NORTH-HOLLAND
MATHEMATICS STUDIES

53

Notas de Matemática
editor: Leopoldo Nachbin

Nine Introductions in Complex Analysis

SANFORD L. SEGAL

NORTH-HOLLAND

NINE INTRODUCTIONS
IN COMPLEX ANALYSIS

NORTH-HOLLAND
MATHEMATICS STUDIES **53**

Notas de Matemática (80)

Editor: Leopoldo Nachbin

Universidade Federal do Rio de Janeiro
and University of Rochester

Nine Introductions in Complex Analysis

SANFORD L. SEGAL

Department of Mathematics
University of Rochester
Rochester, N.Y. 14620, U.S.A.

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM • NEW YORK • OXFORD

ISBN: 0 444 86226 9

Publishers:
NORTH-HOLLAND PUBLISHING COMPANY
AMSTERDAM • NEW YORK • OXFORD

Sole distributors for the U.S.A. and Canada:
ELSEVIER NORTH-HOLLAND, INC.
52 VANDERBILT AVENUE, NEW YORK, N.Y. 10017

Library of Congress Cataloging in Publication Data

Segal, Sanford L., 1937-
Nine introductions in complex analysis.

(North-Holland mathematics studies ; 53)
(Notas de matemática ; 80)
Bibliography: p.
Includes index.
1. Functions of complex variables. I. Title.
II. Series. III. Series: Notas de matemática
(North-Holland Publishing Company) ; 80.
QA1.N86 no. 80 [QA331] 510s [515.9] 81-9568
ISBN 0-444-86226-9 AACR2

PRINTED IN THE NETHERLANDS

FOR THE STUDENT´S WHO I HOPE MAY PROFIT FROM IT

ACKNOWLEDGEMENTS

I wish to thank Marion Lind and Roberta Colon for the preparation of this manuscript. To Marion Lind I am especially grateful for her diligence over many many months, and her typing and retyping until the present form was reached. She has made real what was previously a scrawl. I wish also to thank Michael Miller and Marion Lind for the preparation of the diagrams, and Michael Miller for the construction of the Word List. Much of the material of this book has been used at one time or another in classes at the University of Rochester and students' comment, always welcome, has found itself anonymously incorporated herein. Thanks are also owed to Mary Anne Maher of New Mexico State University, who read much of the manuscript and made many suggestions, especially of diction, which, while not always adopted, were always gratefully received. I also wish to thank Leopoldo Nachbin for suggesting I write this book, and only hope to have fulfilled his faith in it through many delays.

However, beyond all these debts for technical and expert help, thanks are owed to my wife Rima, without whose support, encouragement, and understanding this book would never have been.

FOREWORD

The content of a one-semester course in complex analysis is reasonably certain--except that one would like to include the Riemann Mapping Theorem, but often does not manage to have the time. What a second course should contain is far less clear. In this book we try to give introductions to several (but certainly not all!) of the many topics which naturally present themselves when a first one-semester course has been completed. (There is a presupposition of working familiarity with the Γ-function, Weierstrass Products, and the elements of analytic continuation. Students whose first course failed to include these topics may easily acquire the requisite knowledge from any of the standard texts, e.g. Conway's Introduction to Complex Analysis). In a few places the concept of Lebesgue measure is used, but not in a significant way.

This book is much too large for all topics to be treated in one semester; however, an instructor and students may reasonably select various material to examine or decide on a longer course. It is to be stressed that all chapters are introductions,--indeed the material introduced in almost every chapter has been the subject of book-length presentations (often more than one). Each topic involved may be pursued further in the more specialized books and research papers listed in the course of each chapter. Students are encouraged to do this. Nevertheless, the subject is so vast that no pretense of a complete bibliography is claimed.

