscr{E}$. To reach a contradiction, it suffices to find a curve σ in R, also ending at ξ, along which ψ has asymptotic value x_0. It is not difficult to see that there are points $\{a_j\}_{j=1}^{\infty} \subseteq \gamma$ so that the image points $x_j = \psi(a_j)$ lie in V and converge to x_0. For each j there is a point $y_j \in \mathscr{C} \cap U$ with $x_j \in \Delta(y_j, r_q(y_j)/3)$. Taking a subsequence, if necessary, we have $y_j \to y_0 \in \mathscr{C} \cup \mathscr{E}$. It is evident that $y_0 \notin \mathscr{E}$ since $x_j \to x_0 \notin \mathscr{E}$ and $r_q(y_j) \to r_q(y_0)$. Thus $y_0 \in \mathscr{C}$ and $r_q(y_0) > 0$. In particular, $x_0 \in \partial\Delta(y_0, r_q(y_0)/3)$; thus there is a geodesic τ in $V \subseteq U$ running from y_0 to x_0. Since the discs $\Delta(y_j, r_q(y_j)/3)$ lie in U and $y_j \to y_0$, it is easy to check that τ can be modified with small diversions in the neighborhood of x_0 so that it contains some subsequence of the x_j's and still converges to x_0. The preimage $\psi^{-1}(\tau)$ is then a path σ in R ending at ∂R which contains a subsequence of $\{a_j\}$ and hence ends at ξ. This completes the contradiction and hence the proof. $\square$

We note that the extension of ψ, which we will continue to call ψ, is not necessarily one-to-one on ∂R. However, $\psi(\partial R) \subseteq \partial \mathscr{U}$, and that will be enough for our purposes. In R, we will use the following notations: $\widehat{\mathscr{E}} = \psi^{-1}(\mathscr{E})$, $\widehat{\mathscr{C}} = \psi^{-1}(\mathscr{C})$, and $\widehat{Q}$ will denote a component of $\widehat{\mathscr{C}}$, hence a preimage of a component Q of $\mathscr{C}$.

Remark. We list here some of the properties of these sets which will be useful shortly:

(a) Each $\widehat{Q}$ is relatively closed in R and simply connected.

(b) Each $\widehat{Q}$ has one or more limit points in $\widehat{\mathscr{E}}$.

(c) $\widehat{\mathscr{E}}$ is a closed nowhere dense subset of ∂R. This is due to the fact that $\mathscr{E}$ is finite and that the boundary values of a conformal mapping cannot lie in a finite set on a set of positive logarithmic capacity [2].

(d) Every component Ω of $R \backslash \widehat{\mathscr{C}}$ contains an arc of $\partial R \backslash \widehat{\mathscr{E}}$ in its boundary. For suppose not, then $\psi(\Omega)$ will be a component of $\mathscr{P} = \mathscr{X} \backslash (\mathscr{C} \cup \mathscr{E})$ which

lies entirely in U. By Lemma 5, this would contradict the assumption that $f \notin BQ$.

(e) Every component of ∂R contains a limit point of some component $\widehat{Q}$.

Proposition 5. $R \setminus \widehat{\mathscr{Q}}$ *has only finitely many components.*

Proof. Assuming there are infinitely many components, (b) and (d) above allow us to choose, in each component, one simple arc α_j which starts and ends at points of $\widehat{\mathscr{E}}$ (not necessarily distinct). The images of these curves in $\mathscr{X}$ under the map ψ will comprise infinitely many disjoint simple arcs γ_j in $\mathscr{X}$, each starting and ending at $\mathscr{E}$. In particular, infinitely many have the same pair $e_1, e_2 \in \mathscr{E}$ of endpoints. (Here again, the points e_1 and e_2 may in fact be the same.) $\mathscr{X}$ has finite genus, so one easily verifies that there will be infinitely many configurations of the type illustrated in Figure 5; that is, involving a contractable region Σ bounded by curves γ_1 and γ_5 and containing additional curves γ_2, γ_3, and γ_4. The curves $\gamma_1, \gamma_2, \gamma_3, \gamma_4$, and γ_5 lie in U; e_1 and e_2 are not in U; and U is connected. This ensures that there is a curve from γ_3 to either γ_1 or γ_5 which lies in U; say such a curve runs from γ_3 to γ_1. Since γ_1, γ_2 and γ_3 are in distinct components of $U \setminus \mathscr{Q}$, there are components Q_1 and Q_2 of $\mathscr{Q}$ which lie between γ_1 and γ_2 and between γ_2 and γ_3, respectively. Furthermore, Q_1 separates γ_1 and γ_2 in Σ, while Q_2 separates γ_2 and γ_3 in Σ, so Q_1 and Q_2 each have both e_1 and e_2 as limit points. Define Ω to be the component of $\mathscr{X} \setminus (\overline{Q}_1 \cup \overline{Q}_2)$ which contains γ_2 (shaded). Note the following: Ω is simply connected; $\partial \Omega \subseteq (\overline{Q}_1 \cup \overline{Q}_2)$, which lies in $(\mathscr{Q} \cup \mathscr{E}) \cap \overline{U}$; $\mathscr{X} \setminus (\overline{\Omega} \cup \mathscr{E})$ is nonempty, containing, e.g., points of γ_5; and $\Omega \cap U$ is connected, due to the existence, noted above, of a curve in U joining γ_3 to γ_1. We conclude from Lemma 8 that Ω contains either a branch value or an exceptional value. However, there are infinitely many disjoint sets like Ω and only finitely many branch and exceptional values. This contradiction proves the result. □

It is convenient to continue our work in R rather than U. We now know that $R \setminus \widehat{\mathscr{Q}}$ has at most finitely many
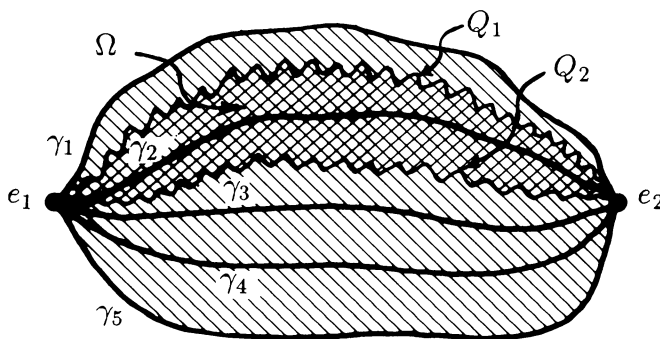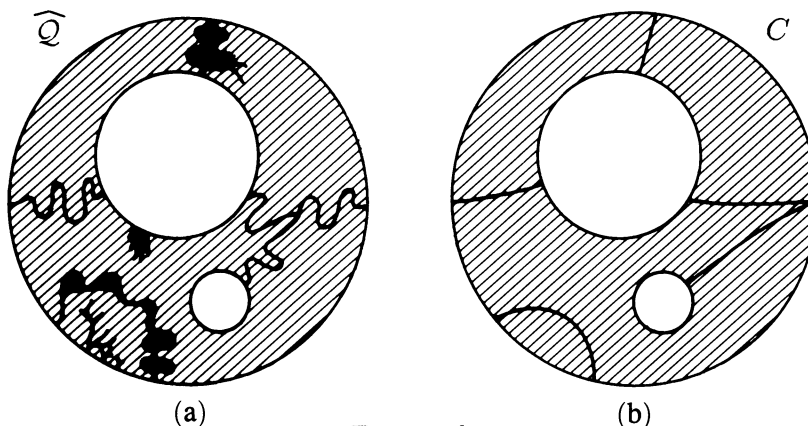

