

ITERATION AND THE SOLUTION OF FUNCTIONAL EQUATIONS FOR FUNCTIONS ANALYTIC IN THE UNIT DISK

BY

CARL C. COWEN¹

ABSTRACT. This paper considers the classical functional equations of Schroeder $f \circ \varphi = \lambda f$, and Abel $f \circ \varphi = f + 1$, and related problems of fractional iteration where φ is an analytic mapping of the open unit disk into itself. The main theorem states that under very general conditions there is a linear fractional transformation Φ and a function σ analytic in the disk such that $\Phi \circ \sigma = \sigma \circ \varphi$ and that, with suitable normalization, Φ and σ are unique. In particular, the hypotheses are satisfied if φ is a probability generating function that does not have a double zero at 0. This intertwining relates solutions of functional equations for φ to solutions of the corresponding equations for Φ . For example, it follows that if φ has no fixed points in the open disk, then the solution space of $f \circ \varphi = \lambda f$ is infinite dimensional for every nonzero λ . Although the discrete semigroup of iterates of φ usually cannot be embedded in a continuous semigroup of analytic functions mapping the disk into itself, we find that for each z in the disk, all sufficiently large fractional iterates of φ can be defined at z . This enables us to find a function meromorphic in the disk that deserves to be called the infinitesimal generator of the semigroup of iterates of φ . If the iterates of φ can be embedded in a continuous semigroup, we show that the semigroup must come from the corresponding semigroup for Φ , and thus be real analytic in t . The proof of the main theorem is not based on the well known limit technique introduced by Koenigs (1884) but rather on the construction of a Riemann surface on which an extension of φ is a bijection. Much work is devoted to relating characteristics of φ to the particular linear fractional transformation constructed in the theorem.

1. Introduction. In this paper we study iteration and solution of functional equations for analytic functions mapping the open unit disk D into itself. In the main theorem (Theorem 3.2) we show that, under very general conditions, such a map can be intertwined with a linear fractional transformation. Thus, problems of iteration and solution of functional equations for the given function are related to the analogous problems for linear fractional transformations, which are generally easier to solve.

Results of this sort are well known and usually depend on the nature of the fixed points of the given function. Although a mapping φ of D into D need not have a fixed point, an old theorem due to A. Denjoy [5] and J. Wolff [25] gives a replacement. Let $\varphi_0(z) = z$, $\varphi_1(z) = \varphi(z)$ and $\varphi_n(z) = \varphi(\varphi_{n-1}(z))$ for $n = 2, 3, 4, \dots$, that is, φ_n is the n th iterate of φ . By a linear fractional (Möbius)

Received by the editors July 23, 1979 and, in revised form, August 4, 1980.

1980 *Mathematics Subject Classification.* Primary 30D05; Secondary 39B10, 60J80.

Key words and phrases. Functional equation, iteration, analytic function, semigroup, infinitesimal generator, Galton-Watson process.

¹Supported in part by National Science Foundation grant MCS-7902018.

© 1981 American Mathematical Society
0002-9947/81/0000-0204/\$07.75

transformation, we mean a nonconstant function of the form

$$\Phi(z) = (\alpha z + \beta)(\gamma z + \delta)^{-1}$$

where α, β, γ and δ are complex numbers.

THEOREM D-W ([5], [25]). *Suppose φ , not a Möbius transformation of D onto D , is analytic and maps D into D . Then there is a point a , $|a| \leq 1$, so that the sequence φ_n , $n = 1, 2, 3, \dots$, converges to a uniformly on compact subsets of D . Moreover, if $|a| = 1$, then $\lim_{r \rightarrow 1^-} \varphi(ra) = a$ and $s = \lim_{r \rightarrow 1^-} \varphi'(ra)$ exists and $0 < s \leq 1$.*

Throughout this paper, we will call the point a referred to in this theorem the *Denjoy-Wolff point* of φ . If φ is continuous at a , then clearly a is a fixed point of φ . In the case $|a| = 1$, an important step in the proof of this theorem (and a fact we need later) is that φ maps small disks in D , tangent to the unit circle at a , into themselves.

The main result of the paper is the following theorem.

THEOREM 3.2. *Let φ be an analytic mapping of D into D , φ nonconstant and not a Möbius transformation of D onto D , and let a be the Denjoy-Wolff point of φ . If $\varphi'(a) \neq 0$, then there is a fundamental set V for φ on D , a domain Ω , either the complex plane or the unit disk, a linear fractional transformation Φ mapping Ω onto Ω , and an analytic mapping σ of D into Ω such that φ and σ are univalent on V , $\sigma(V)$ is a fundamental set for Φ on Ω , and $\Phi \circ \sigma = \sigma \circ \varphi$. Moreover, Φ is unique up to conjugation by a linear fractional transformation mapping Ω onto Ω , and Φ and σ depend only on φ , not on the particular fundamental set V .*

The idea of “fundamental set” in the theorem is a technical concept introduced so that we may describe a set of points near the Denjoy-Wolff point small enough that φ is well behaved, but large enough that $\varphi_n(z)$ is eventually in this set. That $\sigma(V)$ is a fundamental set for Φ gives the uniqueness of Φ . We give the definition.

DEFINITION. If ψ maps a domain Δ into itself, we say V is a *fundamental set* for ψ on Δ if V is an open, connected, simply connected subset of Δ such that $\psi(V) \subset V$ and for every compact set K in Δ , there is a positive integer N so that $\psi_N(K) \subset V$.

Clearly the Denjoy-Wolff point a is in the closure of V if V is any fundamental set for φ on D .

We will see (remark following Theorem 3.2) that the Φ and Ω of the conclusion of Theorem 3.2 fall into four essentially different cases:

- (1) $\Omega = \mathbf{C}$, $\Phi(z) = sz$, $0 < |s| < 1$,
- (2) $\Omega = \mathbf{C}$, $\Phi(z) = z + 1$,
- (3) $\Omega = D$, $\Phi(z) = [(1 + s)z + (1 - s)][(1 - s)z + (1 + s)]^{-1}$, $0 < s < 1$, or
- (4) $\Omega = D$, $\Phi(z) = [(1 \pm 2i)z - 1][z - 1 \pm 2i]^{-1}$.

(Case (3) may be easier to understand in its equivalent guises as translation by 1 on a horizontal strip or multiplication by s on the right half-plane. In the same way, case (4) is equivalent to translation by 1 on the upper (−) or lower (+) half-plane.) A major portion of §3 is devoted to determining which of these cases occurs for a particular φ . Loosely put, we find case (1) occurs when a is in D and $\varphi'(a) = s$, case (3) occurs when $|a| = 1$ and $\varphi'(a) = s < 1$, and cases (2) or (4) occur when

$|a| = 1$ and $\varphi'(a) = 1$. Deciding which of cases (2) or (4) occurs is a fairly subtle problem. A concept related to this distinction is that of nontangential convergence for a sequence.

DEFINITION. Suppose $\{z_n\}_{n=1}^{\infty} \subset D$ and $\lim_{n \rightarrow \infty} z_n = a$, $|a| = 1$. We say the sequence z_n , $n = 1, 2, 3, \dots$, converges to a nontangentially if $\sup_n |\operatorname{Arg}(a - z_n)| < \pi/2$.

If $|a| = 1$ and for some z_0 in D , the sequence $\varphi_n(z_0)$, $n = 1, 2, 3, \dots$, converges to a nontangentially, then case (2) or case (3) occurs according as $\varphi'(a) = 1$ or $\varphi'(a) < 1$.

We should observe that there is a reason why $\varphi'(a) = 1$ is a more difficult case. A fixed point of φ is a zero of the function $\varphi(z) - z$; a is a multiple zero of this function if and only if $\varphi'(a) = 1$. (More generally, difficulties arise when $|\varphi'(a)| = 1$ because $\varphi'(a)^n = 1$ means a is a multiple zero of $\varphi_n(z) - z$, but, except when φ is a linear fractional transformation, $|\varphi'(a)| = 1$, $\varphi'(a) \neq 1$ cannot occur for a function mapping D into itself.)

A motivation for proving Theorem 3.2 is its relation to the solution of functional equations of the form $f \circ \varphi = g \circ f$ where g and φ are given and f is the unknown function. (Schroeder's equation $f \circ \varphi = \lambda f$ and Abel's equation $f \circ \varphi = f + 1$ are two well-known examples.) If F is a solution of $F \circ \Phi = g \circ F$ for the linear fractional transformation Φ , then $f = F \circ \sigma$ is a solution of $f \circ \varphi = g \circ f$ where $\Phi \circ \sigma = \sigma \circ \varphi$. Since Schroeder's and Abel's equations are easily solved for linear fractional transformations, we get results for more general functions. For example (Theorem 4.7) if φ has no fixed points in D then the solution set of the equation $f \circ \varphi = \lambda f$ is infinite dimensional for every $\lambda \neq 0$.

A second motivation is that we can sometimes use Theorem 3.2 to define "fractional iterates" of φ , that is, a real semigroup of functions φ_t for $t \geq 0$, where $\varphi_1 = \varphi$ and $\varphi_{s+t} = \varphi_s \circ \varphi_t$. Since fractional iterates can be defined for linear fractional transformations, we try to define φ_t by $\Phi_t \circ \sigma = \sigma \circ \varphi_t$. But there are difficulties because, in general, σ is not univalent, and even if it were, σ^{-1} need not be defined on $\Phi_t(\sigma(D))$ unless $\sigma(D)$ meets some geometric condition. We do find, though, that for each z in D , $\varphi_t(z)$ can be defined for t sufficiently large. This is enough that we can find a meromorphic function on the disk that deserves to be called the infinitesimal generator of the semigroup. If φ is an analytic mapping of D into itself that is real valued on the interval $(-1, 1)$ and x_1 and x_2 are such that $-1 \leq x_1 \leq a \leq x_2 \leq 1$ and $\varphi'(x) > 0$ for $x_1 < x < x_2$, then $\varphi_t(x)$ can be defined for $x_1 < x < x_2$ and $t \geq 0$ to be real analytic in both variables (corollary to Theorem 5.1). On the other hand, we show (Theorem 5.2) that if φ_t is a continuous semigroup, then φ is univalent, the fixed points of φ_t (including those on ∂D) are the same for all t , and φ_t arises as above. A consequence of the latter is that φ_t is real analytic in t .

The history of this circle of ideas is very long. In his paper of 1871 [21], Schroeder considers the problem of fractional iteration and relates it to the solution of various functional equations. In 1884, Koenigs [14] constructed a solution of Schroeder's equation $f \circ \varphi = sf$ near a fixed point a when φ is analytic at a and

$\varphi'(a) = s$, $|s| \neq 0, 1$. In the years since these classic works, much has been written; the interested reader might consult the bibliography at the end of this paper or that of Kuczma [15]. For the most part, the literature has concentrated on cases when φ is analytic in a neighborhood of the fixed point or dropped analyticity completely and considered continuous function in \mathbf{R}^n .

These problems are related to the theory of discrete branching processes (Galton-Watson processes) in probability theory through the probability generating function. The probability generating function for a Galton-Watson process is the function $\varphi(z) = \sum_{k=0}^{\infty} p_k z^k$ where $p_k \geq 0$ and $\sum p_k = 1$. (Harris' book [8] begins with a very readable discussion of these ideas.) The coefficient p_k is to be interpreted as the probability that an individual will have k offspring and the k th coefficient of φ_n is interpreted as the probability that there will be k individuals in the n th generation. Questions concerning the eventual population size are related to the asymptotic behavior of the iterates of φ . We note that every probability generating function for a Galton-Watson process is an analytic function mapping D into D , and if $p_0 + p_1 \neq 0$, φ satisfies the hypotheses of Theorem 3.2. The bibliographies of Harris [8] and Athreya and Ney [1] are a guide to work on branching processes.

In view of the extensive literature concerning these problems, it seems reasonable to inquire what contribution this paper makes. In case the Denjoy-Wolff point a is in D or when φ is a probability generating function, little has been added to the statements of the results, merely a slight reduction of hypotheses or a slight improvement in the conclusions in certain cases. However, the cases for which the Denjoy-Wolff point a is on the unit circle and φ is not analytic at a seem to have been considered rarely except when φ is a probability generating function. In part this is probably due to our specific interest in functions analytic in the unit disk whereas most of the existing literature deals with functions analytic everywhere, nowhere, or on some (unspecified) neighborhood of the fixed point.

Perhaps the most important contribution of this paper is the introduction of a more geometric point of view. In most of the existing literature, the techniques used are modifications of the limit technique introduced by Koenigs [14], namely, consideration of $\lim_{n \rightarrow \infty} s^{-n}(\varphi_n(z) - a)$ where $s = \varphi'(a)$. This technique works well when φ is analytic at a and $|s| \neq 0, 1$, but is harder to apply in other cases. In particular, this necessitates special consideration of the case $\varphi'(a) = 1$.

In the proof of Theorem 3.2 we begin with the more geometric observation that φ is univalent on a fundamental set V . For each z in V , $\varphi_n(z)$ is a point of V for $n = 0, 1, 2, 3, \dots$; we construct a Riemann surface by adjoining to V ideal points $\varphi_n(z)$, $n = -1, -2, -3, \dots$, and extend the action of φ to a univalent map of the surface onto itself. This surface is simply connected and noncompact, so is equivalent to either the unit disk or the whole plane and this equivalence transforms the extension of φ into a linear fractional transformation. Thus, in this proof, the case $\varphi'(a) = 1$ is treated in the same way as $\varphi'(a) \neq 1$. The distinction comes in that, if φ is analytic at a and $\varphi'(a) = 1$, there is no small disk centered at a that φ maps into itself, so that a must be on the boundary of a fundamental set for φ on

any region, whereas if $0 < |\varphi'(a)| < 1$, there is a small disk centered at a that is a fundamental set for φ on some region.

The remainder of the paper is divided into five sections. In §2, we present some lemmas needed in the later sections. In §3, we prove the main theorem and several other theorems which explain the conclusion of the main theorem.

In §§4 and 5 we use the results of §3 to give complete solutions (in terms of σ) to Schroeder's and Abel's functional equations and to discuss fractional iteration. In §6 we point out some problems raised by this work.

I would like to thank David Drasin, Richard Penney and Allen Weitsman of the Purdue Mathematics Department for many helpful conversations and Gerald T. Cargo of Syracuse University for the use of the (unpublished) notes of his talk given in the fall of 1978 [4]. I would also like to express gratitude to Christian Pommerenke for his help in improving the first version of this work (see [20]), and to Maurice Heins for pointing out an error.

2. Some lemmas. In this section we will prove some lemmas needed in later sections. Although the statements are perhaps not well known, their proofs involve standard techniques. The reader may wish to take these results on faith and proceed immediately to §3.

We record the following well-known theorem of Julia and Carathéodory. A proof may be found in Nevanlinna's book [16].