The text consists of two interspersed parts: Theorems with Proofs, and Notes. Each is numbered consecutively in each chapter; so that a Note 3.5 say, means that there is a Note 3.4 preceding it within section three of that chapter. The Notes consist of glosses on the text, references to the literature and further extensions, historical remarks, and the like. They are more informal in style than the Theorems and Proofs. With only a few exceptions there are no references to Notes other than in other Notes. Thus, by and large,a reader may in fact read just the Definitions, Theorems, and Proofs, skip all the Notes and acquire a coherent presentation of the material; however, he would be very ill-advised to do so, for the Notes contain a context in which the material should be seen. On the other hand, I have felt free to include occasional mention in the Notes of concepts which are not defined in the text and may be unfamiliar to a reader -- in such a case pursuit of the topic will require preliminarily some consultation of an elementary textbook, or at the very least a good mathematical dictionary such as Naas-Schmidt. Internal textual references are preceded by a Roman numeral indicating chapter number if and only if the reference is not to the chapter in which it appears; thus a reference in Chapter IV to Theorem 3.1 would refer to the first theorem of section 3 of that chapter, while a reference to Theorem III.3.1 would be to the first theorem of section 3 of Chapter III. Above all the attempt has been made to provide a book which can be *read* for profit and is not just a shelf adornment. The material in this book is not original, but it is basic to various areas of one complex variable; I hope it may stimulate students to pursue further such topics.

Chapter I discusses some material on conformal mapping and fills the frequent lacuna of the Riemann Mapping Theorem. The proof using normal families is given, and the topic (usually not discussed) of construction of the mapping for general regions is treated. The Schwarz-Christoffel formula for polygons is also discussed.

Chapter II deals with Picard's Theorem, both from the Bloch-Landau point of view, and using the elliptic modular function (constructed in an *ad hoc* manner). The problem of Bloch's and Landau's constants and the Ahlfors-Grunsky bounds for them is also discussed.

Chapter III presents an introduction to the basic ideas of entire function theory: order, type, the Phragmén-Lindelöf indicator, and relationships elementary in that theory.

Chapter IV presents an introduction to Nevanlinna theory including some of the more initially startling standard applications such as the identity of two functions which assume five distinct values at the same points in the complex plane, or the existence of fixed points of order n. As Nevanlinna Theory may be regarded as a far-reaching deepening of Picard's Theorem, it follows naturally after Chapters II and III. On the other hand, Chapter IV, at least as regards the proof of Nevanlinna's Second Fundamental Theorem, which is the kernel of all later developments, is probably more difficult than any of the preceding material, and seems inevitably to involve a somewhat denser style of exposition. It is possible (although not necessarily recommended) for a reader to omit Section IV.2 and "take on faith" the "second version" of Nevanlinna's Second Fundamental Theorem which appears as Theorem IV.3.1, and which is the version often used in applications.

Chapter V returns to entire functions from a slightly different point of view and presents results on asymptotic values; in particular, Julia's Theorem which deepens Picard's Theorem in a different direction, and the Denjoy-Carleman-Ahlfors Theorem limiting the number of asymptotic values an entire function of finite order can have.

Chapter VI is a change of pace in that it is concerned with functions represented by power series with a finite radius of convergence. Here we discuss some problems of analytic continuation and the many seemingly different kinds of conditions which produce natural boundaries. The Hadamard and Fabry Gap Theorems, overconvergence, the Pólya-Carlson Theorem on power series with integral coefficients, and Pólya's converse of Fabry's Gap Theorem are among the several topics discussed. Nevertheless a certain continuity of ideas with some of those in other chapters should be apparent.

Chapter VII provides an introduction to what has become the classic problem, still unsolved, in the theory of functions univalent in a disk. Several distinct methods have been developed to attack this problem, the so-called Bieberbach conjecture. Among those discussed here are the Grunsky inequalities, a proof that $|a_4| \leq 4$, and the Dieudonné-Rogosinski result for real coefficients. Convexity and starlike conditions are also discussed. The first section is concerned with distortion theorems in general, even if they are not used for investigations of the Bieberbach conjecture in the remainder of the chapter. For reasons of space it has been necessary to omit Loewner's method, despite its importance and "classical" status, except for some references where the student may find expositions. These several sections contain, it is hoped, ample references to the large and various research literature, which is steadily growing.

In Chapter VIII elliptic functions are discussed both from Weierstrass's and Jacobi's point of view. Throughout the emphasis is on the structure of this area of analysis. Because of the antiquity of the subject of elliptic functions, and the way the subject grew, it often seems in its classical analytic aspect like a welter of intriguing but incoherently linked formulas, while, paradoxically, abstract algebraic versions of some of these analytic ideas are in the forefront of contemporary research. The treatment in the rather lengthy Chapter VIII, which, nevertheless, hardly contains all the relevant details, is both "classical" and, I hope, coherent.