FIGURE 5

FIGURE 6

components, to be denoted $\Omega_1, \Omega_2, \ldots, \Omega_n$, and that $\widehat{\mathscr{E}} \subseteq \partial R$ is closed and nowhere dense. This implies that the components of $\widehat{\mathscr{E}}$ which have more than one accumulation point on ∂R comprise a finite set, which we denote as $\widehat{Q}_1, \widehat{Q}_2, \ldots, \widehat{Q}_m$, and that their accumulation points on ∂R will be finite in number. We will construct a finite collection of pairwise disjoint smooth simple curves α in R, with endpoints among these same points on ∂R, whose union $\mathscr{E}$ will replace $\widehat{\mathscr{E}}$. More precisely, we develope an isotopy which carries $R \backslash \widehat{\mathscr{E}}$ to $R \backslash \mathscr{E}$. Figure 6 illustrates a domain R with the complicated set $\widehat{\mathscr{E}}$ in (a) and its replacement $\mathscr{E}$ in (b).

Definition. *Let A and B be open subsets of a plane open set W, F a subset of $\partial A \cap \partial B \cap \partial W$. We say that A is isotopic to B in W with F fixed if there is a continuous map $\Gamma : A \cup F \times [0, 1] \to \overline{W}$ for which*:

   (a) $\Gamma(\cdot, 0) = identity$,
   (b) $\Gamma(\cdot, t)$ *maps A one-to-one into W, $t \in [0, 1]$,*
   (c) $\Gamma(\cdot, 1)$ *maps A one-to-one onto B, and*
   (d) $\Gamma(x, t) = x$, $x \in F$, $t \in [0, 1]$.

Note that conditions (a)–(c) are the usual ones associated with an isotopy (see [7]), and that (d) is the added ingredient.

Lemma 10. *There exists a set $\mathscr{E}$, a union of finitely many arcs as described above, so that $R \backslash \widehat{\mathscr{E}}$ is isotopic to $R \backslash \mathscr{E}$ in R with $\partial R \backslash \widehat{\mathscr{E}}$ fixed.*

The isotopy Γ will be obtained as a succession of three others, Γ_1, Γ_2, and Γ_3. The first serves to eliminate all components of $\widehat{\mathscr{E}}$ except $\widehat{Q}_1, \widehat{Q}_2, \ldots, \widehat{Q}_m$, allowing us to define the arcs in $\mathscr{E}$. The second shrinks the components of $R \backslash (\bigcup \widehat{Q}_j)$ so that they also lie in $R \backslash \mathscr{E}$. This ensures that we do not violate the global one-to-one condition during the third stage, when we expand these shrunken components, but now to fill out $R \backslash \mathscr{E}$. The details are rather cumbersome. However, in view of the difficulties which topological pathologies have caused in past studies, we feel that care is warranted. First is the central existence result.

Lemma 11. *Let A be a simply-connected open subset of the unit disc D, F a union of open arcs in $\partial A \setminus \partial(D \setminus A) \subseteq \partial D$. Then A is isotopic to D in D with F fixed.*

Proof. Let σ denote the desired homotopy. If $F = \varnothing$, it is well known that there is an isotopy of A onto D. Therefore, assume $F \neq \varnothing$. We obtain σ from countably many ambient isotopies of $\overline{D}$ of the following type: Let M and N be simply-connected compact subsets of $\overline{D}$, M in the (relative) interior of N, and $k > 0$. Then there is a continuous map $\tau : \overline{D} \times [0, 1] \to \overline{D}$ so that

  (i) $\tau(\cdot, 0) = $ identity,
  (ii) $\tau(\cdot, t)$ fixes $M \cup \{1 - 2^{-k-1} \leq |z| \leq 1\}$, $t \in [0, 1]$,
  (iii) $\tau(\cdot, t)$ is a homeomorphism of $\overline{D}$, $t \in [0, 1]$, and
  (iv) $\tau(\cdot, 1)$ maps $D \setminus N$ into $\{1 - 2^{-k} < |z| < 1 - 2^{-k-1}\}$.

Choose a sequence $t_n \to 1$, $0 = t_0 < t_1 < \cdots < 1$. We define σ in stages, $\sigma_n \equiv \sigma|_{[t_n, t_{n+1}]}$. Exhaust $A \cup F$ with a sequence $\{K_n\}_{n=0}^{\infty}$ of compact subsets, each in the relative interior of the next. Begin with $M = K_0$, $N = K_1$, $k = 1$, and obtain τ as above. Define

$$\sigma_0(\cdot, t) = \tau\left(\cdot, \frac{t}{(t_1 - t_0)}\right), \qquad t \in [t_0, t_1].$$

Assume as the induction hypothesis that σ_{n-1} has been defined on $[t_{n-1}, t_n]$. Let $M = \sigma_{n-1}(K_n, t_n)$, $N = \sigma_{n-1}(K_{n+1}, t_n)$, $k = n + 1$, and obtain τ as above. Define

$$\sigma_n(\cdot, t) = \tau\left(\sigma_{n-1}(\cdot, t_n), \frac{(t - t_n)}{(t_{n+1} - t_n)}\right), \qquad t \in [t_n, t_{n+1}].$$

We arrive inductively at a sequence $\{\sigma_n\}_{n=0}^{\infty}$. Now we define σ as follows:

$$\sigma(z, t) = \begin{cases} \sigma_n(z, t), & z \in A \cup F, \ t \in [t_n, t_{n+1}], \\ \lim_{n \to \infty} \sigma_n(z, t_n), & z \in A \cup F, \ t = 1. \end{cases}$$

Verification of most properties of σ is fairly routine, but we point out these facts: Every compact $K \subseteq A \cup F$ lies in K_n for some n, hence its image $\sigma_{n-1}(K, t_n)$ is fixed by σ_m for $m \geq n$. This is the essential ingredient in proving continuity and injectivity of σ at the endpoint $t = 1$. It also shows, along with the fact that $\sigma(\cdot, t)$ fixes ∂D for all $t \in [0, 1)$, that $\sigma(\cdot, t)$ fixes F, $t \in [0, 1]$. Since $\{|z| < 1 - 2^{-n-1}\} \subseteq \sigma_n(A, t) \subseteq D$, $t \in [t_n, t_{n+1}]$, one verifies that $\sigma(\cdot, 1)$ maps A onto D. This completes the proof of Lemma 11. $\square$

Of course, this result can be transferred from D to another simply connected domain G by use of a homeomorphism, although care must be taken that the homeomorphism extends continuously to the set $F \subseteq \partial G$ of interest. In our case G will always be a simply-connected subset of the circle domain R and F will consist of *free arcs* in ∂G; that is, arcs in $\partial R \cap \partial G$ which avoid $\overline{R \setminus G}$. Consequently, homeomorphisms $\rho : G \to D$ which extend to be continuous and

one-to-one on F are available (e.g., by using conformal mappings). The isotopy σ of Lemma 11 can be lifted to G as $\rho^{-1} \circ \sigma \circ \rho$. With this in mind, we will simply apply Lemma 11 directly to such regions G.