THEOREM J-C. *Suppose φ is an analytic mapping of D into D and b is a point on the unit circle such that $\lim_{r \rightarrow 1^-} \varphi(rb) = b$. Then $\lim_{z \rightarrow b} (b - \varphi(z))/(b - z) = c$ exists and $\lim_{z \rightarrow b} \varphi'(z) = c$ when z tends to b nontangentially. Moreover, $0 < c < \infty$ and*

$$\frac{1 - |\varphi(z)|^2}{|b - \varphi(z)|^2} \geq \frac{1}{c} \frac{1 - |z|^2}{|b - z|^2} \quad \text{for all } z \text{ in } D.$$

The latter inequality says φ maps disks tangent to the unit circle at b into other disks (smaller if $c < 1$) tangent to the unit circle at b .

LEMMA 2.1. *Suppose φ is analytic, maps D into D , and has Denjoy-Wolff point a , $|a| = 1$. If $\lim_{r \rightarrow 1^-} \varphi'(ra) = s < 1$, then for any z in D , the sequence $\varphi_n(z)$, $n = 1, 2, 3, \dots$, converges to a nontangentially.*

PROOF. The transformation $w = (a + z)(a - z)^{-1}$ maps the disk D onto the right half-plane $\operatorname{Re} \zeta > 0$, and

$$\tilde{\varphi}(w) = \left[a + \varphi(a(w - 1)(w + 1)^{-1}) \right] \left[a - \varphi(a(w - 1)(w + 1)^{-1}) \right]^{-1}$$

is the map of the right half-plane into itself that corresponds to φ . We have $\lim_{x \rightarrow \infty} \tilde{\varphi}(x) = \infty$, $c = \lim_{x \rightarrow \infty} \tilde{\varphi}'(x) = s^{-1} > 1$ and Nevanlinna's proof of the theorem of Julia and Carathéodory [16, pp. 57-60] shows that $c = \inf_{x > 0} x^{-1} \operatorname{Re}(\tilde{\varphi}(x + iy))$. We give the proof for $\tilde{\varphi}$ on the right half-plane.

Given z_0 in the right half-plane, write z_n for $\tilde{\varphi}_n(z_0)$ and let $z_n = x_n + iy_n$. Thus we have $x_{n+1} = \operatorname{Re} \tilde{\varphi}(z_n) \geq cx_n$ for $n = 0, 1, 2, \dots$. Let $r = |(z_0 - z_1)(\bar{z}_0 + z_1)^{-1}|$. Since z_1 is in the half-plane $\operatorname{Re} \zeta > cx_0$, we find that $c < (1 + r)(1 - r)^{-1}$.

We need the following, but omit the straightforward proof.

Claim. Suppose $0 < r < 1 < c < (1 + r)(1 - r)^{-1}$. If $\zeta = x + iy$ is in the non-euclidean disk centered at $\alpha = u + iv$, $|(\alpha - \zeta)(\bar{\alpha} + \zeta)^{-1}| \leq r$, and the half-plane $\text{Re } \zeta \geq cu$, then $|y - v|(x - u)^{-1} \leq [r^2(c + 1)^2 - (c - 1)^2]^{1/2} (c - 1)^{-1}(1 - r)^{-1} = M$.

By the Schwarz-Pick theorem we have

$$|(z_n - z_{n+1})(\bar{z}_n + z_{n+1})^{-1}| \leq |(z_{n-1} - z_n)(\bar{z}_{n-1} + z_n)^{-1}| < \dots < r$$

for $n = 1, 2, 3, \dots$.

The claim shows that $|y_{n+1} - y_n|(x_{n+1} - x_n)^{-1} \leq M$ for $n = 0, 1, 2, \dots$ from which it follows that

$$\begin{aligned} |y_{N+1} - y_0|(x_{N+1} - x_0)^{-1} &\leq \left(\sum_{k=0}^N |y_{k+1} - y_k| \right) \left(\sum_{k=0}^N x_{k+1} - x_k \right)^{-1} \\ &\leq \max_k |y_{k+1} - y_k|(x_{k+1} - x_k)^{-1} \leq M. \end{aligned}$$

Thus, the sequence $\tilde{\varphi}_n(z_0)$, $n = 1, 2, 3, \dots$, tends to infinity inside the sector $|\text{Arg}(\zeta - z_0)| \leq \arctan M$. \square

LEMMA 2.2. *Suppose φ is analytic, maps D into D , and has Denjoy-Wolff point a , $|a| = 1$. If, for some z_0 in D , the sequence $\varphi_n(z_0)$, $n = 1, 2, 3, \dots$, converges to a nontangentially, then for any compact set K in D , the sequence $\varphi_n(K)$, $n = 1, 2, 3, \dots$, converges to a nontangentially.*

PROOF. As in the proof of Lemma 2.1, we find it more convenient to prove the corresponding assertion for $\tilde{\varphi}$ in the right half-plane.

If $\zeta = x + iy$ and $\alpha = u + iv$ satisfy $|(\zeta - \alpha)(\bar{\zeta} + \alpha)^{-1}| \leq r < 1$, then

$$x \geq u(1 - r)(1 + r)^{-1} \quad \text{and} \quad |y| \leq |v| + 2ru(1 - r^2)^{-1}$$

so that $|y/x| \leq (|v/u| + 2)(1 - r)^{-2}$.

Now suppose for some z_0 in the right half-plane, the sequence $\tilde{\varphi}_n(z_0)$, $n = 1, 2, 3, \dots$, converges nontangentially to ∞ , that is, letting

$$\tilde{\varphi}_n(z_0) = z_n = x_n + iy_n,$$

that $|y_n/x_n| \leq M < \infty$ for all positive integers n .

If K is a compact subset of the right half-plane, then there is a positive number r , $r < 1$, so that $|(z - z_0)(\bar{z} + z_0)^{-1}| \leq r$ for all z in K . It follows from the Schwarz-Pick theorem that $|(z - z_n)(\bar{z} + z_n)^{-1}| \leq r$ for all z in $\tilde{\varphi}_n(K)$. From the above we see that if $z = x + iy$ is in $\tilde{\varphi}_n(K)$ for some n , then $|y/x| \leq (M + 2)(1 - r)^{-2}$, so that $\tilde{\varphi}_n(K)$ converges nontangentially to infinity. \square

The following lemma is due to Wolff [26], Warschawski [24], and Noshiro [17] but we include its simple proof.

LEMMA 2.3. *If U is a convex open subset of the plane, f is analytic on U and $\text{Re } f'(z) > 0$ for all z in U , then f is univalent on U .*

PROOF. Suppose z and w are in U . Then

$$\begin{aligned} f(z) - f(w) &= \int_w^z f'(\xi) d\xi = \int_0^1 f'(tz + (1-t)w)(z-w) dt \\ &= (z-w) \int_0^1 f'(tz + (1-t)w) dt. \end{aligned}$$

But $\operatorname{Re}[\int_0^1 f'(tz + (1-t)w) dt] > 0$ so $f(z) = f(w)$ if and only if $z = w$. \square

The following lemma is proved under more restrictive hypotheses in Kamowitz's *Spectra of composition operators on H^p* [9, Lemmas 1.1 and 1.2].

LEMMA 2.4. *Suppose φ , not a Möbius transformation of D onto D , is analytic, maps D into D , and has Denjoy-Wolff point a . If $b \neq a$, $|b| = 1$, and $\lim_{r \rightarrow 1^-} \varphi(rb) = b$ then $\lim_{r \rightarrow 1^-} \varphi'(rb) > 1$.*

PROOF. We note that by the theorem of Julia and Carathéodory, $\lim_{r \rightarrow 1^-} \varphi'(rb) = q$ exists and $0 < q \leq \infty$.

Suppose first that a is in D . Without loss of generality we assume $a = 0$ and $b = 1$. Let $\psi(z) = z^{-k}\varphi(z)$ where k is the order of the zero of φ at 0, so that $k \geq 1$, ψ is an analytic mapping of D into D , and $\lim_{r \rightarrow 1^-} \psi(r) = 1$. Thus $\lim_{r \rightarrow 1^-} \varphi'(r) = \lim_{r \rightarrow 1^-} kr^{k-1}\psi(r) + r^k\psi'(r) = k + \lim_{r \rightarrow 1^-} \psi'(r) > 1$ (since $\varphi \not\equiv z$).

Now if $|a| = 1$, we may assume, without loss of generality that $a = -1$ and $b = 1$. Let $\psi(z) = \frac{1}{2}(\varphi(z) - \varphi(-z))$. Then ψ is an analytic mapping of D into D , $\psi(0) = 0$ and $\lim_{r \rightarrow 1^-} \psi(r) = 1$. (Moreover, $\psi \not\equiv z$ since $\varphi \not\equiv z$.) By the above, $\lim_{r \rightarrow 1^-} \psi'(r) > 1$. The definition of ψ yields $\lim_{r \rightarrow 1^-} \varphi'(ra) + \lim_{r \rightarrow 1^-} \varphi'(rb) > 2$, but since a is the Denjoy-Wolff point of φ , $\lim_{r \rightarrow 1^-} \varphi'(ra) \leq 1$ and our conclusion follows. \square

Let d be the "noneuclidean distance" function on the disk,

$$d(\alpha, \beta) = |\alpha - \beta| |1 - \alpha\bar{\beta}|^{-1}.$$

The properties of this function are well known, for example there is the "triangle inequality" $d(\alpha_1, \alpha_2) = r_1 < 1$ and $d(\alpha_2, \alpha_3) = r_2 < 1$ implies

$$d(\alpha_1, \alpha_3) \leq (r_1^2 + r_2^2)^{1/2} (1 + r_1^2 r_2^2)^{-1/2} < 1.$$

The Schwarz-Pick theorem says that if φ is an analytic mapping of D into itself then $d(\varphi(\alpha), \varphi(\beta)) \leq d(\alpha, \beta)$. We will also need the following lemma.

LEMMA 2.5. *Suppose $\{z_n\}_{n=1}^\infty \subset D$ and $\lim_{n \rightarrow \infty} z_n = 1$. The sequence z_n converges to 1 nontangentially if and only if $\sup_n d(z_n, \bar{z}_n) < 1$.*

PROOF. For $k > 0$, the set $\{z | d(z, \bar{z}) < (1 + k^2)^{-1/2}\}$ is the eye-shaped intersection of the disks with centers $\pm ik$ and radius $(1 + k^2)^{1/2}$. \square

The following result is due to Pommerenke [20, Theorem 2].

THEOREM P. *Let f be an analytic function mapping the right half-plane into itself with angular derivative $c \geq 1$ at ∞ . Let S be a Stolz angle at ∞ , $0 < \delta < 1$, and $0 < \lambda < \infty$, and let $f_n(1) = x_n + iy_n$. If $G = S \cup \bigcup_{n=1}^\infty \{x + iy : \delta x_n \leq x < \infty \text{ and } |y - y_n| \leq \lambda x_n\}$ then there is $\rho > 0$ so that f is univalent on $G \cap \{w : |w| > \rho\}$.*

3. The construction. In this section we will prove the main result of the paper (Theorem 3.2). Proposition 3.1 constructs a fundamental set needed in the main theorem and Theorems 3.3–3.8 amplify the conclusions somewhat. Throughout this section, φ will be a nonconstant analytic mapping of the unit disk D into itself, not an elliptic Möbius transformation of D onto D , and a will denote the Denjoy-Wolff point of φ .

DEFINITION. If ψ maps a domain Δ into itself, we say V is a *fundamental set* for ψ on Δ if V is an open, connected, simply connected subset of Δ such that $\psi(V) \subset V$ and for every compact set K in Δ , there is a positive integer N so that $\psi_N(K) \subset V$.

PROPOSITION 3.1. *Let φ be an analytic mapping of D into D with Denjoy-Wolff point a . If $\varphi'(a) \neq 0$ then there is a fundamental set V for φ on D such that φ is univalent on V . Moreover, if for some point z_0 in D , the sequence $\varphi_n(z_0)$, $n = 1, 2, 3, \dots$, converges to a nontangentially, then V may be chosen so that it contains small sectors with vertex at a and opening θ for all $\theta < \pi$.*

PROOF. If the Denjoy-Wolff point is in D , then $0 < |\varphi'(a)| < 1$ and we can find ϵ , $0 < \epsilon < 1 - |a|$, so that if $V = \{z : |z - a| < \epsilon\}$ then $\varphi(V) \subset V$ and φ is one-to-one on V . By Theorem D-W, V is a fundamental set for φ on D .

Henceforth we assume $|a| = 1$. By the theorem of Julia and Carathéodory (Theorem J-C), $s = \lim_{r \rightarrow 1^-} \varphi'(ra)$ exists and $0 < s \leq 1$. We give the proof in two steps: we use Theorem P to show that for every compact set K in D , there is an integer N so that φ is univalent on $\cup_{n=N}^\infty \varphi_n(K)$ and then use this condition to construct V .

Let $\mu(z) = (a + z)(a - z)^{-1}$ and let $f(w) = \mu(\varphi(\mu^{-1}(w)))$. Then f maps the right half-plane into itself and has angular derivative $s^{-1} \geq 1$ at ∞ , so Theorem P applies. Given any compact set K in D , choose $r < 1$ so that z in K means $|z| < r$. We let $\delta = (1 - r)(1 + r)^{-1}$, $\lambda = 2r(1 - r)^{-3/2}(1 + r)^{-1/2}$, $D_r = \mu(\{z : |z| < r\})$, and $f_n(1) = w_n = x_n + iy_n$. Now the Schwarz-Pick inequality shows that $f_n(D_r) \subset \{w : |w - w_n| |w + \bar{w}_n|^{-1} < r\}$ and an elementary calculation shows that

$$\{w : |w - w_n| |w + \bar{w}_n|^{-1} < r\} \subset \{w : \operatorname{Re} w > \delta x_n \text{ and } |\operatorname{Im} w - y_n| < \delta x_n\}.$$

Thus for some N , $\cup_{n=N}^\infty \mu(\varphi_n(K)) \subset \cup_{n=N}^\infty f_n(D_r)$ is contained in the set $G \cap \{w : |w| > \rho\}$ and, by Theorem P, f univalent there, so φ is univalent on $\cup_{n=N}^\infty \varphi_n(K)$.

We now use this condition on compact sets to construct V . Let $K_k = \{z : |z| \leq 1 - 1/k\}$ and $K_k^0 = \{z : |z| < 1 - 1/k\}$ for $k = 2, 3, \dots$. Let m be the smallest integer so that $\varphi(0) \in K_m^0$.

Let N_m be a positive integer large enough that φ restricted to $\cup_{n=N_m}^\infty \varphi_n(K_m)$ is one-to-one and let $U_m = \cup_{n=N_m}^\infty \varphi_n(K_m^0)$. By the choice of m , U_m is connected, indeed, $\varphi(0)$ is in K_m^0 and $\varphi(K_m^0)$, so $\varphi_2(0)$ is in $\varphi(K_m^0)$ and $\varphi_2(K_m^0)$, etc., and each $\varphi_n(K_m^0)$ is connected.