Chapter IX first presents a classical proof of the Prime Number Theorem as an example of using complex analysis and as motivation for discussing the Riemann Zeta-function. It concludes with a discussion of Riemann's famous unproved hypothesis concerning the Riemann Zeta-function. The prime number theorem was one of Hadamard's chief motivations in creating entire function theory and so questions solved and unsolved which are related to it seem especially appropriate to a book of this sort. Nevertheless, the chapter is an introduction to the Riemann Zeta-function, and not to the theory of prime numbers, let alone analytic number theory -- thus, there is, for example, no mention of sieve methods, nor even of L-functions. The notes do contain relevant information about results in prime number theory which seem related, but here again a very great deal has been omitted without mention; how much can be seen by referring to some of the standard works cited in the chapter.

The book concludes with an Appendix in which proofs are given of some of those standard tools which rarely find their way into a first course: The Area Theorem, the Borel-Carathéodory Lemma, The Schwarz Reflection Principle, Hadamard's Three Circles Theorem, the Poisson Summation Formula, the Poisson Integral Formula; as well as a special case of the Osgood-Carathéodory Theorem which finds application in Chapter II, and a special case of the Fourier Integral Theorem used in Chapter IX. It also contains brief discussions of Farey Series and Bernoulli numbers.

Throughout this book multiple proofs of the same major result are frequently given in the belief that the demonstration of different points of view can only serve to elucidate a subject. A consequence, of course, is that a subject matter cannot be followed in detail to the same depth that it might otherwise be.

The various chapters are largely independent, though appropriate cross-references are usually given. Some basic ideas that appear throughout the book such as the growth of entire functions, normal families, univalence, are not always cross-referenced after their first introduction and definition as to do so would be excessive. The reader in doubt should be able to use the Word List and Table of Contents to find appropriate definitions if they are not known or the chapters are not being read in sequence. The first four chapters form a natural sequence, and might be considered as a unit for a one-semester course, with, perhaps, some additional material selected from one of the later chapters.

Every effort has been made to eliminate errors, typographical and otherwise; nevertheless it is too much to hope especially in a book this size, that all have been found. Although several colleagues have made suggestions about one point or another, needless to say all such errors are my own. I can only hope that they are neither too frequent nor egregious, and welcome any corrections from readers.

A NOTE ON NOTATIONAL CONVENTIONS

We list here a few notations used throughout this book usually without explicit definition. A <u>region</u> always refers to an open connected set in the plane. $\mathbb{C}$ denotes the complex plane, and $\mathbb{C}_\infty$ the usual "extended plane" $\mathbb{C} \cup \{\infty\}$ which maps onto the Riemann sphere under stereographic projection. If X is a set, $\mathrm{Bd}X$ denotes the boundary of X and $\overline{X}$ the closure of X. $B(a,r)$ denotes the disk with center a and radius r, that is the set $\{z: |z-a| < r\}$ and $C(a,r) = \mathrm{Bd}B(a,r)$ the circle with center a and radius r, namely the set $\{z: |z-a| = r\}$. Thus $\overline{B(a,r)} = B(a,r) \cup C(a,r) = \{z: |z-a| \leq r\}$. All contour integrals are assumed to be taken in the positive (counterclockwise) direction unless explicitly mentioned otherwise. $[y]$ invariably refers to the greatest integer $\leq y$. $\sum'$ indicates a summation in which the term corresponding to 0 has been omitted. $M(r,f)$ (or $M(r)$ if there is no danger of confusion) indicates the maximum modulus of the function f in $B(0,r)$. The Bachmann-Landau O,o notation for error terms is used; namely,

$$f(x) = O(g(x) \quad \text{as} \quad x \to a$$

means that $\left|\frac{f(x)}{g(x)}\right|$ is bounded as $x \to a$,

and $f(x) = o(g(x))$ as $x \to a$

means that $\left|\frac{f(x)}{g(x)}\right| \to o$ as $x \to a$,

$$f(x) \sim g(x) \quad \text{as} \quad x \to a$$

means that $\frac{f(x)}{g(x)} \to 1$ as $x \to a$.

Curves, unless explicitly stated otherwise, are assumed to be rectifiable and without self-crossings.