Proof of Lemma 10. Designate a point z_j in each component Ω_j of $R \backslash \widehat{\mathscr{C}}$, for $j = 1, 2, \dots, n$. Our first isotopy, Γ_1, will eliminate the components of $\widehat{\mathscr{C}}$ other than $\widehat{Q}_1, \widehat{Q}_2, \dots, \widehat{Q}_m$.

For each $j = 1, 2, \dots, n$, let S_j be the unbounded component of $C \backslash \overline{\Omega_j}$ and define Σ_j to be that component of $R \backslash (S_j \cup \widehat{Q}_1 \cup \cdots \cup \widehat{Q}_m)$ containing z_j. One can verify that Σ_j is a simply-connected open set; that $\Omega_j \subseteq \Sigma_j$; that free arcs of $\partial \Omega_j$ are also free arcs of $\partial \Sigma_j$; and that $\Sigma_j \cap \Omega_k = \varnothing$, $k \neq j$. (For this last, see our earlier Remark (d).) Let F_j be the union of the free arcs of $\partial \Omega_j$, $F = \bigcup F_j$. By Lemma 11, Ω_j is isotopic to Σ_j in Σ_j with F_j fixed. Since the Σ_j, $j = 1, 2, \dots, n$, are pairwise disjoint, these isotopies can proceed simultaneously, showing that $R \backslash \widehat{\mathscr{C}} = \bigcup \Omega_j$ is isotopic to $\bigcup \Sigma_j$ in R with $F = \partial R \backslash \widehat{\mathscr{E}}$ fixed. The map implementing this isotopy is our Γ_1. Note that if $\widehat{Q}$ is a component of $\widehat{\mathscr{C}}$ which has a single accumulation point on ∂R, then $\widehat{Q}$ lies in one of the sets Σ_j. Therefore, $R \backslash \bigcup \Sigma_j = \bigcup_{k=1}^m \widehat{Q}_k$, so Γ_1 removes from consideration a possibly uncountable number of bothersome components; this will prove important in the next stage. Also note that if we eliminate all of $\widehat{\mathscr{C}}$, i.e., if $m = 0$, then take $\Gamma = \Gamma_1$ and we are finished. So assume $m \geq 1$.

We next need to construct the arcs forming $\mathscr{E}$. Consider one of the components Σ_j and let $\phi : D \to \Sigma_j$ be a conformal map with $\phi(0)$ as the distinguished point z_j. The boundary of D decomposes into a finite number of closed connected arcs I having disjoint interiors, each maximal with the property that the cluster set of ϕ on I lies in the closure of only one of $\widehat{Q}_1, \widehat{Q}_2, \dots, \widehat{Q}_m$, or in ∂R. This occurs because these latter sets are continua which intersect only at isolated points of $\widehat{\mathscr{E}}$. Standard arguments in conformal mapping imply that ϕ is continuous at the endpoints of each I and map them to $\widehat{\mathscr{E}}$. Let $I_1, \dots, I_l$ be those intervals on which ϕ has cluster set in the closure of one of the $\widehat{Q}_i$ rather than in ∂R. Construct a system of pairwise disjoint simple smooth arcs $\gamma_1, \dots, \gamma_l$ in D so that γ_i connects the endpoints of I_i and separates I_i both from the origin and from $\partial D \backslash I_i$. The curves $\alpha_i = \phi \circ \gamma_i$, $i = 1, \dots, l$, in Σ_j will start and end at points of $\widehat{\mathscr{E}}$, will be pairwise disjoint, will fail to separate z_j from the free arcs of $\partial \Sigma_j$, yet will separate z_j from $\partial \Sigma_j \cap R$.

Let the finite collection of these arcs for all $j = 1, \dots, n$ be denoted by $\mathscr{A}$. Any curve in R joining distinct points among the z_j or any simple closed non-null homotopic curve will encounter $\bigcup (\partial \Sigma_j \cap R)$. Hence it will also encounter some curve in $\mathscr{A}$. Therefore, the full collection $\mathscr{A}$ will

    (a) separate the points $z_1, \dots, z_n$, and
    (b) break R into simply connected components.

However, we do not want all the curves of $\mathscr{A}$; choose a minimal subcollection $\mathscr{A}_1$ which has properties (a) and (b). The union of these curves is denoted by $\mathscr{C}$. From (a) and (b), components Θ of $R\backslash\mathscr{C}$ are simply connected, and each z_j is in a unique one which we will denote by Θ_j, $j = 1, \ldots, n$. We claim that, in fact, $\Theta_1, \ldots, \Theta_n$ are the only components. If not, there is a component Θ containing none of the points z_j but sharing a boundary arc α with, say, Θ_1. Since α has its endpoints in ∂R, $\Theta \cup \alpha \cup \Theta_1$ is open, simply connected, and disjoint from $\Theta_2, \ldots, \Theta_n$. Thus α could be removed from the collection $\mathscr{A}_1$, contradicting its minimality. We conclude that $R\backslash\mathscr{C} = \bigcup_{j=1}^n \Theta_j$.

Now to describe Γ_1 and Γ_2: Define $A_j = \Sigma_j \cap \Theta_j$, $j = 1, \ldots, n$. For each j, $z_j \in A_j$ and, by our construction of the arcs α, the free arcs of $\partial\Sigma_j$ are also free arcs of ∂A_j. Let E_j be the union of these free arcs, and $E = \bigcup E_j$. Applying Lemma 11, only reversing the isotopy, we see that Σ_j is isotopic to A_j in Σ_j with E_j fixed. As the Σ_j are pairwise disjoint, these isotopies can proceed simultaneously for $j = 1, 2, \ldots, n$, showing that $\bigcup\Sigma_j$ is isotopic to $\bigcup A_j$ in R with E fixed. The map implementing this isotopy is Γ_2.

The free arcs in E_j are also free arcs of $\partial\Theta_j$, $j = 1, \ldots, n$. Applying Lemma 11, A_j is isotopic in Θ_j to Θ_j with E_j fixed. Again, the Θ_j are pairwise disjoint, these isotopies can proceed simultaneously, and we see that $\bigcup A_j$ is isotopic to $\bigcup\Theta_j = R\backslash\mathscr{C}$ in R with E fixed. The map implementing this is Γ_3.