Suppose that integers $N_m \leq N_{m+1} \leq \dots \leq N_{k-1}$ and open sets $U_m \subset U_{m+1} \subset \dots \subset U_{k-1}$ have been chosen inductively so that U_j is a connected subset of D ,

φ is univalent on U_j , $\varphi(U_j) \subset U_j$, and $\bigcup_{n=N_j}^{\infty} \varphi_n(K_j^0) \subset U_j$ for $j = m, m + 1, \dots, k - 1$. Let N'_k be an integer so that $N'_k \geq N_{k-1}$ and φ restricted to $\bigcup_{n=N'_k}^{\infty} \varphi_n(K_k)$ is one-to-one. Now the set $L = U_{k-1} \setminus \left(\bigcup_{n=N'_k}^{\infty} \varphi_n(K_k^0) \right)$ is a compact subset of D so $\varphi(L)$ is also. Since φ_n converges uniformly on K_k to the constant a (Theorem D-W) and a is not in $\varphi(L)$, we can find $N_k > N'_k$ so that $\left(\bigcup_{n=N_k}^{\infty} \varphi_n(K_k^0) \right) \cap \varphi(L) = \emptyset$. Let $U_k = U_{k-1} \cup \left(\bigcup_{n=N_k}^{\infty} \varphi_n(K_k^0) \right)$. Clearly we have U_k is a connected subset of D , $\varphi(U_k) \subset U_k$, and $\bigcup_{n=N_k}^{\infty} \varphi_n(K_k^0) \subset U_k$. We claim that φ restricted to U_k is one-to-one. Suppose α, β are in U_k with $\alpha \neq \beta$. If both α and β are in U_{k-1} or both α and β are in $\bigcup_{n=N'_k}^{\infty} \varphi_n(K_k^0)$ then clearly $\varphi(\alpha) \neq \varphi(\beta)$, so we suppose α is in L and β is in $\bigcup_{n=N_k}^{\infty} \varphi_n(K_k^0)$. (The reader should note the occurrence of N'_k and N_k .) But if α is in L , $\varphi(\alpha)$ is in $\varphi(L)$ which is disjoint from $\bigcup_{n=N_k}^{\infty} \varphi_n(K_k^0)$ so that $\varphi(\alpha) \neq \varphi(\beta)$.

Thus we construct the open sets U_k inductively for $k = 2, 3, \dots$. Let $V' = \bigcup_{k=m}^{\infty} U_k$, and let $V = V' \cup$ (holes of V'). By the construction, V' is a connected open subset of D , $\varphi(V') \subset V'$ and φ is univalent on V' . By the maximum principle and the argument principle $\varphi(V) \subset V$ and φ is univalent on V as well, so V is a fundamental set for φ on D such that φ is univalent on V .

We now suppose z_0 is in D and $\varphi_n(z_0)$ converges to a nontangentially. We sketch a slightly different proof to establish the special form of V . Without loss of generality we assume $a = 1$. Let d be the distance-like function defined in §2,

$$d(\alpha, \beta) = |\alpha - \beta| |1 - \alpha\bar{\beta}|^{-1}.$$

For $n = 1, 2, 3, \dots$, let $c_n = 1 - 1/n$ and $z_n = \varphi_n(0)$. By the hypothesis, Lemma 2.2 and our simplifying assumption that $a = 1$, the sequence z_n converges to 1 nontangentially, which means (Lemma 2.5) $c_0 = \sup_n \{d(z_n, \bar{z}_n)\} < 1$. For positive integers n and j , let $\Delta_{n,j} = \{z | d(z, z_n) < c_j\}$. For a fixed integer j , since $d(z_n, \bar{z}_n) < c_0$ and $d(z, z_n) = d(\bar{z}, \bar{z}_n) < c_j$ for z in $\bigcup_{n=1}^{\infty} \Delta_{n,j}$, it follows that there is $\bar{c} < 1$ so that $d(z, \bar{z}) < \bar{c}$ for z in $\bigcup_{n=1}^{\infty} \Delta_{n,j}$. Therefore, for each j , $\bigcup_{n=1}^{\infty} \Delta_{n,j}$ is contained in a sector of opening less than π with vertex at 1, and by Theorem J-C and Lemma 2.3, we can find an integer N so that φ is univalent on $\bigcup_{n=N}^{\infty} \Delta_{n,j}$.

Let m be the smallest integer such that $|\varphi(0)| < c_m$ (so that z_{n-1}, z_n , and z_{n+1} are all contained in $\Delta_{n,j}$ for $j \geq m$) and let N_m be large enough that φ is univalent on $U_m = \bigcup_{n=N_m}^{\infty} \Delta_{n,m}$. Thus U_m is connected, open, simply-connected, and by the Schwarz-Pick theorem, $\varphi(U_m) \subset U_m$.

Suppose integers $N_m \leq N_{m+1} \leq \dots \leq N_{k-1}$ and open sets $U_m \subset U_{m+1} \subset \dots \subset U_{k-1}$ have been chosen inductively so that U_j is an open, connected, simply-connected subset of D , φ is univalent on U_j , $\varphi(U_j) \subset U_j$, and $U_j \supset \bigcup_{n=N_j}^{\infty} \Delta_{n,j}$ for $j = m, m + 1, \dots, k - 1$. Let N'_k be an integer so that $N'_k \geq N_{k-1}$ and φ restricted to $\bigcup_{n=N'_k}^{\infty} \Delta_{n,k}$ is univalent. As above, letting $L = U_{k-1} \setminus \left(\bigcup_{n=N'_k}^{\infty} \Delta_{n,k} \right)$, choose $N_k \geq N'_k$ so that $\varphi(L) \cap \left(\bigcup_{n=N_k}^{\infty} \Delta_{n,k} \right) = \emptyset$, and let $U_k = U_{k-1} \cup \left(\bigcup_{n=N_k}^{\infty} \Delta_{n,k} \right)$. We show as before that U_k has the desired properties and continue the induction for $k \geq n$. We see that $V = \bigcup_{k=m}^{\infty} U_k$ is a fundamental set for φ on D on which φ is univalent.

In addition, we claim that given an angle $\theta < \pi$, then V contains a small sector with vertex at 1 and opening θ . Indeed, suppose $\delta < 1$ is given; we must find $r < 1$ so that if $\operatorname{Re} w > r$ and $d(w, \bar{w}) < \delta$ then $w \in V$. Let $u = \operatorname{Re} w$ and $x_n = \operatorname{Re} z_n$. We have $d(z_n, x_n) \leq d(z_n, \bar{z}_n) < c_0$ and, since $d(z_n, z_{n+1}) < c_m$, we can find $q < 1$ so that $d(x_n, x_{n+1}) < q$ for all positive integers n . Therefore, we can find $q^* < 1$ so that if $d(w, \bar{w}) < \delta$ and $d(u, x_k) < q$ for some integer k , then $d(z_k, w) < q^*$. But since $d(x_n, x_{n+1}) < q$ for all n , and $\lim_{n \rightarrow \infty} x_n = 1$, if $u \geq x_N$ then $d(u, x_k) < q$ for some $k \geq N$. Now find an integer j so that $c_j > q^*$ and let $r = x_{N_j}$. By the above, if $u = \operatorname{Re} w > r$ and $d(w, \bar{w}) < \delta$ then $d(u, x_k) < q$ for some $k > N_j$ so $d(w, z_k) < q^* < c_j$ and $w \in \Delta_{k,j} \subset V$. \square

THEOREM 3.2. *Let φ be an analytic mapping of D into D , φ nonconstant and not a Möbius transformation of D onto D , and let a be the Denjoy-Wolff point of φ . If $\varphi'(a) \neq 0$, then there is a fundamental set V for φ on D , a domain Ω , either the complex plane or the unit disk, a linear fractional transformation Φ mapping Ω onto Ω , and an analytic mapping σ of D into Ω such that φ and σ are univalent on V , $\sigma(V)$ is a fundamental set for Φ on Ω , and $\Phi \circ \sigma = \sigma \circ \varphi$. Moreover, Φ is unique up to conjugation by a linear fractional transformation mapping Ω onto Ω , and Φ and σ depend only on φ , not on the particular fundamental set V .*

PROOF. Since $\varphi'(a) \neq 0$, Proposition 3.1 shows there is a fundamental set V for φ on D such that φ is univalent on V . We will construct an abstract Riemann surface \mathfrak{S} and a map Ψ of \mathfrak{S} onto itself which corresponds to Ω and Φ .

1. *The construction of the Riemann surface \mathfrak{S} .* We introduce \mathfrak{S} as a point set, topologize it, and give it an analytic structure. For this purpose we will restrict our attention to the action of φ on V . Some points of V are not in the image of φ , some are not in the image of $\varphi \circ \varphi$, et cetera. We may think of \mathfrak{S} as being made from V by gluing on (abstract) preimages of points in V so that every point of \mathfrak{S} is in the image of φ_n .

If m and n are integers and z and w are in V we say (z, n) is equivalent to (w, m) and write $(z, n) \sim (w, m)$ if there is an integer $k, k \geq \max\{-n, -m\}$, so that $\varphi_{k+n}(z) = \varphi_{k+m}(w)$. Since φ is one-to-one on V , $\varphi_{k+n}(z) = \varphi_{k+m}(w)$ if and only if $\varphi_{k'+n}(z) = \varphi_{k'+m}(w)$ for any k and k' not smaller than $-n$ and $-m$. The relation \sim is an equivalence relation between pairs; we use the symbol $[(z, n)]$ to denote the equivalence class containing (z, n) . Let $\mathfrak{S} = \{[(z, n)] \mid z \in V, n \text{ an integer}\}$. The pair (z, n) should be thought of as representing $\varphi_n(z)$.

If U is open in V and n is an integer, let $\mathcal{U}_n^U = \{[(z, n)] \mid z \in U\}$. We claim that the sets $\{\mathcal{U}_n^U\}$ form a basis for a Hausdorff topology on \mathfrak{S} . Suppose $z_j^* = [(z_j, n_j)]$, $j = 1, 2$ and $z_1^* \neq z_2^*$. Let $k = \max\{-n_1, -n_2\}$. We have $z_j^* = [(\varphi_{k+n_j}(z_j), k)]$ since $(z_j, n_j) \sim (\varphi_{k+n_j}(z_j), -k)$ and $\varphi_{k+n_1}(z_1) \neq \varphi_{k+n_2}(z_2)$, since $z_1^* \neq z_2^*$. Choose U_1 and U_2 disjoint open sets in V containing $\varphi_{k+n_1}(z_1)$ and $\varphi_{k+n_2}(z_2)$ respectively. Then $\mathcal{U}_{-k}^{U_1}$ and $\mathcal{U}_{-k}^{U_2}$ are disjoint neighborhoods of z_1^* and z_2^* respectively. Now suppose w^* is in $\mathcal{U}_{m_1}^{W_1} \cap \mathcal{U}_{m_2}^{W_2}$, that is suppose $w^* = [(w_1, m_1)] = [(w_2, m_2)]$ where $w_j \in W_j$ for $j = 1, 2$. Thus we have $\varphi_{k+m_1}(w_1) = \varphi_{k+m_2}(w_2)$ where $k = \max\{-m_1, -m_2\}$. Since $\mathcal{U}_n^U = \mathcal{U}_{-k}^{\varphi_k^{-1}(U)}$, if $Y = \varphi_{k+m_1}(W_1) \cap \varphi_{k+m_2}(W_2)$ then $\mathcal{U}_{-k}^Y \subset \mathcal{U}_{m_1}^{W_1} \cap \mathcal{U}_{m_2}^{W_2}$

and w^* is in \mathcal{U}_{-k}^Y . Thus $\{\mathcal{U}_n^U\}$ is the basis for a Hausdorff topology on \mathfrak{S} (see, for example, [6, p. 67]).

We introduce an analytic structure on \mathfrak{S} by defining the coordinate maps $c_n: V \rightarrow \mathfrak{S}$ given by $c_n(z) = [(z, n)]$. The map c_n is easily seen to be one-to-one (since φ is) and continuous. Since V is locally compact and \mathfrak{S} is Hausdorff, c_n is a homeomorphism of V onto $c_n(V)$. If $n = m + l$ where $l \geq 0$ then $c_m^{-1}c_n$ is defined on V and $c_m^{-1}c_n(z) = c_m^{-1}([(z, n)]) = c_m^{-1}([(z, m + l)]) = c_m^{-1}[(\varphi_l(z), m)] = \varphi_l(z)$, so $c_m^{-1}c_n$ is analytic on V . Since φ_l is one-to-one on V , $c_n^{-1}c_m = \varphi_l^{-1}$ which is analytic on $\varphi_l(V)$, where it is defined. Thus \mathfrak{S} is a Riemann surface.

\mathfrak{S} is simply connected, for suppose $\gamma: [0, 1] \rightarrow \mathfrak{S}$ is a loop. Since $\gamma([0, 1])$ is compact, we can find an integer n so that $\gamma([0, 1])$ is in the open set $c_n(V)$. But $c_n(V)$ is homeomorphic to V , which is simply connected, so γ is null homotopic in $c_n(V)$.

\mathfrak{S} is not compact since if z is any point in V different from a then the sequence $z_k^* = [(z, -k)]$, $k = 1, 2, 3, \dots$, does not have a convergent subsequence. Indeed, suppose U is an open subset of V with compact closure and suppose n is a fixed integer. Then $z_k^* \in \mathcal{U}_n^U$ means $z \in \varphi_{n+k}(U)$, for $k > -n$, but φ_m converges uniformly on U to $a \neq z$. Thus, the sequence z_k^* can have at most finitely many terms in any neighborhood with compact closure, and cannot have a convergent subsequence.

Therefore, the Riemann surface \mathfrak{S} is analytically equivalent to either the unit disk or the complex plane. (We will see below that both cases can occur.)

II. *The maps Ψ and π .* Define the map $\pi: V \rightarrow \mathfrak{S}$ by $\pi(z) = [(z, 0)]$ and the map $\Psi: \mathfrak{S} \rightarrow \mathfrak{S}$ by $\Psi([(z, n)]) = [(\varphi(z), n)]$. One easily verifies that Ψ is well defined, that π and Ψ are analytic and that π and Ψ are one-to-one. Ψ is surjective since $[(z, n)] = [(\varphi(z), n - 1)] = \Psi([(z, n - 1)])$. Clearly $\Psi \circ \pi = \pi \circ \varphi$.

We note that $\pi(V)$ is a fundamental set for Ψ on \mathfrak{S} . Indeed if K is a compact subset of \mathfrak{S} , then we can find an integer n so that $K \subset c_n(V)$ and $\Psi_{|n}(K) \subset \pi(V)$.