Any other notation should either be familiar as standard, or defined at appropriate places in each chapter. Some of the chapters contain a summary of special notational conventions of their own.

TABLE OF CONTENTS

CHAPTER I

CONFORMAL MAPPING AND RIEMANN MAPPING THEOREM

Introduction: Without question, the basic theorem in the theory of conformal mapping is Riemann's mapping theorem. It is one of those results one would like to present in a one-semester introductory course in complex variable, but often does not for lack of sufficient time. Indeed, depending on the time available and the text used, elementary conformal mapping in general is a subject which in an introductory course may not be adequately treated. This chapter begins therefore with an introduction to some basic results on conformal mapping especially those involving univalent functions. The concept of normal families is then introduced and developed far enough to be able to give the well-known elegant existence proof of the Riemann Mapping Theorem resulting from the reworking of ideas of Carathéodory and Koebe by Fejér and F. Riesz. We then turn to the related construction of the mapping, following the work of Koebe and Ostrowski. Finally the Schwarz-Christoffel Theorem giving explicit mappings of polygonal regions is treated. Examples are given throughout.

I.1 We begin with a semi-informal review of the basic mapping properties of an analytic function f at a point z_0 where $f'(z_0) \neq 0$.

Suppose C_1 and C_2 are two continuous curves intersecting in a point z_0, and such that each has definite tangents at z_0 (i.e. they represent functions differentiable at z_0). An "angle between C_1 and C_2" is an angle formed by the tangents at z_0.

Let $f(z)$ be an analytic function in a region D of the z-plane. Let z_0 be an interior point of D and C_1, C_2, two continuous curves passing through z_0 which have definite tangents there. Suppose $f'(z_0) \neq 0$. Then an angle between $f(C_1)$ and $f(C_2)$ at z_0 exists, since well-defined tangents exist.

If the tangent to C_1 at z_0 makes the angle α_1 with the real axis and the tangent to C_2 makes the angle α_2 (both measured on the right side of the tangent), clearly $\alpha_2-\alpha_1$ is the "interior" angle between C_1 and C_2 (see Diagram I.1); then it is furthermore true that $\alpha_2-\alpha_1$ as so defined is also the angle between $f(C_1)$ and $f(C_2)$.

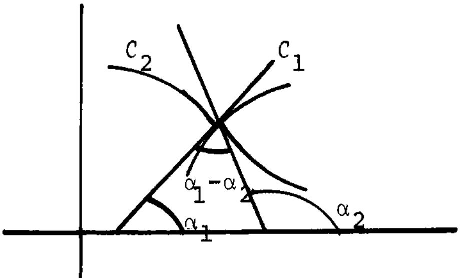

Diagram I.1

This may be seen as follows: If $z_1 \in C_1$ and $z_2 \in C_2$ are two variable points on the curves near z_0 such that $|z_1-z_0| = |z_2-z_0| = r$ say, then

$$z_1 = z_0 + re^{i\theta_1}$$

$$z_2 = z_0 + re^{i\theta_2}$$

and, as $r \to 0$, $z_1 \to z_0$, $z_2 \to z_0$, $\theta_1 \to \alpha_1$, $\theta_2 \to \alpha_2$. Hence since f is analytic in D,

$$f'(z_0) = \lim_{r\to 0} \frac{f(z_1)-f(z_0)}{z_1-z_0} = \lim_{r\to 0} \frac{f(z_2)-f(z_0)}{z_2-z_0} ,$$

and since $f'(z_0) \neq 0$, we can write $f'(z_0) = Re^{i\delta}$ say where $R \neq 0$ and $\delta = \arg f'(z_0)$ is a specific fixed number.

Now let $\phi_1 = \arg(f(z_1)-f(z_0))$, so we may write $f(z_1) - f(z_0) = \rho_1 e^{i\phi_1}$, say.

Then $\lim\limits_{r\to 0} \dfrac{\rho_1 e^{i\phi_1}}{re^{i\theta_1}} = Re^{i\delta}$

and so $\lim\limits_{r\to 0} (\phi_1-\theta_1) = \delta$, whence $\lim\limits_{r\to 0} \phi_1 = \alpha_1 + \delta$.