Note that $R\backslash\widehat{\mathscr{C}} = F \subseteq E$, so that Γ_1, Γ_2, and Γ_3 all leave $R\backslash\widehat{\mathscr{C}}$ fixed. Applying them in succession, we obtain the desired isotopy Γ. As a final note, observe that we can shift the curves in $\mathscr{C}$ slightly, if necessary, so that they avoid the points of R associated with branch values in U, i.e., $\mathscr{B}\cap U$. This completes our proof of Lemma 10. $\square$

Now we can return to U. Define the homotopy λ on U by

$$\lambda(x, t) = \psi(\sigma(\psi^{-1}(x), t)), \qquad x \in \overline{U}, \ t \in [0, 1].$$

Of course, ψ^{-1} is not well defined as a point function on ∂U, but for $x \in \partial U$, the set $\psi(x)$ lies in ∂R, and σ is the identity there; we conclude that λ is well defined. We consider the components of $\mathscr{U}$ individually to arrive at the homotopy λ defined on $\mathscr{U}$. We use this to define the $\mathscr{U}$-homotopy Λ described earlier between $f(\cdot) = \Lambda(\cdot, 0)$ and $h(\cdot) = \Lambda(\cdot, 1)$. The set $\mathscr{U}$ is still a fattened q-set for h, although, having moved $\mathscr{C}$ in such a way that branch values formerly in $\mathscr{C}\cap\mathscr{U}$ are now in $\mathscr{P}\cap\mathscr{U}$, it may no longer be *regular*. This new $\mathscr{C}$ consists, in each component of $\mathscr{U}$, of a finite number of disjoint simple arcs (the $\psi\circ\alpha_j$) starting and ending at points of $\mathscr{E}$. The last step is to observe that the total number of arcs comprising the new $\mathscr{C}$, irrespective of components of $\mathscr{U}$, is finite.

Proposition 6. *Suppose the set $\mathscr{C}$ is a collection of disjoint simple arcs starting and ending at points of $\mathscr{E}$. Then the number of such arcs is finite.*

Proof. Let $\{\Delta_e: e \in \mathscr{E}\}$ denote a collection of topological discs with disjoint closures about the points of $\mathscr{E}$ chosen so that all points of $\mathscr{B} \setminus \mathscr{E}$ lie in $K = \mathscr{X} \setminus \bigcup_{\mathscr{E}} \Delta_e$. Note that K is compact and disjoint from $\mathscr{E}$, so the depth function r_q is bounded below on K, say $r_q(y) > \delta > 0$, $y \in K$. For $e \in \mathscr{E}$, let γ be a simple closed curve in K which separates e from the other points of $\mathscr{E}$ and suppose it has finite length L in the metric ρ. If U is a component of $\mathscr{U}$ containing a curve from e to some other point of $\mathscr{E}$, then $\gamma \cap U$ contains a point of $\mathscr{Q}$; hence by the construction of $\mathscr{U}$, $\gamma \cap (\text{int}(\overline{U}))$ has length at least $\frac{2\delta}{3}$. Therefore γ can intersect at most a finite number of such components, and each has only finitely many curves in $\mathscr{Q}$. We conclude that $\mathscr{Q}$ contains at most finitely many curves running between distinct points of $\mathscr{E}$.

Now, if $\mathscr{Q}$ contains infinitely many arcs, there is some point $e \in \mathscr{E}$ so that infinitely many start and end at e. Indeed, a slight modification of the argument above shows that infinitely many such arcs will lie in Δ_e. If α is such an arc, then $\alpha \cup \{e\}$ bounds a topological disc $\Omega \subseteq \Delta_e$. By our choice of Δ_e it contains no branch or exceptional values, and one checks that it meets the hypotheses of Lemma 8. This contradiction proves Proposition 6. □

   We have established the following:

Theorem 2. *Let $f: D \to \mathscr{X}$ be a (p, q)-map, $f \notin BQ$. Then there is a (p, q)-map h which is (p, q)-homotopic to f whose q-set consists of a finite number of simple arcs starting and ending at exceptional points. Aside from the endpoints, the arcs are pairwise disjoint and do not contain branch values.*

BQ-maps. The parallel result for maps in BQ is considerably easier. Using the notation from the proof of Lemma 5, $\mathscr{P}$ and $\mathscr{U}$ consist of single simply connected components P and U which overlap in a topological annulus A, and $\mathscr{Q}$ is a simply connected compact subset Q of $\mathscr{U}$. In the compactification argument, we pasted to the image surface of f a $(p - q)$-fold covering surface U_0 of U having a branch point of order $p - q - 1$ over $z_0 \in \mathscr{Q}$. Note that the surface $\widetilde{\mathscr{X}}$ which we obtain in this case is the Riemann sphere C^∞; indeed, $\beta_f = p + q - 1$, and we have added $p - q - 1$ additional branch points, so $\beta_\pi = 2p - 2$, and by the Riemann-Hurwitz relations, $g(\widetilde{\mathscr{X}}) = 0$. Let Δ be a metric disc (the spherical metric) centered at z_0 and of sufficiently small radius that its closure lies in U. The set $C = U \setminus Q$ is a topological annulus, and it is evident that there is an isotopy $\lambda: C \times [0, 1] \to U$ of the type indicated at the beginning of this section so that $\lambda(\cdot, 1)$ maps C onto $U \setminus \overline{\Delta}$. The associated $\mathscr{U}$-homotopy Λ proves the following:

Theorem 3. *Let $f: D \to C^\infty$ be a (p, q)-map in BQ. Then there is a (p, q)-map $h \in BQ$ which is (p, q)-homotopic to f and a metric disc $\Delta \subseteq C^\infty$ so that h has p-set $C^\infty \setminus \overline{\Delta}$ and q-set $\overline{\Delta}$.*

Summary. We have shown that arbitrary (p, q)-maps are (p, q)-homotopic to ones having very regular q-sets. Since such homotopies preserve valence

structure, we are justified in the sequel in assuming that the particular map f we are studying has a q-set with the regularity properties of Theorem 2 or 3, as appropriate.

5. SIMPLICIAL MAPS

We will start with a (p, q)-map f which is not in BQ. In accordance with the results of the last section, we assume its q-set consists of a finite number of simple arcs which start and end at exceptional points and are, aside from the endpoints, mutually disjoint and free of branch values. The image surface of f will be embedded in the compact covering $(\widetilde{\mathscr{X}}, \pi)$ of $\mathscr{X}$ in accordance with Theorem 1.

Choose a finite triangulation T of $\mathscr{X}$ with the following properties:

(1) Each point of $\mathscr{E}$ is a 0-simplex (vertex) of T,
(2) Each point of $\mathscr{B}$ is a 0-simplex (vertex) of T,
(3) each arc in $\mathscr{Q}$ forms, along with its endpoints in $\mathscr{E}$, a 1-simplex (edge) of T.