III. *The maps Φ and σ , and the domain Ω .* We have already noted that the Riemann surface \mathfrak{S} is analytically equivalent to either the unit disk or the complex plane: let Ω denote this domain, and let ρ be a Riemann map of \mathfrak{S} onto Ω . We define σ on V by $\sigma = \rho \circ \pi$ and define Φ on Ω by $\Phi = \rho \circ \Psi \circ \rho^{-1}$. One easily verifies that Φ is a one-to-one map of Ω onto itself (hence a linear fractional transformation), that σ is a univalent map of V into Ω , that $\sigma(V)$ is a fundamental set for Φ , and that $\Phi \circ \sigma = \sigma \circ \varphi$.

We extend σ to all of D by letting $\sigma(z) = \Phi_{-n}(\sigma(\varphi_n(z)))$ where n is a positive integer large enough that $\varphi_n(z)$ is in V , and Φ_{-n} denotes the n th iterate of Φ^{-1} . The extension is well defined because, for any positive integer k , $\Phi_{-k} \circ \sigma \circ \varphi_k = \sigma$ on V so that if $\varphi_m(z)$ and $\varphi_n(z)$ are in V and $m = n + k$ then $\Phi_{-m}(\sigma(\varphi_m(z))) = \Phi_{-n}(\Phi_{-k}(\sigma(\varphi_k(\varphi_n(z)))) = \Phi_{-n}(\sigma(\varphi_n(z)))$. The extension is single valued, analytic, and maps D into Ω because Φ_{-n} , σ , and φ_n are single valued and analytic and Φ^{-1} maps Ω into Ω .

To prove uniqueness, we suppose that \tilde{V} is a fundamental set for φ on D such that φ is univalent on \tilde{V} and that $\tilde{\Omega}$, $\tilde{\Phi}$ and $\tilde{\sigma}$ are the corresponding objects as

above. Let $K = \{t\varphi(0) | 0 \leq t \leq 1\}$. Now K is compact, connected and has the property that for each N , $\cup_{n=N}^{\infty} \varphi_n(K)$ is connected. Since V and \tilde{V} are each fundamental sets for φ , there is an integer N so that $\cup_{n=N}^{\infty} \varphi_n(K) \subset V \cap \tilde{V}$. Let W be the connected component of $V \cap \tilde{V}$ that contains $\cup_{n=N}^{\infty} \varphi_n(K)$. Clearly W is a fundamental set for φ on D . It follows that $\sigma(W)$ and $\tilde{\sigma}(W)$ are fundamental sets for Φ and $\tilde{\Phi}$ on Ω and $\tilde{\Omega}$ respectively. Define $\tau: \Omega \rightarrow \tilde{\Omega}$ by $\tau(z) = \tilde{\Phi}_n^{-1}(\tilde{\sigma}(\sigma^{-1}(\Phi_n(z))))$ where n is an integer large enough that $\Phi_n(z) \in W$. The map τ is well defined since if $\Phi_n(z)$ is in W and p is a positive integer then

$$\tilde{\Phi}_{n+p}^{-1}(\tilde{\sigma}(\sigma^{-1}(\Phi_{n+p}(z)))) = \tilde{\Phi}_{n+p}^{-1}(\tilde{\sigma}(\varphi_p(\sigma^{-1}(\Phi_n(z)))) = \tilde{\Phi}_n^{-1}(\tilde{\sigma}(\sigma^{-1}(\Phi_n(z)))).$$

One easily verifies that τ is a univalent map of Ω onto $\tilde{\Omega}$, so $\Omega = \tilde{\Omega}$, τ is a linear fractional transformation and $\tilde{\Phi} = \tau \circ \Phi \circ \tau^{-1}$. We note also that $\tilde{\sigma} = \tau \circ \sigma$ on V .

□

In the next section, we will give complete solutions of Schroeder's and Abel's functional equations for certain special linear fractional transformations Φ and use the above theorem to draw conclusions for general maps of the disk into the disk.

REMARK. In the theorem above we showed that Ω and Φ are unique up to conjugacy and depend only on φ . Naturally, if we are given φ we would like to know what Ω and Φ are. By an examination of fixed points, we show that σ can be defined at the Denjoy-Wolff point a so as to be continuous on $V \cup \{a\}$ and that Ω and Φ can be chosen from one of the four cases below. (Which of these cases applies is the subject of Theorems 3.3–3.6.)

Case 1. $\Omega = \mathbb{C}$, $\sigma(a) = 0$, and $\Phi(z) = sz$ where $0 < |s| < 1$.

Case 2. $\Omega = \mathbb{C}$, $\sigma(a) = \infty$, and $\Phi(z) = z + 1$.

Case 3. $\Omega = D$, $\sigma(a) = 1$, and $\Phi(z) = [(1 + s)z + (1 - s)][(1 - s)z + (1 + s)]^{-1}$ where $0 < s < 1$.

Case 4. $\Omega = D$, $\sigma(a) = 1$, and $\Phi(z) = [(1 \pm 2i)z - 1][z - 1 \pm 2i]^{-1}$.

We note that if p is a fixed point of Φ in Ω , then $p = \rho([(z, n)])$ where $[(z, n)]$ is a fixed point of Ψ on \mathfrak{S} , that is, that $(z, n) \sim (\varphi(z), n)$, hence that z is a fixed point of φ in V . Conversely, if φ has any fixed points in D , the Denjoy-Wolff point a is the only one. In this case, a is in V , $(a, n) \sim (a, m)$ for all integers n, m so that $[(a, n)]$ is a fixed point for Ψ on \mathfrak{S} and $p = \rho([(a, n)])$ is a fixed point for Φ in Ω . Summarizing, we see that Φ has a fixed point in Ω if and only if φ has a fixed point in D , and in this case the fixed point of Φ is $\sigma(a)$.

A linear fractional transformation ψ has two fixed points p, p' on the sphere and $\psi'(p)\psi'(p') = 1$ or else ψ has one double fixed point p and $\psi'(p) = 1$. If $|\psi'(p)| < |\psi'(p')|$ and $w \neq p'$ then $\lim_{n \rightarrow \infty} \psi_n(w) = p$. If $\psi'(p) = 1$, then $\lim_{n \rightarrow \infty} \psi_n(w) = p$ for all w in the sphere. If $|\psi'(p)| = 1$ but $\psi'(p) \neq 1$ then ψ is a rotation of the sphere and for every point w in the sphere the sequence $\psi_n(w)$, $n = 1, 2, 3, \dots$, has w as a limit point.

If n is a positive integer and z is in V , we have $\Phi_n(\sigma(z)) = \sigma(\varphi_n(z))$. Since the sequence $\varphi_n(z)$, $n = 1, 2, 3, \dots$, does not have z as a limit point if $z \neq a$, we see that the sequence $\Phi_n(\sigma(z))$ cannot have $\sigma(z)$ as a limit point, so that Φ is not a rotation. If we define $\sigma(a)$ by $\sigma(a) = \lim_{n \rightarrow \infty} \Phi_n(\sigma(z))$ for any z in V , then σ is continuous on $V \cup \{a\}$.

Elementary calculations show that if a linear fractional transformation ψ maps the disk onto the disk and is not a rotation of the sphere then the fixed points are on the unit circle and $\psi'(p) > 0$.

Conjugating Φ by an appropriate linear fractional transformation to move its fixed points, we reduce to one of the four cases mentioned above. (The functions Φ of Case 4 with the positive and negative signs are inverses of each other, and are not conjugate to each other by a linear fractional transformation mapping the disk onto itself.) We note that in Cases 1 and 3, $\Phi'(\sigma(a)) = s$ and that in Cases 2 and 4, $\Phi'(\sigma(a)) = 1$.

Each of the four cases can actually occur. For Cases 1, 3 and 4 we may take $\varphi = \Phi$ and $\sigma(z) = z$. For Case 2, we may take $\varphi(z) = (1+z)(3-z)^{-1}$ and $\sigma(z) = (1+z)(1-z)^{-1}$.

Deciding which of the four cases a particular φ falls into may be difficult, but we do have the following results.

THEOREM 3.3. *Let φ , Φ , and Ω be as in Theorem 3.2. Then $\Omega = \mathbf{C}$ and $\Phi(z) = sz$ (or conjugate to this) if and only if the Denjoy-Wolff point a is in D and $\varphi'(a) = s$.*

PROOF. Case 1 of the above remark is the only case for which Φ has a fixed point in Ω . Since fixed points of Φ in Ω correspond to fixed points of φ in D , we see that this case occurs if and only if a is in D . Now σ is analytic at a and, since σ is univalent on V , $\sigma'(a) \neq 0$. We have $\sigma'(a)\varphi'(a) = \sigma'(\varphi(a))\varphi'(a) = (\sigma \circ \varphi)'(a) = (\Phi \circ \sigma)'(a) = s\sigma'(a)$ so that $\varphi'(a) = s$. \square

THEOREM 3.4. *Let φ , Ω , Φ and σ be as in Theorem 3.2 and suppose the Denjoy-Wolff point a has modulus 1. Suppose, in addition, either that φ' can be defined at a so as to be continuous on $D \cup \{a\}$ or that for some z_0 in D , the sequence $\varphi_n(z_0)$, $n = 1, 2, 3, \dots$, converges to a nontangentially. Then $\Phi'(\sigma(a)) = \lim_{r \rightarrow 1^-} \varphi'(ra)$.*

COROLLARY. *If φ is an analytic mapping of the disk into the disk with Denjoy-Wolff point a on the unit circle and $\lim_{r \rightarrow 1^-} \varphi'(ra) = s < 1$, then we may take $\Omega = D$ and $\Phi(z) = [(1+s)z + (1-s)][(1-s)z + (1+s)]^{-1}$ in the conclusion of Theorem 3.2.*

We recall that by Theorem J-C, if $|a| = 1$ then $s = \lim_{r \rightarrow 1^-} \varphi'(ra)$ exists and $0 < s \leq 1$.

PROOF. If U is a simply connected region and β a point of U , let $G_U(z, \beta)$ denote Green's function for U with pole at β . In the proof of the theorem we will estimate Green's functions for V and $\sigma(V)$ and show

$$\begin{aligned} \varphi'(a) &= \lim_{n \rightarrow \infty} [G_V(\varphi_n(0), \beta)]^{1/n} \\ &= \lim_{n \rightarrow \infty} [G_{\sigma(V)}(\Phi_n(\sigma(0)), \sigma(\beta))]^{1/n} = \Phi'(\sigma(a)). \end{aligned}$$

We begin by noting a few elementary facts required below. (See, for example, [23].) If U is a simply connected region and ω is a univalent map defined on U then $G_U(z, \beta) = G_{\omega(U)}(\omega(z), \omega(\beta))$. If $U \subset W$ then $G_U(z, \beta) \leq G_W(z, \beta)$. If K is a compact subset of U and β and β' are in K , then there are constants C_1 and C_2 so

that $C_1 G_U(z, \beta) \leq G_U(z, \beta') \leq C_2 G_U(z, \beta)$, for z not in K . Therefore, if $z_n, n = 1, 2, 3, \dots$, is a sequence converging to a point on the boundary of U , $\lim_{n \rightarrow \infty} [G_U(z_n, \beta)]^{1/n} = \lim_{n \rightarrow \infty} [G_U(z_n, \beta')]^{1/n}$ and we can afford to be somewhat cavalier about the location of the pole.

Without loss of generality, we assume $a = 1$, and we let $s = \lim_{r \rightarrow 1^-} \varphi'(r)$.

For ρ and θ positive, let $S_{\rho, \theta} = \{z \mid |z - 1| < \rho \text{ and } |\arg(1 - z)| < \theta\}$. Suppose φ is as in the hypothesis and for some z_0 in D , the sequence $\varphi_n(z_0), n = 1, 2, 3, \dots$, converges nontangentially to 1. By Lemma 2.2, we may take $z_0 = 0$ and we define θ^* to be the smallest angle such that $\varphi_n(0) \in S_{2, \theta^*}$ for all $n = 1, 2, 3, \dots$. Let V be the fundamental set for φ constructed in Proposition 3.1 that contains sectors. Then for any $\theta, \theta^* < \theta < \pi/2$, we can find ρ so that $S = S_{\rho, \theta} \subset V$. For a certain β' in S , we have

$$G_V(z, \beta') \geq G_S(z, \beta') = -\log \left| 1 - 2 \left(\frac{1-z}{\rho} \right)^{\pi/2\theta} \left[\left(\frac{1-z}{\rho} \right)^{\pi/2\theta} - \left(\frac{1-z}{\rho} \right)^{\pi/\theta} + 1 \right]^{-1} \right|.$$

Since the sequence $\varphi_n(0), n = 1, 2, 3, \dots$, converges to 1 nontangentially, there is a constant C so that $G_S(\varphi_n(0), \beta') \geq C|1 - \varphi_n(0)|^{\pi/2\theta}$, and it also follows, by Theorem J-C, that $\lim_{k \rightarrow \infty} |(1 - \varphi_{k+1}(0))/(1 - \varphi_k(0))| = s$. Now

$$\lim_{n \rightarrow \infty} |1 - \varphi_n(0)|^{1/n} = \lim_{n \rightarrow \infty} \left[\prod_{k=0}^{n-1} |(1 - \varphi_{k+1}(0))/(1 - \varphi_k(0))| \right]^{1/n} = s$$

so for any β in V we have

$$\begin{aligned} \lim_{n \rightarrow \infty} [G_V(\varphi_n(0), \beta)]^{1/n} &\geq \lim_{n \rightarrow \infty} [G_S(\varphi_n(0), \beta')]^{1/n} \\ &\geq \lim_{n \rightarrow \infty} [C|1 - \varphi_n(0)|^{\pi/2\theta}]^{1/n} = s^{\pi/2\theta}. \end{aligned}$$

Since this holds for every $\theta, \theta^* < \theta < \pi/2$, we have in fact that $\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), \beta)]^{1/n} \geq s$. On the other hand, $V \subset D$ so $G_V(z, \beta) \leq G_D(z, \beta)$ and $\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), \beta)]^{1/n} \leq \lim_{n \rightarrow \infty} [G_D(\varphi_n(0), 0)]^{1/n} = \lim_{n \rightarrow \infty} [-\log|\varphi_n(0)|]^{1/n} = \lim_{n \rightarrow \infty} (1 - |\varphi_n(0)|)^{1/n} = s$. Thus $\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), \beta)]^{1/n} = s$ when the iterates converge nontangentially.