Since the limit exists, $f(C_1)$ has a definite tangent at $f(z_0)$ which makes the angle $\alpha_1 + \delta$ with the real axis. Similarly $f(C_2)$ makes the angle $\alpha_2 + \delta$ with the real axis, whence the result follows. We should further note that f preserves the sense of the angle.

Also, clearly

$$\lim_{r\to 0} \rho_1/r = |f'(z_0)| \neq 0.$$

In fact, the above shows that the angle between $f(C_1)$ and $f(C_2)$ in question is obtained by a rotation by δ, and any small subregion of D containing z_0 goes into a "similar" subregion of $f(D)$ determined by this rotation and a "stretching" by $|f'(z_0)|$.

Because of this similarity, maps which preserve angles as above are called <u>conformal</u>.

<u>Note 1.1</u>: Such a definition of <u>conformal</u> includes the possibility of a conformal map preserving the magnitude but not the sense of angles. An example of such a map (which is not analytic) is reflection in the real axis $f(z) = \bar{z}$, or, more generally, the map obtained by taking the complex-conjugate of any analytic conformal map. Some authors call such maps "indirect-

ly conformal".

If we suppose f' has a zero of order n at z_0, then $f(C_1)$ and $f(C_2)$ still have definite tangents at z_0, but the angle between them is the angle between C_1 and C_2 multiplied by $n + 1$. For if f' has a zero of exact order n at z_0 then in a neighborhood of z_0,

$$f(z) = f(z_0) + a(z-z_0)^{n+1} + \ldots \quad \text{where} \quad a \neq 0,$$

and so in the previous notation

$$\rho_1 e^{i\phi_1} = f(z_1) - f(z_0) = a(z_1-z_0)^{n+1} + \ldots =$$

$$= |a| r^{n+1} e^{i(\eta+(n+1)\theta_1)} + \ldots \text{ (higher powers of } r)$$

where $\eta = \arg a$. So

$$\lim_{r\to 0} \frac{\rho_1 e^{i\phi_1}}{r^{n+1}} = |a| e^{i(\eta+(n+1)\alpha_1)}$$

and

$$\lim_{r\to 0} \phi_1 = \eta + (n+1)\alpha_1 \ , \quad \lim_{r\to 0} \phi_2 = \eta + (n+1)\alpha_2 .$$

Before turning to a brief study of linear fractional transformations, we give some examples of non-linear transformations.

Example 1.1: $f(z) = z^2$.

If $z = re^{i\theta}$ and $w = f(z) = \rho e^{i\phi}$, then

$$\rho = r^2 \quad \text{and} \quad \phi = 2\theta;$$

hence the angular region $\alpha < \arg z < \beta$ is mapped onto $2\alpha < \arg w < 2\beta$, and if $\beta - \alpha > \pi$ the image region covers part of the plane twice. (The ambiguity can be removed by consideration of the Riemann surface for $\sqrt{w}$).

If $z = x + iy$ and $w = u + iv$, then $u + iv = x^2 - y^2 + 2ixy$.

Hence the pre-image of the straight line $u = a$ is the hyperbola $x^2 - y^2 = a$ and the pre-image of the straight line $v = b$ is the hyperbola $2xy = b$. Suppose a and b are not both 0 and let $\xi \neq 0$ be a point where the hyperbolas meet, and θ the angle at ξ. Since $f'(z) = 2z$ and $2\xi \neq 0$, θ is also the angle between $u = a$ and $v = b$; but this angle is $\pi/2$; hence $\theta = \pi/2$. If $a = 0$ and $b = 0$, then since $f'(z)$ has a simple zero at the origin, if θ is the angle between $x^2 - y^2 = 0$ and $2xy = 0$, 2θ is the angle between $u = 0$ and $v = 0$ which is $\pi/2$; hence $\theta = \pi/4$ (as can also be deduced directly).

Similarly the lines $x = c$, $c \neq 0$ and $y = c$, $c \neq 0$ map onto parabolas meeting at an angle of $\pi/2$, while $x = 0$ and $y = 0$ (the axes) map onto the halflines $v = 0$, $u \leq 0$, and $v = 0$, $u \geq 0$ each described twice. These half lines meet, as is expected, at an angle of π.

<u>Example 1.2</u>: $f(z) = \frac{1-\cos z}{1+\cos z}$.