Now, lift T via the projection π to a triangulation $\widetilde{T}$ of $\widetilde{\mathscr{X}}$. This is well defined because of (2) above. The image surface of f is somehow embedded as a disc $\widetilde{D}$ in this complex. However, to describe $\widetilde{D}$, it seems best to go one additional step: The universal covering surface of $\widetilde{\mathscr{X}}$ is the sphere, the plane, or the disc, depending on the genus of $\widetilde{\mathscr{X}}$. If we use the notation $(\widehat{\mathscr{X}}, \kappa)$ for this covering surface, then we can lift $\widetilde{T}$ via κ to a triangulation $\widehat{T}$ of $\widehat{\mathscr{X}}$. As it is simply connected, $\widetilde{D}$ lifts in a one-to-one fashion to a disc $\widehat{D}$ in $\widehat{\mathscr{X}}$. It is this latter disc which we can describe most easily: The points of $\mathscr{Q} \cup \mathscr{E}$ lie in the 0- and 1-simplices of T, so the interior of every 2-simplex belongs to $\mathscr{P}$. Each of these 2-simplices is covered by p of the 2-simplices of $\widetilde{\mathscr{X}}$, so the interiors of all 2-simplices of $\widetilde{T}$ lie in $\widetilde{D}$ (along with certain of the 0-simplices and interiors of certain 1-simplices of $\widetilde{T}$). In particular, in $\widehat{T}$, the disc $\widehat{D}$ contains the interior of a 2-simplex over the interior of each 2-simplex of $\widetilde{T}$. We conclude that $\widehat{D}$ is a fundamental domain for the covering $(\widehat{\mathscr{X}}, \kappa)$.

The boundary of $\widehat{D}$ is a 1-complex which we denote $\widehat{K}$ and is projected under κ to the 1-complex $\widetilde{K}$ which forms the boundary of $\widetilde{D}$. This, in turn, projects under π to the 1-complex we will denote by K consisting of $\mathscr{Q} \cup \mathscr{E}$. The covering $(\widehat{K}, \kappa)$ of $\widetilde{K}$ is two-to-one, with each 1-simplex of $\widetilde{K}$ being covered once in one direction and once in the other, in the usual fashion for fundamental domains. The covering $(\widetilde{K}, \pi)$ of K is $(p - q)$-to-one on the open 1-simplices; counting 0-simplices is more difficult, as we will see shortly.

Theorem 4. *Let $f: D \to \mathscr{X}$ be a (p, q)-map. Then f is (p, q)-homotopic to a map h which extends continuously to a simplicial map of a triangulation of $\overline{D}$ and a triangulation of $\mathscr{X}$.*

Proof. Consider the case of $f \notin BQ$. Observe that the valence properties of f are determined entirely by the properties of the embedded disc $\widehat{D} \subseteq \widehat{T}$

described above. Since $\widehat{D}$ is a proper, open, simply connected subset of $\widetilde{\mathscr{X}}$, it is conformally equivalent to the unit disc D; let $\psi\colon D \to \widehat{D}$ be a conformal mapping. As $\partial \widehat{D}$ consists of a finite number of points and a finite number of pairwise disjoint simple arcs, ψ can be extended to be continuous and one-to-one on ∂D. We also refer to the extended function by ψ. Using ψ, we can pullback the triangulation induced on the closure of $\widehat{D}$ by $\widehat{T}$ to a triangulation S of $\overline{D}$. The map $h\colon \overline{D} \to \mathscr{X}$ defined by $h \equiv \pi \circ \kappa \circ \psi$ is a simplicial map of S to T. Moreover, the image surface of the restriction of h to D is precisely that of f. By Lemma 3, there exists an orientable homeomorphism η of D for which $f \circ \eta \equiv h$. This is isotopic to the identity in D, and the isotopy induces, by composition with f, a homotopy between f and h, all consisting of (p, q)-maps with image surface $\widetilde{D}$, hence with identical valence structures. This is the (p, q)-homotopy we desired.

For $f \in BQ$, we proceed differently, since we can reach a stronger conclusion. By Theorem 3, we may assume there is a metric disc Δ in C^{∞} so that the p-set of f is $C^{\infty} \setminus \overline{\Delta}$ and the q-set is $\overline{\Delta}$. Let $\Theta\colon C^{\infty} \times [0, 1] \to C^{\infty}$ be an isotopy of C^{∞} (using Möbius transformations, for instance) for which $\Theta(\cdot, 0)$ is the identity and $\Theta(\cdot, 1)$ maps $C^{\infty} \setminus \overline{\Delta}$ onto the open unit disc. Define $\Lambda(\cdot, t) = \Theta(f(\cdot), t)$, $t \in [0, 1]$. This is evidently a (p, q)-homotopy. The image surface of $g(\cdot) = \Theta(\cdot, 1)$ covers points of the unit disc p times and the points of its complement q-times. One can show that this is also the image surface of some quotient of finite Blaschke products as described in Example 1; in particular, if h is a meromorphic function of D onto this surface, then its cluster set on ∂D lies in ∂D. Dividing out the zeros using a p-fold Blaschke product, the cluster set continues to lie in ∂D. An application of the maximum principle shows that we are left with the reciprocal of a q-fold Blaschke product. Thus h is a quotient of Blaschke products as desired. One can also show that h has no branch points on ∂D.

Now, apply the argument of the previous case to conclude that g is (p, q)-homotopic to the meromorphic function h, hence that f is (p, q)-homotopic to h. Since h is analytic across ∂D, it can be shown by elementary means to be a simplicial map (with appropriate choices of triangulations). This completes the proof. □

Recall that the total deficiency and total branch order are invariants under (p, q)-homotopies. Therefore, in proving the following result we need only consider simplicial mappings.

Theorem 5. *If $f\colon D \to \mathscr{X}$ is a (p, q)-map, then*

$$\beta_f + \delta_f \geq p + q - 1 - 2qg(\mathscr{X}),$$

where $g(\mathscr{X})$ is the genus of $\mathscr{X}$. This inequality is sharp for $g(\mathscr{X}) = 0$ or 1, and in any case $\beta_f + \delta_f \geq 1$. For $f \in BQ$, $\delta_f = 0$ and $\beta_f = p + q - 1$.

Proof. If f is in BQ, then we may assume it is a quotient of finite Blaschke

products; the results follow easily as in Example 1. Henceforth, assume $f \notin BQ$. In particular, $\beta_f + \delta_f \geq \delta_f \geq 1$.

We will treat $\mathscr{X}$ and $\widetilde{\mathscr{X}}$ as (topological) cell complexes, using some of the notation we established above. Let r be the number of components of $\mathscr{P}$ in $\mathscr{X}$. Since these are simply connected, they serve as 2-cells, and $\mathscr{X}$ is the 2-complex consisting of these r 2-cells along with the 1-complex K. Likewise, $\widetilde{D}$ is a 2-cell in $\widetilde{\mathscr{X}}$, and $\widetilde{\mathscr{X}}$ consists of this along with $\widetilde{K}$.