Now suppose φ is as in the hypotheses and φ' can be defined at 1 to be continuous on $D \cup \{1\}$. Then, if $\varphi_n(0)$ is eventually in arbitrarily small disks tangent to the unit circle at 1, we may take the fundamental set V to be a disk $\Delta = \{z \mid |z - r| < 1 - r\}$ where $0 < r < 1$. Since

$$G_\Delta(z, r) = -\log|1 - (1 - r)^{-1}(1 - z)|$$

we have

$$\lim_{n \rightarrow \infty} [-\log|1 - (1 - r)^{-1}(1 - \varphi_n(0))|]^{1/n} = \lim_{n \rightarrow \infty} [G_V(\varphi_n(0), r)]^{1/n}.$$

Clearly, for large n , we have

$$\begin{aligned} \frac{1}{2} [1 - |1 - (1 - r)^{-1}(1 - \varphi_n(0))|] &\leq -\log|1 - (1 - r)^{-1}(1 - \varphi_n(0))| \\ &\leq 2 [1 - |1 - (1 - r)^{-1}(1 - \varphi_n(0))|]. \end{aligned}$$

Now, if $0 < s < 1$, the iterates $\varphi_n(0)$ converge nontangentially and there are constants C_1 and C_2 so that $C_1|1 - \varphi_n(0)| \leq 1 - |1 - (1 - r)^{-1}(1 - \varphi_n(0))| \leq C_2|1 - \varphi_n(0)|$, so we have $\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), r)]^{1/n} = \lim_{n \rightarrow \infty} |1 - \varphi_n(0)|^{1/n} = s$ as before. On the other hand, if $s = 1$, the iterates $\varphi_n(0)$ may not converge nontangentially and the above argument fails. However, $\varphi_n(0)$ is in successively smaller disks tangent to the unit circle at 1, and we have, in any case, a constant C so that $C|1 - \varphi_n(0)|^2 \leq 1 - |1 - (1 - r)^{-1}(1 - \varphi_n(0))|$ so we have

$$\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), r)]^{1/n} \geq \lim_{n \rightarrow \infty} |1 - \varphi_n(0)|^{2/n} = s^2 = 1.$$

Since $\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), r)]^{1/n} \leq 1$ a priori, we have $\lim_{n \rightarrow \infty} [G_V(\varphi_n(0), r)]^{1/n} = s$ in this case as well.

On the other hand, if $\varphi_n(0)$ is not eventually in arbitrarily small disks tangent to the unit circle at 1, then $\varphi'(1) = 1$. For convenience, we regard φ as mapping the upper half-plane into itself with fixed point at infinity. It follows from Theorem P and Proposition 3.1 that V contains half strips of arbitrary width and as in Cases 2 and 4 below

$$\lim_{n \rightarrow \infty} G_V(\varphi_n(0), \beta)^{1/n} = 1 = \varphi'(1).$$

We now turn our attention to Φ on Ω . By Theorem 3.3, we may exclude Case 1 of the remark, where $\Omega = \mathbb{C}$ and $\Phi(z) = sz$.

Suppose we have Case 3, $\Omega = D$ and

$$\Phi(z) = [(1 + \tilde{s})z + (1 - \tilde{s})][(1 - \tilde{s})z + (1 + \tilde{s})]^{-1}$$

for some \tilde{s} , $0 < \tilde{s} < 1$. Since $\sigma(V)$ is a fundamental set for Φ on D , $\sigma(V)$ contains sectors $S_{\rho, \theta}$, for θ arbitrarily close to $\pi/2$ and we show, as above, that $\lim_{n \rightarrow \infty} [G_{\sigma(V)}(\Phi_n(\sigma(0)), \sigma(\beta))]^{1/n} = \tilde{s} = \Phi'(1)$.

Suppose we have Case 2, $\Omega = \mathbb{C}$ and $\Phi(z) = z + 1$. For x and y arbitrary real numbers and $b > 0$, let $R_{x,y,b} = \{z | \operatorname{Re} z > x, y - b < \operatorname{Im} z < y + b\}$. Since $\sigma(V)$ is a fundamental set for Φ on \mathbb{C} , given $b > 0$, we can find x and y so that $R = R_{x,y,b} \subset \sigma(V)$. For an appropriate β' and complex constant ζ ,

$$\begin{aligned} G_R(\Phi_n(\sigma(0)), \beta') &= G_R(n + \sigma(0), \beta') \\ &= -\log \left[\left[\sinh\left(\frac{\pi}{2b}n + \zeta\right) - 1 \right] \left[\sinh\left(\frac{\pi}{2b}n + \zeta\right) + 1 \right]^{-1} \right]. \end{aligned}$$

An explicit calculation yields

$$\lim_{n \rightarrow \infty} [G_{\sigma(V)}(\Phi_n(\sigma(0)), \sigma(\beta))]^{1/n} \geq \lim_{n \rightarrow \infty} [G_R(n + \sigma(0), \beta')]^{1/n} = \exp(-\pi/2b).$$

Since b was an arbitrary positive number, we have

$$\lim_{n \rightarrow \infty} [G_{\sigma(V)}(\Phi_n(\sigma(0)), \sigma(\beta))]^{1/n} > 1,$$

and as before, we conclude this limit is equal to 1, i.e. equal to $\Phi'(\infty)$.

Suppose we have Case 4, $\Omega = D$ and $\Phi(z) = [(1 \pm 2i)z - 1][z - 1 \pm 2i]^{-1}$. By means of the mapping $\rho(z) = \pm i(z + 1)(z - 1)^{-1}$ we transform this case to translation by 1 on the upper (negative sign) or lower (positive sign) half-plane, i.e. $\rho \circ \Phi \circ \rho^{-1}(z) = z + 1$. Calculations similar to those of Case 2 above yield $\lim_{n \rightarrow \infty} [G_{\sigma(V)}(\Phi_n(\sigma(0)), \sigma(\beta))]^{1/n} = 1 = \Phi'(1)$.

Combining the results of the above calculations with the identity $\sigma(\varphi_n(0)) = \Phi_n(\sigma(0))$ and the mapping property of Green's functions, we obtain the desired result:

$$s = \lim_{n \rightarrow \infty} [G_V(\varphi_n(0), \beta)]^{1/n} = \lim_{n \rightarrow \infty} [G_{\sigma(V)}(\Phi_n(\sigma(0)), \sigma(\beta))]^{1/n} = \Phi'(\sigma(a)).$$

The corollary follows from the theorem because, by Lemma 2.1, if $s < 1$ then the sequence $\varphi_n(0), n = 1, 2, 3, \dots$, converges to a nontangentially. \square

We now prove a theorem which allows us, in some cases, to distinguish Case 2 and Case 4 of the remark.

THEOREM 3.5. *Let φ be an analytic mapping of the unit disk into itself with Denjoy-Wolff point a of modulus 1. Suppose that defining $\varphi'(a)$ to be 1 allows φ' to be continuous on $D \cup \{a\}$. If, for some point z_0 in D , the sequence $\varphi_n(z_0), n = 1, 2, 3, \dots$, converges to a nontangentially, then the domain Ω obtained from φ in Theorem 3.2 is the whole plane.*

PROOF. By Theorem 3.4, $\Phi'(\sigma(a)) = 1$ so either Case 2 or Case 4 of the remark applies. We will show that if Case 4 applies, then the sequence of iterates cannot converge nontangentially to a .

Under the given hypotheses, we may choose V to be a disk tangent to the unit circle ∂D , at a , and on a smaller such disk σ is analytic on the boundary circle except at a . Therefore, we may assume, without loss of generality, that $a = 1, V = D$, and σ is analytic on ∂D except at 1.

Now suppose $\Omega = D$ and Φ is as in Case 4, say,

$$\Phi(z) = [(1 - 2i)z - 1][z - 1 - 2i]^{-1}.$$

Using the mapping $\rho(z) = i(1 + z)(1 - z)^{-1}$, we transform this case into translation on the upper half-plane: $\tilde{\Phi}(z) = \rho \circ \Phi \circ \rho^{-1}(z) = z + 1$. Thus, letting $\tilde{\sigma} = \rho \circ \sigma$, we have $\tilde{\sigma}$ maps D onto a fundamental set for $\tilde{\Phi}$ on the upper half-plane, and $\tilde{\sigma} \circ \varphi = \tilde{\Phi} \circ \tilde{\sigma}$. Moreover, $\tilde{\sigma}$ maps $\partial D \setminus \{1\}$ onto an analytic curve in the upper half-plane and the image of $\Gamma = \{e^{i\theta} | -\pi < \theta < 0\}$ is a curve tangent to the positive real axis. (Since $\tilde{\sigma}(V)$ is a fundamental set for translation, $\tilde{\sigma}(\Gamma)$ is eventually below any ray parallel to the positive real axis.)

Let $h(z) = \text{Arg}[(1 - z)(1 + z)^{-1}] + \pi/2$. Then h is a positive harmonic function in $D, h(\Gamma) = 0$ and $h(\Gamma') = \pi$ (where $\Gamma' = \{e^{i\theta} | 0 < \theta < \pi\}$). It is easily seen that if γ is a smooth curve in D so that $\lim_{t \rightarrow 1^-} \gamma(t) = 1$ then γ makes an angle θ with the interval $(-1, 1)$ at 1 if and only if $\lim_{t \rightarrow 1^-} h(\gamma(t)) = \theta - \pi/2$. We will show that for any z_0 in D , then $\lim_{n \rightarrow \infty} h(\varphi_n(z_0)) = 0$, that is, the sequence $\varphi_n(z_0), n = 1, 2, 3, \dots$, converges to 1, tangent to Γ .

For x and b positive real numbers, let $R_{x,b} = \{z | \text{Re } z > x, 0 < \text{Im } z < b\}$, and let $g(z) = g_{x,b}(z) = \text{Arg}[\cosh(\pi b^{-1}(z - x)) - 1]$. Then g is a positive harmonic function in $R_{x,b}$ and $g(z) = 0$ if z is real, $z > x$ and $g(z) = \pi$ if z is any nonreal point of the boundary. Let $\tilde{h}(z) = h(\sigma^{-1}(z))$ for z in $\tilde{\sigma}(V)$.

Suppose z_0 is in D and $\tilde{\sigma}(z_0) = \alpha + i\beta$. Given $b > \beta$, find x large enough that $x > \alpha$, that the ray $\{z | \text{Re } z \geq x, \text{Im } z = b\}$ is in $\tilde{\sigma}(V)$, and that the segment

$\{z | \operatorname{Re} z = x, 0 \leq \operatorname{Im} z \leq b\}$ does not meet $\tilde{\sigma}(\Gamma')$. For this x and b , we see that $g(z) \geq \tilde{h}(z)$ everywhere on the boundary of the set $R_{x,b} \cap \tilde{\sigma}(V)$, so that $g(z) \geq \tilde{h}(z)$ in this set, in particular, for $z = \tilde{\Phi}_n(\tilde{\sigma}(z_0)) = n + \alpha + i\beta$, n large. (To justify this precisely, we should replace $R_{x,b}$ by the finite rectangle $\{z | x < \operatorname{Re} z < x', 0 < \operatorname{Im} z < b\}$ and take the limit as $x' \rightarrow \infty$. We omit these details.)

Therefore, we have

$$\lim_{n \rightarrow \infty} h(\varphi_n(z_0)) = \lim_{n \rightarrow \infty} \tilde{h}(\tilde{\Phi}_n(\tilde{\sigma}(z_0))) \leq \lim_{n \rightarrow \infty} g(n + \alpha + i\beta) = \beta b^{-1}.$$

Since b was arbitrary, we see that $\lim_{n \rightarrow \infty} h(\varphi_n(z_0)) = 0$.

By an analogous argument for the lower half-plane, we see that if

$$\Phi(z) = [(1 + 2i)z - 1][z - 1 + 2i]^{-1}$$

that the sequence $\varphi_n(z_0)$ converges to a tangent to the arc of ∂D on the left of a .

□

The converse of this theorem is false, as may be seen by examining the map φ defined on D by $\varphi(z) = \sigma^{-1}(\sigma(z) + 1)$ where σ is a conformal map of D onto $\{z | \operatorname{Re} z > 0 \text{ and } -\operatorname{Re} z < \operatorname{Im} z < \sqrt{\operatorname{Re} z}\}$.

We next examine the important special case when φ is real valued on the real axis.

THEOREM 3.6. *Suppose φ is an analytic function that maps D into D and is real valued on the interval $(-1, 1)$. Suppose also that there is no fixed point of φ at which φ' vanishes. Then, for V , σ , and Φ as in Theorem 3.2, the set $V_{\mathbf{R}} = V \cap (-1, 1)$ is nonempty, and we may choose σ and Φ to be real valued on $V_{\mathbf{R}}$ and \mathbf{R} respectively.*

In particular, Case 4 of the remark above cannot occur. Since a function φ satisfying the hypotheses above satisfies either the hypotheses of Theorem 3.3 or those of Theorem 3.4, we may decide which of the remaining cases occurs by locating the Denjoy-Wolff point a and finding $\varphi'(a)$.

PROOF. Since $\varphi_n(0)$ is real for $n = 1, 2, 3, \dots$, the Denjoy-Wolff point $a = \lim_{n \rightarrow \infty} \varphi_n(0)$ is real, and if $a = \pm 1$ then the sequence $\varphi_n(0)$ converges to a nontangentially. Moreover, $\varphi_n(0)$ is in V for n sufficiently large, so $V_{\mathbf{R}} \neq \emptyset$.

Suppose that the function Φ constructed in Theorem 3.2 has been normalized as in the remark. In addition, in Cases 1 and 2, we suppose that for some particular x in V , $x \neq a$, $\sigma(x)$ is real. (We note that in Case 1, by Theorem 3.3, $\Phi'(0) = \varphi'(a)$ which is real since $-1 < a < 1$.) Clearly \bar{V} (i.e. the complex conjugate of V) is also a fundamental set for φ on which φ is univalent, and we define $\tilde{\sigma}$ on \bar{V} by $\tilde{\sigma}(z) = \overline{\sigma(\bar{z})}$.

Now in Cases 1, 2, or 3 of the remark, since φ and Φ are real valued on the real axis, we have $\tilde{\sigma}(\varphi(z)) = \overline{\sigma(\overline{\varphi(z)})} = \overline{\sigma(\varphi(\bar{z}))} = \overline{\Phi(\sigma(\bar{z}))} = \Phi(\sigma(\bar{z})) = \Phi(\tilde{\sigma}(z))$. That is both σ , V , and Φ and $\tilde{\sigma}$, \bar{V} and Φ are as in the conclusion of Theorem 3.2. By the uniqueness property, this means there is a linear fractional transformation τ mapping Ω onto itself such that $\Phi = \tau \circ \Phi \circ \tau^{-1}$ and $\tilde{\sigma} = \tau \circ \sigma$. From the relation $\Phi = \tau \circ \Phi \circ \tau^{-1}$ we see that $\tau(z) = \zeta z$ in Case 1, $\tau(z) = z + \zeta$ in Case 2, or

$\tau(z) = [(1 + \zeta)z + (1 - \zeta)][(1 - \zeta)z + (1 + \zeta)]^{-1}$ in Case 3, where ζ is an appropriate constant. From the equation $\tilde{\sigma} = \tau \circ \sigma$ (and, in Cases 1 and 2, from $\sigma(x) = \tilde{\sigma}(x)$), we see that, in fact, $\tau(z) = z$. Therefore, $\overline{\sigma(\bar{z})} = \sigma(z)$ and σ is real valued on $V_{\mathbb{R}}$.

In Case 4 where, say $\Phi(z) = [(1 + 2i)z - 1][z - 1 + 2i]^{-1}$, we have

$$\tilde{\sigma}(\varphi(z)) = \overline{\Phi(\sigma(\bar{z}))} = \tilde{\Phi}(\overline{\sigma(\bar{z})}) = \tilde{\Phi}(\tilde{\sigma}(z))$$

where $\tilde{\Phi}(z) = [(1 - 2i)z - 1][z - 1 - 2i]^{-1}$. That is, both σ , V , and Φ and $\tilde{\sigma}$, \bar{V} , and $\tilde{\Phi}$ are as in the conclusion of Theorem 3.2. By the uniqueness property, this means there is a linear fractional transformation mapping D onto D such that $\tilde{\Phi} = \tau \circ \Phi \circ \tau^{-1}$. But there is no such τ : the τ 's satisfying this equation are $\tau(z) = [-\zeta z + 1 + \zeta][(1 - \zeta)z + \zeta]^{-1}$, but none of these map D onto D . We have the same conclusion when Φ has the negative sign in Case 4. This contradiction means that, under our hypotheses, Case 4 cannot occur. \square

We conclude this section by proving two specialized theorems on the nature of σ . The proofs illustrate techniques that can be useful in studying σ for φ more general than stated in the hypotheses.

THEOREM 3.7. *Suppose φ is a finite Blaschke product with Denjoy-Wolff point a such that $|a| = 1$ and $\varphi'(a) < 1$. Then σ is an inner function that is continuous near a .*

PROOF. Let $s = \varphi'(a)$. By Theorem 3.4 we may take $\Omega = D$ and

$$\Phi(z) = [(1 + s)z + (1 - s)][(1 - s)z + (1 + s)]^{-1}.$$

Choose $\delta > 0$ small enough that $|\varphi'(z) - s| < \min\{s/2, (1 - s)/2\}$ if $|z - a| < \delta$. We take V , in Theorem 3.2, to be

$$V = \{z \in D \mid |z - a| < \delta\}$$

and let $I = \{z \mid |z| = 1, |z - a| < \delta\}$ and $\tilde{V} = V \cup I$. Thus φ is a univalent map of I into itself and $\bigcap_{n=0}^{\infty} \varphi_n(I) = \{a\}$.

Since σ maps D into D , $\lim_{r \rightarrow 1^-} \sigma(re^{i\theta})$ exists for almost all θ , and we write $\sigma(e^{i\theta})$ for this limit when it exists.

Now suppose $e^{i\theta}$ is in I and $\sigma(e^{i\theta}) = \beta$. Since σ is univalent on V and $e^{i\theta} \notin V$, $\beta \notin \sigma(V)$. It follows, from the continuity of φ and Φ , that $\sigma(\varphi_n(e^{i\theta})) = \Phi_n(\sigma(e^{i\theta})) = \Phi_n(\beta)$ and since $\varphi_n(e^{i\theta}) \notin V$, that $\Phi_n(\beta) \notin \sigma(V)$, for $n = 1, 2, 3, \dots$. But $\sigma(V)$ is a fundamental set for Φ on D , so $\Phi_n(\beta) \notin \sigma(V)$ for $n = 1, 2, 3, \dots$ implies $|\Phi_n(\beta)| = 1$ for all n .

Thus σ is a univalent map of V into D and $\sigma(I)$ is a subset of the unit circle. It follows from standard theorems in conformal mapping that σ is continuous on I .

Let $W = \bigcup_{n=0}^{\infty} (\varphi_n)^{-1}(\bar{V}) = \bigcup_{n=0}^{\infty} (\varphi_n)^{-1}(\tilde{V})$ (where $f^{-1}(A) = \{x \mid f(x) \in A\}$), so W is open in \bar{D} and $W \supset D$. For $e^{i\theta}$ in W , say $\varphi_k(e^{i\theta})$ is in I for some $k = 0, 1, 2, \dots$, we have $\sigma(e^{i\theta}) = \Phi_k^{-1}(\sigma(\varphi_k(e^{i\theta})))$ so $|\sigma(e^{i\theta})| = 1$. On the other hand, if $e^{i\theta}$ is not in W , we have $\varphi_n(e^{i\theta}) \notin \bar{V}$ for every $n, n = 0, 1, 2, \dots$. Given n , for f near 1 we have $\varphi(re^{i\theta}) \notin \bar{V}$ so that $\sigma(re^{i\theta}) \notin \Phi_n^{-1}(\sigma(\bar{V}))$. Since $\bigcup_{n=1}^{\infty} \Phi_n^{-1}(\sigma(\bar{V})) = \bar{D} \setminus \{-1\}$, this shows $\sigma(e^{i\theta}) = -1$ for $e^{i\theta} \notin W$.

Thus, we have shown $|\lim_{r \rightarrow 1^-} \sigma(re^{i\theta})| = 1$ for every θ and σ is continuous near a . \square

From the above we see that $\bar{D} \setminus W$ has measure zero. It seems likely that $\bar{D} \setminus W = \{e^{i\theta} | \varphi_{n+1}(e^{i\theta}) = \varphi_n(e^{i\theta}) \neq a \text{ for some } n = 0, 1, 2, \dots\}$, but I found no easy proof of this.

THEOREM 3.8. *Suppose φ is an analytic mapping of D into itself, with no fixed points in D . Let Φ , σ , and Ω be as in 3.2 and the subsequent remark, let $\alpha = \lim_{n \rightarrow \infty} \Phi_n(0)$, and let b be a fixed point of φ , different from the Denjoy-Wolff point a of φ , such that φ' has a continuous (and bounded) extension to $\{b\} \cup D$. Let K be a compact subset of \bar{D} , $b \in K$, such that $K \setminus \{a, b\} \subset D$ and $\varphi(K) \subset K$. If $\{z_j\}_{j=1}^\infty \subset K \cap D$ and $\lim_{j \rightarrow \infty} z_j = b$ then $\lim_{j \rightarrow \infty} \sigma(z_j) = \alpha$.*

PROOF. By Lemma 2.4, $q = \varphi'(b) > 1$. Choose $\delta > 0$ so that $|\varphi'(z)| < q + 1$ and $|\varphi(z) - a| > \delta$ for $|z - b| < \delta$. Let $L = \{z | (q + 1)^{-1}\delta \leq |z - b| \leq \delta\} \cap K$. From the hypothesis on K , we see that L is a compact subset of D .

Now if $|z - b| < (q + 1)^{-1}\delta$, then $|\varphi(z) - b| = |(z - b) \int_0^1 \varphi'(tz + (1 - t)b) dt| \leq |z - b|(q + 1) < \delta$. Thus if z is in K and $|z - b| < (q + 1)^{-1}\delta$ then either $\varphi(z) \in L$ or $|\varphi(z) - b| < (q + 1)^{-1}\delta$. For each z in K , $|z - b| < \delta$, since $\varphi_n(z) \rightarrow a$, we see that there is an integer k , $k = 0, 1, 2, \dots$, so that $\varphi_k(z) \in L$. Moreover, if $|z - b| < (q + 1)^{-N}\delta$, then $\varphi_n(z) \notin L$ for $n < N$. Let $L' = \sigma(L)$, so L' is a compact subset of Ω and the sequence Φ_n^{-1} converges uniformly to α on L' . If $\varphi_n(z) \in L$ then $\Phi_n(\sigma(z)) = \sigma(\varphi_n(z)) \in L'$ so $\sigma(z) \in \Phi_n^{-1}(L')$.

Combining the above, we see that $\lim_{j \rightarrow \infty} \sigma(z_j) = \alpha$; indeed if $|z_j - b| < (q + 1)^{-N}\delta$ then $\sigma(z_j) \in \cup_{n=N}^\infty \Phi_n^{-1}(L')$. \square

COROLLARY. *Suppose φ is an analytic mapping of the disk D into itself and is real valued on the interval $(-1, 1)$. Suppose the Denjoy-Wolff point of φ is 1 and that $\lim_{x \rightarrow -1} \varphi(x) = -1$ and $\lim_{x \rightarrow -1} \varphi'(x) = q < \infty$. Then $\lim_{x \rightarrow -1} \sigma(x) = \alpha$, where in each case $-1 < x < 1$.*

PROOF. In the proof of the theorem above, take $K = [-1, 1]$. Although φ' need not be continuous in $\{-1\} \cup D$, φ' is continuous on K . Since K is convex, the integral estimate still holds and the proof proceeds as above. \square

If Φ has two fixed points, the conclusion of the theorem can be paraphrased: $\sigma(a)$ is one fixed point of Φ , $\sigma(b)$ is the other. If Φ has only one fixed point, the proof of the theorem yields the more subtle conclusion that $\{\sigma(z_j)\}$ converges to the fixed point of Φ in the same way as $\{\Phi_n^{-1}(0)\}$ does.

The hypotheses of Theorem 3.8 may strike the reader as unduly complex, and no doubt this is not the best possible theorem along these general lines. However, some hypothesis regarding the set from which we choose the sequence z_k is essential. Consider the Blaschke product $\varphi(z) = (3z + 2)^3(2z + 3)^{-3}$. The Denjoy-Wolff point of φ is 1 and the other fixed points are -1 and $(-9 \pm 5i\sqrt{7})/16$ (all on the unit circle) and $\varphi'(1) = 3/5 < 1$. By Theorem 3.4,

$$\Phi(z) = (4z + 1)(z + 4)^{-1}$$

and $\Omega = D$ is a possible choice. By the corollary above, $\lim_{x \rightarrow -1} \sigma(x) = -1$. On the other hand there is a sequence of points λ_k , $|\lambda_k| = 1$, $\lim_{k \rightarrow \infty} \lambda_k = -1$ such that $\varphi_k(\lambda_k) = 1$ for each k , so that $\lim_{r \rightarrow 1^-} \sigma(r\lambda_k) = 1$ for each k . Clearly, the behavior of the inner function σ is very bad near -1 .

4. The solution of Schroeder’s and Abel’s equations. In the previous section we have seen how to associate a mapping φ of the disk into itself with a linear fractional transformation, Φ . We begin by showing how this produces a one-to-one correspondence between solutions of the functional equations for φ and Φ . Since the functional equations are easily solved for linear fractional transformations, we obtain information about the solutions for more general functions.

LEMMA 4.1. *Suppose $\varphi, V, \Phi, \sigma,$ and Ω are as in Theorem 3.2 and suppose $g(z) = \lambda z + \mu,$ where λ and μ are complex numbers, $\lambda \neq 0.$ Then the correspondence $F \leftrightarrow f = F \circ \sigma$ is a one-to-one correspondence between solutions F of $F \circ \Phi = g \circ F$ on Ω and solutions f of $f \circ \varphi = g \circ f$ on $D.$*

PROOF. If F satisfies $F \circ \Phi = g \circ F$ then letting $f = F \circ \sigma$ we have $f \circ \varphi = F \circ \sigma \circ \varphi = F \circ \Phi \circ \sigma = g \circ F \circ \sigma = g \circ f.$

Conversely, if f is analytic and satisfies $f \circ \varphi = g \circ f$ on $D,$ since σ is univalent on $V,$ we may define \tilde{F} on $\sigma(V)$ by letting $\tilde{F} = f \circ \sigma^{-1}$ so that $\tilde{F} \circ \Phi = g \circ \tilde{F}$ on $\sigma(V).$ Since $\sigma(V)$ is a fundamental set for Φ on $\Omega,$ we may define F on Ω by $F(z) = g_{-n}(\tilde{F}(\Phi_n(z))),$ where n is a positive integer large enough that $\Phi_n(z)$ is in $\sigma(V)$ and g_{-n} is the n th iterate of $g^{-1}.$ The function F is well defined because if $\Phi_m(z)$ and $\Phi_n(z)$ are both in $\sigma(V)$ and $m = n + k,$ for k positive, then, from the equation $\tilde{F} \circ \Phi = g \circ \tilde{F},$ we have $g_{-m}(F(\Phi_m(z))) = g_{-n}(g_{-k}(\tilde{F}(\Phi_k(\Phi_n(z))))) = g_{-n}(\tilde{F}(\Phi_n(z))).$ We see that F is analytic and satisfies the equation $F \circ \Phi = g \circ F$ on Ω and that $F \circ \sigma = f.$ \square

The reader should note that no matter what the form of $g,$ if $F \circ \Phi = g \circ F$ then $f = F \circ \sigma$ is a solution of $f \circ \varphi = g \circ f,$ and that this is a one-to-one correspondence between solutions on $\sigma(V)$ and V respectively.

We now give complete, explicit solutions of Schroeder’s equation for the linear fractional transformations that concern us. The following four propositions correspond to the four cases in the remark following Theorem 3.2. In addition, for completeness, we include in Proposition 4.2 the solution for a rotation, i.e. $\Phi(z) = sz,$ where $|s| = 1.$ (Schroeder’s equation is the case $\mu = 0$ in the functional equation of Lemma 4.1.)

PROPOSITION 4.2. *If $\Phi(z) = sz, 0 < |s| \leq 1,$ then $F \circ \Phi = \lambda F$ has a nonzero solution F analytic in \mathbb{C} (or D) if and only if $\lambda = s^n$ for $n = 0, 1, 2, \dots.$ If s is not a root of unity and $\lambda = s^n$ for some nonnegative integer $n,$ then $F(z) = cz^n$ for some constant $c.$ If s is a primitive k th root of unity and $\lambda = s^n$ for $n = 0, 1, 2, \dots, k - 1$ then $F(z) = z^n g(z^k)$ where g is analytic in \mathbb{C} (or D).*

PROOF. Compare the Taylor series of $F(sz)$ and $\lambda F(z).$ \square

PROPOSITION 4.3. *If $\Phi(z) = z + 1,$ the function $F,$ analytic in the entire plane, is a solution of $F \circ \Phi = \lambda F$ ($\lambda \neq 0$) if and only if $F(z) = e^{\alpha z} g(e^{2\pi iz})$ where $e^\alpha = \lambda$ and g is analytic in the punctured plane $\{w: 0 < |w| < \infty\}.$*

PROOF. Clearly a function of the form given satisfies the functional equation. Conversely, if F is entire and $F \circ \Phi = \lambda F$ ($\lambda \neq 0$), choose α such that $e^\alpha = \lambda$ and define g on $\{w: 0 < |w| < \infty\}$ by $g(w) = e^{-\alpha z} F(z)$ where $w = e^{2\pi iz}.$ The function g

is well defined and analytic because F satisfies the functional equation and is analytic. \square

PROPOSITION 4.4. *Suppose $\Phi(z) = [(1 + s)z + (1 - s)][(1 - s)z + (1 + s)]^{-1}$ where $0 < s < 1$. Let $\delta = (\text{Log } s)^{-1}$ and $\zeta(z) = \delta \text{Log}[(1 - z)(1 + z)^{-1}]$. The function F , analytic in D , is a solution of $F \circ \Phi = \lambda F$ ($\lambda \neq 0$) if and only if $F(z) = e^{\alpha\zeta(z)}g(e^{2\pi i\zeta(z)})$ where $e^\alpha = \lambda$ and g is analytic in the annulus $\{w: |\text{Log } w| < \pi^2|\text{Log } s|^{-1}\}$.*

PROOF. Observe that ζ maps D onto the strip $\{w: |\text{Im } w| < (\pi/2)|\text{Log } s|^{-1}\}$ and that $\zeta \circ \Phi = \zeta + 1$. We now repeat the proof of 4.2. \square

PROPOSITION 4.5. *Suppose $\Phi(z) = [(1 \pm 2i)z - 1][z - 1 \pm 2i]^{-1}$. Let $\zeta(z) = i(1 + z)(1 - z)^{-1}$. The function F , analytic in D , is a solution of the equation $F \circ \Phi = \lambda F$ ($\lambda \neq 0$) if and only if $F(z) = e^{\mp \alpha\zeta(z)}g(e^{2\pi i\zeta(z)})$ where $e^\alpha = \lambda$ and g is analytic on the punctured disk $\{w: 0 < |w| < 1\}$.*

PROOF. Observe that ζ maps D onto the upper half-plane and that $\zeta \circ \Phi = \zeta \mp 1$. We now repeat the proof of 4.3. \square

We can now give a proof of Königs' 1884 theorem on solutions of the functional equation $f \circ \varphi = \lambda f$ when φ has a fixed point in D , and describe the situation when φ has no fixed points in D .