Recall the quantities δ_f, the total deficiency of f, and β_f, the total branch order of f. In addition, we establish the following notations:

(a) $g(\cdot)$ denotes the genus of a surface (or of its 2-complex).

(b) $\chi(\cdot)$ denotes the Euler characteristic of a complex, vertices minus edges plus faces. Recall that $\chi(\cdot) = 2 - 2g(\cdot)$.

(c) $\beta_{\widetilde{K}}$ and β_π denote the total branch orders of the projection π on $\widetilde{K}$ and on $\widetilde{T}$, respectively.

(d) v and e denote the number of vertices and edges, respectively, of K.

(e) $\tilde{v}$ and $\tilde{e}$ denote the number of vertices and edges, respectively, of $\widetilde{K}$.

Note that every vertex x of K lies in $\mathscr{E}$, hence is a deficient point of f and is covered $q - \delta_f(x)$ times (counting multiplicities) by vertices in $\widetilde{D}$. As it is covered a total of p times in $\widetilde{\mathscr{X}}$, it is covered $p - q + \delta_f(x)$ times by vertices in $\widetilde{K}$. Adding over all points of $\mathscr{E}$, which number v, and using the fact that the number of vertices in $\widetilde{K}$, counting multiplicities, is $\tilde{v} + \beta_{\widetilde{K}}$, we obtain

$$\delta_f = \tilde{v} + \beta_{\widetilde{K}} - (p - q)v.$$

Observe that all branch points of π lie in $\widetilde{D}$ or in $\widetilde{K}$, giving the identity

$$\beta_f + \beta_{\widetilde{K}} = \beta_\pi.$$

By the Riemann-Hurwitz relation,

$$\beta_\pi = 2[(g(\widetilde{\mathscr{X}}) - 1) - p(g(\mathscr{X}) - 1)] = 2p - 2 + 2g(\widetilde{\mathscr{X}}) - 2pg(\mathscr{X}).$$

Since we have only the one 2-cell in $\widetilde{T}$,

$$2g(\widetilde{\mathscr{X}}) = 1 - \chi(\widetilde{K}) = 1 - \tilde{v} + \tilde{e}.$$

Finally, note that each edge of K is covered by $p - q$ edges in $\widetilde{K}$, so

$$\tilde{e} = (p - q)e.$$

Applying these equalities in succession, we obtain

$$
\begin{aligned}
\beta_f + \delta_f &= \beta_\pi + \tilde{v} - (p-q)v \\
&= 2p - 2 + 2g(\widetilde{\mathscr{X}}) - 2pg(\mathscr{X}) + \tilde{v} - (p-q)v \\
&= 2p - 1 - \tilde{v} + \tilde{e} - 2pg(\mathscr{X}) + \tilde{v} - (p-q)v \\
&= 2p - 1 + (p-q)(e-v) - 2pg(\mathscr{X}) \\
&= 2p - 1 + (p-q)(e-v-r+2) - 2pg(\mathscr{X}) + (p-q)(r-2) \\
&= 2p - 1 - 2qg(\mathscr{X}) + (p-q)(r-2) \\
&= p + q - 1 - 2qg(\mathscr{X}) + (p-q)(r-1) \\
&\geq p + q - 1 - 2qg(\mathscr{X}).
\end{aligned}
$$

Example 2 shows this inequality is sharp for the genus 1 case. Examples 1 and 4 show it is sharp in the genus zero case, both for BQ and non-BQ functions. This completes the proof. □

We used $r > 0$ to get the final inequality above, so the result could be improved with better information on the number of components of $\mathscr{P}$. We will discuss this further in the closing comments. For now, let us list some results from earlier papers which, when put in our terminology, follow from Theorem 5.

The main line we have taken up begins with Theorem 2 in Brannan-Kirwan [3], then the Theorem of Lyzzaik-Styer [8], Theorem 1 of Ortel-Smith [10], Theorem 1 of Carne-Ortel-Smith [5], and Theorem 2 of Brannan-Lyzzaik [4]. In these results, $\mathscr{X}$ was the Riemann sphere, and ∞ was omitted, giving a deficiency of q there. With minor adjustments, such as moving between the plane and the disc with homeomorphisms, these results are subsumed by Theorem 5. The result closest to ours is Theorem 2 of Brannan-Lyzzaik [4]. Among other things, we see that the "arcwise connected" and the "locally univalent" hypotheses were not needed there. Their proof of the last statement of that theorem, regarding branch values, goes through in our setting, although we have not pursued it.

The results of Srebro [12] and Srebro-Wajnryb [13, 14] concern mainly multiply connected domains and $q = p - 1$. In the case of the disc, certain of these follow from our results; for instance, (i) and (ii)(a) of Covering Theorem 2. The "a-coverings" of Srebro-Wajnryb are, when they have the disc as domain, BQ-maps with $q = p - 1$. Therefore, Corollary 5.5 of [13] is a special case of conclusion (b) of our Lemma 6.

6. FURTHER EXAMPLES.

Examples are important in valence studies. In this section we present several examples which show the need for a variety of our hypotheses; they also show both the scope and the limits of our results.

Other domains. Our functions have had the unit disc D as domain, but one may certainly consider other Riemann surfaces. First, note that we cannot replace D by a compact surface, since it is well known that the valence function will then be constant (counting multiplicities) on $\mathscr{X}$. As for simply connected open Riemann surfaces, they are homeomorphic, hence interchangeable since we are working with open continuous mappings. If we impose a conformal structure on the domain which makes the function under consideration analytic, then we must distinguish between the disc and the plane, but our work is otherwise unaffected. An instance of this is Theorem 1 of Brannan-Kirwan [3], which requires analyticity rather than just open and continuous. An open continuous map can take every value in the plane exactly $p > 1$ times; start with $z \to z^p$ and follow it by a homeomorphism of D onto C. However, the cluster set would consist of the single point at infinity; and for a holomorphic function, this would contradict the fact that it is of bounded characteristic.

For surfaces which are noncompact and are not simply connected, our results do not apply directly, although many of the techniques should be useful. The example below illustrates new phenomena which can arise with a finitely-connected domain Ω in place of D.

Example 5. Let $\mathscr{X}$ be the Riemann sphere, $p = 3$, $q = 1$, and consider the surface illustrated in Figure 7. For purposes of display, the plane is drawn as a disc, its outer boundary being identified as the point ∞; the smaller disc represents the unit disc. The image surface is constructed from the five pieces $\Omega_1, \dots, \Omega_5$ (each with its natural projection to $\mathscr{X}$) attached via branch points as indicated: in particular, the annuli Ω_1 and Ω_2 are attached by cutting each on a segment from a point over the unit circle to a point over x_1 and cross connecting; Ω_3 is then attached to Ω_2 by cutting to a point over x_2 and cross connecting; the smaller discs, Ω_4 and Ω_5, are connected to one another and to Ω_3 by cutting each on a segment running between points over x_3 and x_4 and cross connecting, to form two double branch points.

Denote the resulting surface by Ω, with projection $\rho: \Omega \to \mathscr{X}$. It is planar and has 5 boundary components, so there is a homeomorphism $\phi: R \to \Omega$, where R is a circle domain; namely, the unit disc with four closed subdiscs removed. Define $f: R \to \mathscr{X}$ by $f \equiv \rho \circ \phi$. Then Ω is the image surface of f and we can make the following observations:

(1) f assumes each value on the unit circle once, this is labeled as the set $\mathscr{C}$.

(2) f omits the value ∞, so $\mathscr{E} = \{\infty\}$.

(3) f assumes all other values in $\mathscr{X}$ three times, counting multiplicities; the set $\mathscr{P}$ has two components, P_1 and P_2.

(4) f has total branch order $\beta_f = 6$ and total deficiency $\delta_f = 1$.

Note that $\mathscr{C} \cup \mathscr{E}$ is no longer connected, components of $\mathscr{P}$ may be multiply connected, etc. To proceed as we have in this paper, we would hope to embed this image surface in a compact 3-fold covering surface $(\widetilde{\mathscr{X}}, \pi)$ of $\mathscr{X}$. The
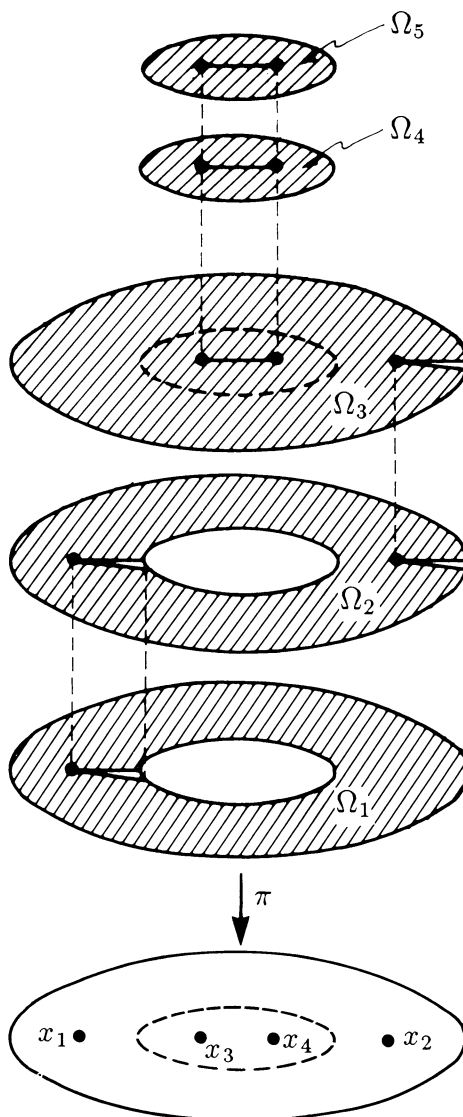
FIGURE 7

boundary of Ω which lies over the unit circle has three components. Compactification would require attaching these edges in a consistent manner. However, since two wrap around the circle in a one-to-one fashion while the other wraps around it two-to-one, compactification requires the inclusion of an unexpected branch point over $\mathcal{Q}$. The details are left to the interested reader. □

Note that there are results by Srebro [12] which use an embedding theorem for functions defined on finitely-connected domains. However, these cases correspond with $q = p - 1$; in particular, the obstruction to embedding which we designed into this example (namely, inconsistent wrapping numbers) cannot arise. The need for an extra branch point in the above example seems related to the multiple connectivity of a component of $\mathcal{P}$. The authors feel that examples

like this may foreshadow a version of our embedding theorem which applies to more general domains.

Hypotheses. In this section we construct a class of examples showing the need for several of our hypotheses. In all these, $p = 2$ and $q = 1$, $\mathscr{X}$ is the Riemann sphere C^∞, and ∞ is an omitted point. The basic construction material consists of the following: Ω is a copy of the plane from which the half-lines $l_n = \{z: \Re z = n, \ \Im z \geq 0\}$ have been removed; and for each integer n, Ω_n is the vertical strip $\{z: n - \frac{1}{4} < \Re z < n + \frac{1}{4}\}$, also with the half-line l_n removed.

Example 6. Construct a simply-connected Riemann surface S by attaching the Ω_n to Ω; in particular, identify one edge of l_n on Ω_n to the opposite edge of l_n on Ω, for each integer n. See the illustration in Figure 8. The function $f: D \to C^\infty$ having this as its image surface would qualify as a $(2, 1)$-map except that the points $\{\infty\} \cup \{0, \pm 1, \pm 2, \ldots\}$ are omitted; that is, f has infinitely many exceptional points. Observe that the 1-set of f is the union of the half-lines l_n and the strips $\{z: n + \frac{1}{4} \leq \Re z \leq n + \frac{3}{4}\}$. Thus, $\mathscr{Q}$ has infinitely many components, each having more than one exceptional point in its closure. The embedding theorem fails for f, since if we embed S in a 2-fold covering surface of C or C^∞, it is evident that branch points will be introduced over the points $z = n$. Were the covering surface compact, these branch points would have an accumulation point. The example shows the need for the finiteness hypothesis on $\mathscr{E}$. Much more sophisticated examples are possible; with infinitely many exceptional points, almost anything is possible.   □

Example 7. We perform the same construction as above, only now we also attach the remaining edge of the slit on each of the sheets Ω_n to its opposite
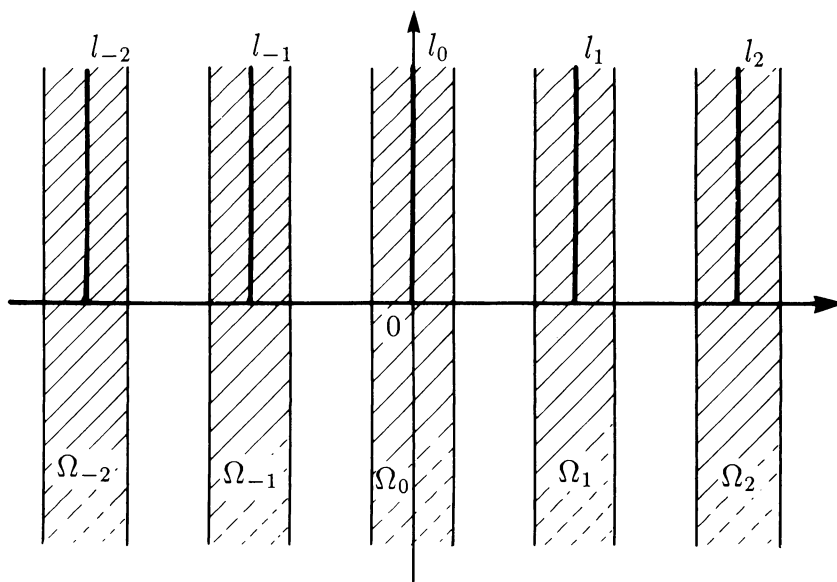