THEOREM 4.6 [14]. *Suppose φ is a nonconstant analytic function mapping D into D with fixed point a in D (but φ not a Möbius transformation mapping D onto D).*

(a) *If $\varphi'(a) = 0$, then the only nonzero solution of $f \circ \varphi = \lambda f$ is $\lambda = 1$ and f constant.*

(b) *If $\varphi'(a) \neq 0$, then $f \circ \varphi = \lambda f$ has nonzero solutions if and only if $\lambda = \varphi'(a)^n$, $n = 0, 1, 2, \dots$. When $\lambda = \varphi'(a)^n$, every solution is a multiple of the solution f whose values are $f(z) = \sigma(z)^n$, where σ and Φ are as in Theorem 3.2 and Case 1 of the remark.*

PROOF. (a) Clearly $\lambda = 1$, f constant is a solution of the functional equation. Since $f \equiv 0$ is the only solution when $\lambda = 0$, we assume $\lambda \neq 0$.

Suppose $f \circ \varphi = \lambda f$. Because a is a fixed point of φ , we have $f(a) = f(\varphi(a)) = \lambda f(a)$, so either $f(a) = 0$ or $\lambda = 1$. In any case, we see that $f - f(a)$ is also a solution of the functional equation so we may assume $f(a) = 0$.

Suppose $f^{(j)}(a) = 0$ for $j = 0, 1, 2, \dots, k - 1$. Taking k th derivatives on both sides of our equation yields $(f^{(k)} \circ \varphi)(\varphi')^k + (\text{terms involving } f^{(j)}, j < k) = \lambda f^{(k)}$. Hence $\lambda f^{(k)}(a) = f^{(k)}(a)(\varphi'(a))^k + 0 = 0$ and $f^{(k)}(a) = 0$ as well.

Therefore the only solutions of $f \circ \varphi = \lambda f$ are $\lambda = 1$ with f constant and $f \equiv 0$.

(b) Since φ' does not vanish at the Denjoy-Wolff point a , φ satisfies the hypotheses of Theorem 3.2. The function Φ has $\Phi(\sigma(a)) = \varphi'(a) = s$, $0 < |s| < 1$, and we may assume that $\Phi(z) = sz$. The conclusion follows from Proposition 4.2 and the correspondence between solutions of the functional equations for φ and Φ (Lemma 4.1). \square

THEOREM 4.7. *If φ is an analytic function mapping D into D with no fixed points in D , then the solution space of $f \circ \varphi = \lambda f$ is infinite dimensional for every $\lambda \neq 0$.*

PROOF. The Φ and Ω of Theorem 3.2 cannot be sz and the plane since φ has no fixed points (Theorem 3.3). The conclusion follows from Propositions 4.3, 4.4, or 4.5 and the correspondence between solutions of the functional equations for φ and Φ . \square

Of course, the information we have developed in this and the preceding section is more complete than stated in the above result. Given a particular function φ , one would want to identify Φ and σ as precisely as possible and examine the nature of the solutions.

Abel's equation is the case $\lambda = 1, \mu = 1$ in the functional equation of Lemma 4.1. We note that if A_1 and A_2 are two solutions of the equation $A \circ \varphi = A + 1$ then $A_1 - A_2$ is a solution of Schroeder's equation with $\lambda = 1$. Thus to give a complete solution to Abel's equation we may give one solution and all solutions to Schroeder's equation with $\lambda = 1$. In the proofs of 4.4 and 4.5 we gave one solution to Abel's equation for these linear fractional transformations, so we obtain one solution for a function φ that has no fixed points in D . We record the following consequence of these observations and Theorem 4.7.

THEOREM 4.8. *If φ has no fixed points in D , then the set of solutions of $A \circ \varphi = A + 1$, where A is analytic in D , is an infinite dimensional hyperplane in the vector space of functions analytic in D .*

If φ has a fixed point a in D , there can be no solutions of $A \circ \varphi = A + 1$ that are analytic in D since $A(a) = \infty$ ($A(a) = A(\varphi(a)) = A(a) + 1$). However, if $\varphi'(a) = s \neq 0$ and σ is as in Case 1 of the remark following Theorem 3.2, the function $A(z) = (\log s)^{-1} \log|\sigma(z)|$ is a harmonic solution of Abel's equation in $D \setminus \{a\}$.

5. Fractional iteration. It is clear that the iterates $\varphi_n, n = 0, 1, 2, \dots$, form a discrete semigroup of functions. It is natural to ask whether this semigroup can be embedded in a continuous (analytic?) semigroup of analytic functions, that is, whether we can define fractional iterates of φ . Depending on what one requires of the semigroup functions, the answer is known to be no in some cases (e.g. Baker [2] or Karlin and McGregor [13]) yes in others (e.g. Szekeres [22]). Using the linear fractional transformation Φ , we construct the "tail" of a semigroup and what deserves to be called the infinitesimal generator of the semigroup interpolating $\{\varphi_n\}$, but we find that the existence of a genuine semigroup depends on the geometry of V and $\sigma(V)$. Berkson and Porta [3] give a detailed discussion of the infinitesimal generator in the semigroup case.

We begin by noting real analytic groups containing the linear fractional transformations Φ . That is, we define, for z in Ω and $-\infty < t < \infty, H(z, t)$ in Ω so that H is complex analytic in the first argument, real analytic in the second and satisfies the identities $H(H(z, t_1), t_2) = H(z, t_1 + t_2)$ and $H(z, 1) = \Phi(z)$. The infinitesimal generator G , of the group is $G(z) = \partial H / \partial t|_{t=0}$. For $0 < |s| < 1$, we write s^t for $e^{t \log s}$. If $\Phi(z) = sz$ for z in \mathbb{C} , where $0 < |s| < 1$, let $H(z, t) = s^t z$, so $G(z) = (\log s)z$. If $\Phi(z) = z + 1$ for z in \mathbb{C} , let $H(z, t) = z + t$, so $G(z) = 1$. If $\Phi(z) = [(1 + s)z + (1 - s)] / [(1 - s)z + (1 + s)]^{-1}$ for z in D , where $0 < s < 1$, let

$$H(z, t) = [(1 + s^t)z + (1 - s^t)] / [(1 - s^t)z + (1 + s^t)]^{-1},$$

so $G(z) = (\frac{1}{2} \log s)(z^2 - 1)$. If $\Phi(z) = [(1 \pm 2i)z - 1][z - 1 \pm 2i]^{-1}$ for z in D , let $H(z, t) = [(t \pm 2i)z - t][tz - t \pm 2i]^{-1}$, so $G(z) = \pm (i/2)(1 - z)^2$. It is not hard to see that these are the only semigroups of functions mapping Ω into Ω and continuous in the sense that $\Phi_t \rightarrow \Phi_{t_0}$ uniformly on compact subsets of Ω .

If φ and V are as in Theorem 3.2, and z is in D , let $\nu(z) = \min\{n: \varphi_n(z) \in V\}$. If Φ is the linear fractional transformation constructed from φ in Theorem 3.2, normalized as in the subsequent remark, and $H(z, t)$ is the group of linear fractional transformations defined above for Φ , we define

$$\tau(z) = \inf\{t: \nu(z) \leq t \text{ and } H(\sigma(z), t_1) \in \sigma(V) \text{ for all } t_1 > t\}.$$

To see that $\tau(z)$ exists we note that $\{H(\sigma(z), t): 0 \leq t \leq 1\}$ is a compact subset of Ω , and since $\sigma(V)$ is a fundamental set for Φ in Ω , there is a positive integer N so that Φ_n maps this set into $\sigma(V)$ for $n > N$. But $\Phi_n(H(\sigma(z), t)) = H(\sigma(z), t + n)$ so the set in the definition of $\tau(z)$ is nonempty.

THEOREM 5.1. *Let φ satisfy the hypotheses of Theorem 3.2 and let τ be defined as above. There is a function $h = h(z, t)$ defined for z in D and $t > \tau(z)$, complex analytic in the first argument, real analytic in the second such that*

$$h(h(z, t_1), t_2) = h(z, t_1 + t_2) \text{ for } t_1 > \tau(z) \text{ and } t_2 > 0$$

and such that $h(z, n) = \varphi_n(z)$ when n is an integer, $n > \tau(z)$. Moreover, there is a function g , meromorphic in D , holomorphic in V , which agrees with the infinitesimal generator of the semigroup $h(z, t)$ on the set $\{z: \tau(z) = 0\}$.

PROOF. Let Φ , σ , V , and Ω be as in Theorem 3.2 and the subsequent remark and let $H(z, t)$ the group for Φ as defined above. Define $h(z, t)$ for $t > \tau(z)$ by $h(z, t) = \sigma_*^{-1}(H(\sigma(z), t))$ where σ_*^{-1} is the inverse of the restriction of σ to V . Since $t > \tau(z)$, $H(\sigma(z), t)$ is in $\sigma(V)$ and σ_*^{-1} is defined there.

Now, if $t_1 > \tau(z)$ and $t_2 > 0$, we have

$$\begin{aligned} h(h(z, t_1), t_2) &= \sigma_*^{-1}\left(H\left(\sigma\left(\sigma_*^{-1}(H(\sigma(z), t_1))\right), t_2\right)\right) = \sigma_*^{-1}(H(H(\sigma(z), t_1), t_2)) \\ &= \sigma_*^{-1}(H(\sigma(z), t_1 + t_2)) = h(z, t_1 + t_2). \end{aligned}$$

We also have, for $n > \tau(z)$, $h(z, n) = \sigma_*^{-1}(H(\sigma(z), n)) = \sigma_*^{-1}(\Phi_n(\sigma(z))) = \sigma_*^{-1}(\sigma(\varphi_n(z))) = \varphi_n(z)$ where the last equality holds since $n > \tau(z) \geq \nu(z)$ and $\varphi_n(z)$ is in V .

The analyticity assertions follow from the analyticity of σ_*^{-1} and σ and the corresponding facts for H .

On the set $\{z: \tau(z) = 0\}$ the functions $h(z, t)$, $t \geq 0$, form a real analytic semigroup. A simple calculation shows that the infinitesimal generator of this semigroup is $g(z) = G(\sigma(z))(\sigma'(z))^{-1}$, where $\tau(z) = 0$. This formula defines a meromorphic function in D that is holomorphic in V , as asserted. \square

COROLLARY. *Suppose φ is an analytic mapping of D into D and is real valued on the interval $(-1, 1)$. Let a denote the Denjoy-Wolff point of φ and let x_1 and x_2 be real numbers $-1 \leq x_1 \leq a \leq x_2 \leq 1$. If $\varphi'(x) > 0$ on the interval (x_1, x_2) there is a function $h(x, t)$, real analytic in each argument, defined for $x_1 < x < x_2$ and $t \geq 0$ such that $h(h(x, t_1), t_2) = h(x, t_1 + t_2)$ for $t_1, t_2 \geq 0$ and $h(x, 1) = \varphi(x)$.*

(This is similar to Theorem 1a of Szekeres [22] for nonanalytic functions.)

PROOF. Since $\varphi' > 0$, φ is univalent on (x_1, x_2) and $(\varphi_k)^{-1}$ is well defined as a function from $(\varphi_k(x_1), \varphi_k(x_2))$ onto (x_1, x_2) for any positive integer k . Given x , $x_1 < x < x_2$, choose an integer n so that $n \geq \tau(x)$. In the theorem above, we defined $h(x, t)$ for $t \geq n$ and for such t we see that $h(x, t)$ is in the interval with end points a and $\varphi_n(x)$. We extend the definition of h , by letting

$$h(x, t) = (\varphi_n)^{-1}(h(x, t + n)) \quad \text{for } t \geq 0. \quad \square$$

If $\varphi(z) = \sum_{k=0}^{\infty} a_k z^k$ is a probability generating function with $a_0 + a_1 > 0$, we may apply the corollary for the interval $(0, 1)$. Of course, $h(x, t)$ need not be a probability generating function if t is not an integer. (Cf. Karlin and McGregor [12].)

In order for the construction in Theorem 5.1 to give a genuine semigroup, φ must be univalent, that is $V = D$ must be possible. This is far from being sufficient to guarantee that, for each $t > 0$, the function $h(z, t)$ is defined for all z in D and maps D into itself. If φ is univalent, but $\sigma(D)$ is not mapped into itself by $H(z, t)$ for $t > 0$ then the functions $h(z, t)$ may be definable on all of D for each t (if σ^{-1} can be extended to $\cup_{t>0} H(\sigma(D), t)$) but fail to map D into D . If in addition $\sigma(D)$ has corners in its boundary, then $h(z, t)$ will not be defined for some z and t . I know of no reasonable criteria for φ that will guarantee that $\{\varphi_n\}$ can be embedded in a real analytic semigroup of analytic functions mapping D into D .

On the other hand, the following theorem shows that the semigroups discussed above are the only ones; in particular it follows that such semigroups are real analytic in t . Moreover, we show that the fixed points of φ_t are the same for all $t > 0$, where we say $e^{i\theta}$ is a fixed point of φ if $\lim_{r \rightarrow 1^-} \varphi(re^{i\theta}) = e^{i\theta}$. (It need not be the case for arbitrary φ that the fixed points of φ_n are all the same. Results of this type are known for φ with fixed point in D , for example see [19].)

THEOREM 5.2. *Let φ, Φ , and σ be as in Theorem 3.2 and suppose $\varphi_t(z) = h(z, t)$ is a one parameter semigroup, $t \geq 0$, z in D , that is continuous in the sense that $\varphi_t \rightarrow \varphi_{t_0}$ uniformly on compact sets as $t \rightarrow t_0$. Then φ is univalent in D , the fixed points of φ_t ($t > 0$) are the same as those of φ , and $h(z, t) = \sigma^{-1}(H(\sigma(z), t))$ where H is a semigroup associated with Φ .*

PROOF. If φ^*, Φ^* , and σ^* are as in Theorem 3.2, then we have $\Phi_2^* \circ \sigma^* = \sigma^* \circ \varphi_2^*$. If we also have $\Psi \circ \nu = \nu \circ \varphi_2^*$ then the uniqueness assertion says Φ_2^* is conjugate to Ψ . If φ_t is a semigroup as above, this remark shows that for t a dyadic rational number, we have $\Phi_t \circ \sigma = \sigma \circ \varphi_t$, so by continuity this holds for all t .

Since $\Phi'(\sigma(z_0))\sigma'(z_0) = \sigma'(\varphi(z_0))\varphi'(z_0)$, $\varphi'(z_0) = 0$ implies $\sigma'(z_0) = 0$. Conversely $\sigma'(z_0) = 0$ implies $0 = \Phi'_t(\sigma(z_0))\sigma'(z_0) = \sigma'(\varphi_t(z_0))\varphi'_t(z_0)$ and since σ' is not identically 0 (and $\sigma'(a) \neq 0$) $\varphi'_t(z_0) = 0$ for all t . Since $\varphi_t \rightarrow z$ uniformly on compact sets as $t \rightarrow 0^+$, we conclude φ' and σ' never vanish on D .

This means σ (and therefore φ) is univalent in D . Indeed, suppose $\sigma(z_0) = \sigma(w_0)$ where z_0 and w_0 are distinct points of D . Denoting $\varphi_t(z_0)$ by z_t and $\varphi_t(w_0)$ by w_t , we have $\sigma(z_t) = \sigma(\varphi_t(z_0)) = \Phi_t(\sigma(z_0)) = \Phi_t(\sigma(w_0)) = \sigma(w_t)$ for all $t \geq 0$. Now for t

large enough, w_t and z_t are in a fundamental set for φ , so for these t , $\sigma(w_t) = \sigma(z_t)$ means $w_t = z_t$, that is $\varphi_t(z_0) = \varphi_t(w_0)$. Letting $t_0 = \inf\{t: \varphi_t(w_0) = \varphi_t(z_0)\}$, we observe that, by continuity, $\varphi_{t_0}(w_0) = \varphi_{t_0}(z_0)$ and $t_0 > 0$ since $w_0 \neq z_0$. As $t \rightarrow t_0^-$, we have $w_t \rightarrow w_{t_0} = z_{t_0}$ and $z_t \rightarrow z_{t_0}$ but $w_t \neq z_t$. Since $\sigma(w_t) = \sigma(z_t)$ for all t , this means $\sigma'(z_{t_0}) = 0$ which contradicts the assertion above.

We are now ready to show that all the iterates have the same fixed points, that is, for s and t positive if $\lim_{r \rightarrow 1^-} \varphi_s(re^{i\theta}) = e^{i\theta}$ then $\lim_{r \rightarrow 1^-} \varphi_t(re^{i\theta}) = e^{i\theta}$. By suitably renaming, it is sufficient to show this for $s = 1$.

For simplicity of notation we will prove it for φ whose Denjoy-Wolff point is on ∂D , and, by using a suitable transformation, we assume $\Phi_t(z) = z + t$ on the whole plane or an appropriate strip or half-plane according as φ falls in Case 2, 3, or 4 of the remark after Theorem 3.2. The proof for φ having a fixed point in D is completely analogous: the spirals e^{bs} , s real, replace the lines $\text{Im } z = y_0$.

Let $\lim_{r \rightarrow 1^-} \varphi(re^{i\theta}) = e^{i\theta}$ and let $\varepsilon > 0$ be given. Find a continuous curve $\gamma: [0, 1] \rightarrow \bar{D}$ so that $|\gamma(0)| = |\gamma(1)| = 1$, $\varepsilon/2 < |\gamma(s) - e^{i\theta}| < \varepsilon$ for $0 < s < 1$, and $\lim_{s \rightarrow 0^+} \sigma(\gamma(s))$ and $\lim_{s \rightarrow 1^-} \sigma(\gamma(s))$ exist and are finite. (That such a curve exists follows from the Riesz theorem.) Let $X = \bigcap_{0 < r < 1} \{\sigma(se^{i\theta}): r < s < 1\}$. Since X is the nested intersection of nonempty, compact, connected subsets of the sphere, X is nonempty, compact and connected, and of course $X \subset \partial(\sigma(D))$.

Now we claim either $X = \{\infty\}$, or else $X = \{x + iy_0: x \geq x_0\}$ for some finite y_0 and $x_0 \geq -\infty$. We observe that $w \in X$ means $w + t \notin \sigma(D)$ for any t since $w + t \in \sigma(D)$ means $w + n \in \sigma(D)$ for any integer $n \geq t$ so that $\varphi_n(re^{i\theta}) \rightarrow \sigma^{-1}(w + n)$ which contradicts $\varphi_n(re^{i\theta}) \rightarrow e^{i\theta}$. If $x_1 + iy_1$ and $x_2 + iy_2$ are in X , $y_1 \neq y_2$, then there is x_3 so that $x_3 + i\frac{1}{2}(y_1 + y_2)$ is in X . Since $\sigma(D)$ is connected and $X \subset \partial(\sigma(D))$ this means that $x_4 + i\frac{1}{2}(y_1 + y_2)$ is in $\sigma(D)$ for some $x_4 > x_3$ which is impossible. Thus either $X = \{\infty\}$ or there is some y_0 so that $X \subset \{z: \text{Im } z = y_0\}$. Now if $x + iy_0 \in X$ implies $x < x_1$ for some finite x_1 , then $\{s + iy_0: s \geq x_1\}$ is a free boundary arc of $\sigma(D)$ which is mapped continuously by σ^{-1} onto an arc in ∂D . We can find a sequence r_k increasing to 1 and an integer n so that $\sigma(r_k e^{i\theta}) \rightarrow x + iy_0$ in X and $x + n > x_1$ so that $\lim_{k \rightarrow \infty} \varphi_n(r_k e^{i\theta}) = \sigma^{-1}(x + n + iy_0) \neq e^{i\theta}$ which is impossible. Thus if X contains any finite point, it must be $\{x + iy_0: x \geq x_0\}$. In this case we observe that $\{x + iy_0: x \in \mathbf{R}\}$ does not intersect $\sigma(D)$ and that X intersects the curve $\sigma(\gamma)$ at most in the single point $x_0 + iy_0$.

Now the curves $\sigma(\gamma)$ and $\sigma(re^{i\theta})$, $1 - \varepsilon/3 < r < 1$, are disjoint, so the arc $\sigma(re^{i\theta})$, $1 - \varepsilon/3 < r < 1$, lies in a single component of $\sigma(D) \setminus \sigma(\gamma)$, and X lies in the boundary of one of these.

If $X = \{\infty\}$ then $|\sigma(re^{i\theta})| \rightarrow \infty$ as $r \rightarrow 1^-$ and since $\sigma(\gamma)$ is bounded, for any fixed t , $\sigma(re^{i\theta})$ and $\sigma(re^{i\theta}) + t$ must be in the same component of $\sigma(D) \setminus \sigma(\gamma)$ for r close to 1.

If $X = \{x + iy_0: x \geq x_0\}$ then, for fixed t , $\sigma(re^{i\theta}) + t$, $1 - \varepsilon/3 < r < 1$, intersects $\sigma(\gamma)$ only finitely often (else $x_0 - t + iy_0 \in X \cap \sigma(\gamma)$). Since $\sigma(re^{i\theta}) + t$ contains points of arbitrarily large real part for r near 1 we again see that $\sigma(re^{i\theta})$ and $\sigma(re^{i\theta}) + t$ lie in the same component of $\sigma(D) \setminus \sigma(\gamma)$ for r near 1.

Thus for r close to 1, $re^{i\theta}$ and $\varphi_t(re^{i\theta})$ lie in the same component of $D \setminus \gamma$, in particular $|\varphi_t(re^{i\theta}) - e^{i\theta}| < \epsilon$. Since ϵ was arbitrary, this shows $\lim_{r \rightarrow 1^-} \varphi_t(re^{i\theta}) = e^{i\theta}$. \square

6. Questions and comments. One would like an effective technique for computing σ . If φ is analytic at the Denjoy-Wolff point a and $|\varphi'(a)| < 1$ (as must happen if a is in D) then, as is well known, one can compute, recursively, the coefficients in the Taylor expansion of σ around a . In case $|a| = 1$, this is not an entirely satisfactory solution because we do not want to assume φ is analytic at a and because we are possibly more interested in σ on D than in some neighborhood of a . Furthermore, when $\varphi'(a) = 1$, the technique of finding Taylor coefficients breaks down because there need not be a solution even in formal power series [18]. For example, if $\varphi(z) = \frac{1}{2} + \frac{1}{2}z^2$, then we have from Theorem 3.5 that we may take $\Omega = \mathbb{C}$ and $\Phi(z) = z + 1$. Changing variables to facilitate computation, we find that there is a function $\tilde{\sigma}$ analytic on $|z + 1| < 1$, $\tilde{\sigma}(0) = 0$, so that $\tilde{\Phi} \circ \tilde{\sigma} = \tilde{\sigma} \circ \tilde{\varphi}$ where $\tilde{\varphi}(z) = z + \frac{1}{2}z^2$ and $\tilde{\Phi}(z) = z(1 - z)^{-1}$. But elementary calculations show that $\tilde{\sigma}$ does not have a formal power series at 0. It would be desirable, but probably difficult, to find techniques to effectively compute σ .

As we have seen, the difference between Case 2 and Case 4 in the remark following Theorem 3.2 (which correspond to $\varphi'(a) = 1$) is rather subtle. Although Theorem 3.5 is helpful in this regard, it cannot be called an effective procedure to distinguish between the cases, because computing iterates is at best a tedious process. There is some evidence that if $a = 1$ and $\varphi'(1) = 1$ then $\text{Arg } \varphi''(1)$ can be used to distinguish the cases; namely $\varphi''(1)$ pure imaginary implies $\Omega = D$. This is the case when φ''' is continuous and $\varphi''(0) \neq 0$, but more work needs to be done along these lines.

Problem 1. How can one decide whether a given φ with $\varphi'(a) = 1$ gives rise to Case 2 or Case 4?

One way of looking at Theorem 3.2 is as a classification theorem: given φ mapping D into D , we (sometimes) associate with it a certain linear fractional transformation which tells us about φ . As noted above, linear fractional transformations cannot be used to classify functions that map D into D and have $\varphi(0) = \varphi'(0) = 0$. How can we classify these functions?

Problem 2. Find a set of functions \mathcal{F} so that for every φ mapping D into D with $\varphi(0) = \varphi'(0) = 0$, there is a unique Φ in \mathcal{F} so that $\Phi \circ \sigma = \sigma \circ \varphi$ where σ is analytic in D .

If the functions of the family \mathcal{F} are well understood, the resulting classification could be helpful in studying problems involving iteration of φ and related functional equations.

It is plausible that the techniques used here to study iteration of analytic functions on the disk could be applied more widely. The crucial facts used here seem to be that φ has a unique attractive fixed point a such that $\varphi_n(z) \rightarrow a$ for all z , that V is simply connected, and φ is univalent on V . If these conditions could be weakened, we might be able to use the techniques to study entire functions or functions on Riemann surfaces.

REFERENCES

1. K. B. Athreya and P. E. Ney, *Branching processes*, Springer-Verlag, Berlin, 1972.
2. I. N. Baker, *Fractional iteration near a fixpoint of multiplier 1*, J. Australian Math. Soc. **4** (1964), 143–148.
3. E. Berkson and H. Porta, *Semigroups of analytic functions and composition operators*, Michigan Math. J. **25** (1978), 101–115.
4. G. T. Cargo, *Fixed points and ideal fixed points of holomorphic functions*, Notices Amer. Math. Soc. **25** (1978), A636.
5. A. Denjoy, *Sur l'itération des fonctions analytiques*, C. R. Acad. Sci. Paris **182** (1926), 255–257.
6. J. Dugundji, *Topology*, Allyn and Bacon, Boston, Mass., 1966.
7. P. Fatou, *Sur les équations fonctionnelles*, Bull. Soc. Math. France **47** (1919), 161–271; **48** (1920), 33–94, 208–314.
8. T. E. Harris, *The theory of branching processes*, Springer-Verlag, Berlin, 1963.
9. H. Kamowitz, *The spectra of composition operators on H^p* , J. Functional Anal. **18** (1975), 132–150.
10. S. Karlin and J. McGregor, *Spectral theory of branching processes*, I, II, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **5** (1966), 6–33, 34–54.
11. _____, *On the spectral representation of branching processes with mean 1*, J. Math. Anal. Appl. **21** (1968), 485–495.
12. _____, *Embeddability of discrete time simple branching processes into continuous time processes*, Trans. Amer. Math. Soc. **132** (1968), 115–136.
13. _____, *Embedding iterates of analytic functions with two fixed points into continuous groups*, Trans. Amer. Math. Soc. **132** (1968), 137–145.
14. G. Koenigs, *Recherches sur les intégrales des certaines equations fonctionnelles*, Ann. École Norm. Sup. (3) **1** (1884), supplement, 3–41.
15. M. Kuczma, *On the Schroeder equation*, Roxprawy Mat. **34** (1963).
16. R. Nevanlinna, *Analytic functions*, Springer-Verlag, Berlin, 1970.
17. K. Noshiro, *On the theory of schlicht functions*, J. Fac. Sci. Hokkaido Univ. Jap. (1) **2** (1934–1935), 129–155.
18. R. C. Penney, conversation with the author, 1979.
19. Ch. Pommerenke, *Über die Subordination analytischer Functionen*, J. Reine Angew. Math. **218** (1965), 159–173.
20. _____, *On the iteration of analytic functions in a half plane. I*, J. London Mat. Soc. (2) **19** (1979), 439–447.
21. E. Schroeder, *Über itierte Funktionen*, Math. Ann. **3** (1871), 296–322.
22. G. Szekeres, *Regular iteration of real and complex functions*, Acta Math. **100** (1958), 203–258.
23. M. Tsuji, *Potential theory in modern function theory*, Maruzen, Tokyo, 1959.
24. S. E. Warschawski, *On the higher derivatives at the boundary in conformal mapping*, Trans. Amer. Math. Soc. **38** (1935), 310–340.
25. J. Wolff, *Sur l'itération des fonctions*, C. R. Acad. Sci. Paris **182** (1926), 42–43, 200–201.
26. _____, *L'intégrale d'une fonction holomorphe et a partie réelle positive dans un demi plan est univalente*, C. R. Acad. Sci. Paris **198** (1934), 1209–1210.

DEPARTMENT OF MATHEMATICS, PURDUE UNIVERSITY, WEST LAFAYETTE, INDIANA 47907