FIGURE 8

on Ω, introducing thereby a simple branch point at each point $z = n$. Here ∞ is the only exceptional point, but there are infinitely many branch values. As above, the embedding theorem fails. $\square$

Example 8. Proceed as in Example 6; however, attach only a finite number of the strips Ω_n, say for $-N \le n \le N$. The resulting function f is a locally univalent $(2, 1)$-map and the embedding theorem does hold. Note that the compact covering surface $\widetilde{\mathscr{X}}$ in which the image surface is embedded must have branch points over $z = n$, $n = -N, \ldots, N$. By the Riemann-Hurwitz formula, $\widetilde{\mathscr{X}}$ is not the Riemann sphere. Here, f can be chosen to be holomorphic and $f \equiv \pi \circ \phi$, where $\phi: D \to \widetilde{\mathscr{X}}$ is analytic in the Riemann surface sense, but ϕ is not a classical holomorphic function since its range is not in the Riemann sphere. Our point is that even in the locally univalent holomorphic setting, more general analytic functions may necessarily become involved. $\square$

We have assumed throughout our work that $\mathscr{Q}$ is nonempty. Of course, one can always define q to be the next smallest valence after p. The only concern might be the case that the next smallest valence is 0, but then, $\mathscr{Q} \cup \mathscr{E} = \mathscr{E}$ which, being connected, must be a singleton. Thus f assumes each value in its range p times, so the range is $\mathscr{P}$, a simply connected planar set. Up to possible homeomorphisms on the domain and range, f is simply a polynomial mapping of degree p.

7. COMMENTS AND QUESTIONS

First, some comments about (p, q)-maps on D: The main question remaining concerns the inequality of Theorem 5, which cannot be sharp for $g(\mathscr{X}) > 1$ since the right-hand side may be negative. What is needed is an estimate for r, the number of components of $\mathscr{P}$, which is better than the value $r = 1$ we used there. Suppose we consider a (p, q)-map f obtained by the method of Example 3, which can be used with any p, q, and $g(\mathscr{X}) \ge 1$. It requires only a single exceptional point, which is omitted, implying $\delta_f = q$. The construction involves p distinct copies of $\mathscr{X}$ and k families of curves, $F_1, F_2, \ldots, F_k$. Each family is removed from exactly $p - q$ copies of $\mathscr{X}$, while each copy of $\mathscr{X}$ has at least one family removed from it. Consequently, k will be the smallest integer for which $k(p - q) \ge p$. Each family generates the fundamental group of $\mathscr{X}$, hence contains $2g(\mathscr{X})$ curves. One can verify that in this situation, $r = 2g(\mathscr{X})(k - 1) + 1$. If $s \ge 0$ is such that $k(p - q) = p + s$, then the equalities in the proof of Theorem 5 lead to

$$\beta_f + \delta_f = p + q - 1 + 2sg(\mathscr{X}).$$

For $g(\mathscr{X}) > 1$, this seems to be the most "efficient" (p, q)-map from D to $\mathscr{X}$. Though we do not know if it is the sharp lower bound on $\beta_f + \delta_f$, we can show by an inductive pruning method that the sharp lower bound will occur for a (p, q)-map having a single exceptional point.

Now, some comments about broader valence issues. In reviewing the papers of the bibliography, one is struck by the diversity of questions and formulations; it is difficult at times to see the interconnections. It is clear, however, that replacing general mappings by simplicial ones is a worthwhile goal. Examples in this paper show that one cannot expect our results *per se* to extend; nonetheless, several of the techniques should be widely applicable. Among these are the depth function, the notion of metric uniformity for fattened valence sets, and especially the use of restricted types of homotopies which maintain valence structure. For instance, the proof of Proposition 5 concerning the structure of $\mathscr{C}$ in components of the standard fattened q-set and the resulting regularization using homotopies should go through in more general circumstances — only later is the fact that the image surface lies in a compact covering surface needed. Example 5 suggests that there may be useful extensions of the embedding technique. Indeed, the finitely connected domain setting is probably the next one where simplicial maps should be investigated.

In concentrating on p and q, we have largely ignored the fact that the exceptional points also have valences. This was exploited, for instance, by Brannan and Lyzzaik [3], and in the combinatorial setting of our simplicial maps, similar results should be accessible.

Finally, there are natural analogues to the questions we have discussed for functions defined on higher dimensional domains. As a first instance, one might ask whether there are any locally univalent open continuous maps of the ball in R^n to itself having exactly two valences p, and q. Even the initial steps in our approach are not clear. Is the p-set simply connected?

References

1. L. V. Ahlfors and L. Sario, *Riemann surfaces*, Princeton Univ. Press, Princeton, N. J., 1960.
2. A. Beurling, *Ensembles exceptionnels*, Acta Math. **72** (1940), 1–13.
3. D. A. Brannan and W. E. Kirwan, *Some covering theorems for analytic functions*, J. London Math. Soc. (2) **19** (1979), 93–101.
4. D. A. Brannan and A. K. Lyzzaik, *Some covering properties of locally-univalent functions*, Ann. Acad. Sci. Fenn. Ser. A I Math. **13** (1988), 3–23.
5. K. Carne, M. Ortel, and W. Smith, *Multiplicities of locally one-to-one continuous maps of the plane*, Bull. London Math. Soc. **18** (1987), 438–442.
6. E. F. Collingwood and A. J. Lohwater, *The theory of cluster sets*, Cambridge Univ. Press, 1966.
7. L. V. Keldyš, *Topological imbeddings in Euclidean space*, Proc. Steklov Inst. Math. **8** (1966).
8. A. Lyzzaik and D. Styer, *A covering surface conjecture of Brannan and Kirwan*, Bull. London Math. Soc. **14** (1982), 39–42.
9. W. S. Massey, *Algebraic topology: An introduction*, Graduate Texts in Math., vol. 56, Springer-Verlag, New York, 1977.
10. M. Ortel and W. Smith, *A covering theorem for continuous locally univalent maps of the plane*, Bull. London Math. Soc. **18** (1986), 359–363.
11. U. Srebro, *Deficiences of immersions*, Pacific J. Math. **113** (1984), 493–496.
12. _____, *Covering theorems for meromorphic functions*, J. Analyse Math. **44** (1984/85), 235–250.

13. U. Srebro and B. Wajnryb, *Covering theorems for Riemann surfaces*, J. Analyse Math. **46** (1986), 283–303.

14. _____, *Covering theorems for open surfaces,* Proceedings of the 1985 Georgia Topology Conference: Geometry and Topology, Dekker, Amsterdam, 1987, pp. 265–275.

15. S. Stoïlow, *Principes topologiques de la theorie des fonctions analytiques*, Gauthier-Villars, Paris, 1938.

DEPARTMENT OF MATHEMATICS, KUWAIT UNIVERSITY, P.O. BOX 5969, 13060-SAFAT-KUWAIT

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TENNESSEE, KNOXVILLE, TENNESSEE 37996-1300