Writing $z = x + iy$, the strip between $x = 0$ and $x = \pi/2$ is mapped into the open unit disk with the interval $(-1,0]$ deleted. The two bounding lines map on the boundary of the slit disk. The line $\{z: x=0\}$ maps onto the real interval $[-1,0]$ described twice: as y goes from ∞ through positive real values to 0, $f(iy) = \frac{1-\cosh y}{1+\cosh y}$ goes from -1 to 0, and then back from 0 to -1 as y goes from 0 to ∞ through negative real values. The line $\{z: x=\pi/2\}$ maps onto the unit circle described once (as y goes from ∞ to 0 through negative real values, the lower semicircle is described, and as y goes from 0 to ∞ through positive real values the upper semi-circle is described.) One may note in particular that ∞ is mapped onto -1. The line $x = a$, $0 < a < \pi/2$ is mapped onto a loop which cuts the real axis at -1 and at another point where $\operatorname{Im}\left(\frac{1-\cos(a+iy)}{1+\cos(a+iy)}\right) = 0$ which may be easily computed. Each loop contains the slit $(-1,0]$ in its Jordan interior, and is contained in $B(0,1)$. If $0 < a < b < \pi/2$, the loop for $x = b$ contains the loop for $x = a$ in its Jordan interior.

Example 1.3: $f(z) = z + 16/z$

The circle $\{z: |z+1-i| = \sqrt{10}\}$ maps onto the curve of Diagram I.2.

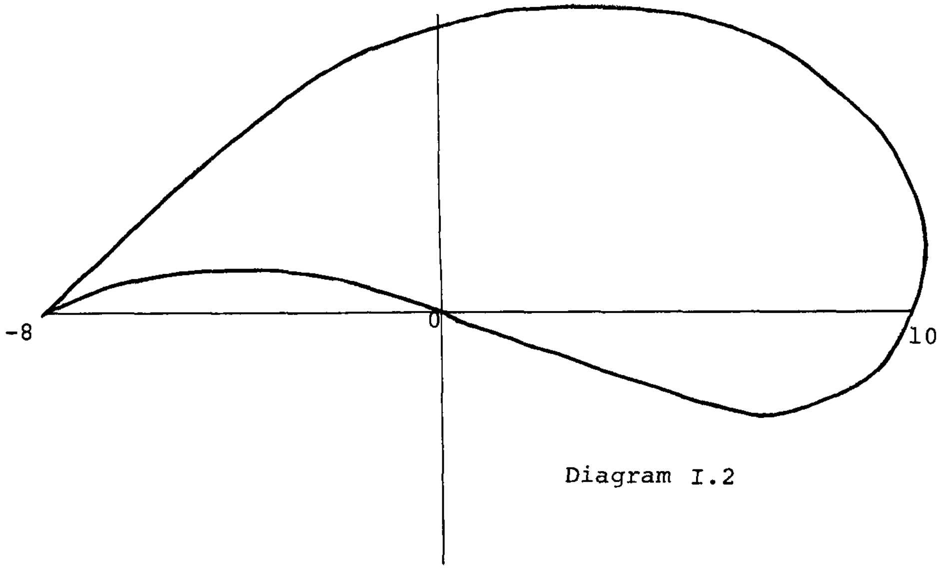

Diagram I.2

Curves of this kind are known as aerofoils, and have had some importance in aerodynamic studies. The interior of the circle maps onto the exterior of the aerofoil.

For further examples with diagrams of the mapping properties of a great variety of functions, the reader is referred to A Dictionary of Conformal Mapping by H. Kober. This was originally prepared for the British Admiralty in 1944-48, and was reissued by Dover Press, New York in 1952.

I.2 Linear fractional transformations

Definition 2.1 A function of the form

$$f(z) = \frac{az+b}{cz+d}, \quad ad - bc \neq 0$$

is called a non-singular linear fractional transformation (or, sometimes, Möbius transformation).

Theorem 2.1. If $f(z)$ is meromorphic in the extended complex plane $\mathbb{C}_\infty$ and maps it one-to-one onto itself, then $f(z)$ is a non-singular linear fractional transformation, and conversely.

Proof: Since $f(z)$ is meromorphic in $\mathbb{C}_\infty$, it is rational. So $f(z) = \frac{p(z)}{q(z)}$ where $p(z)$ and $q(z)$ are polynomials with no common zeros. For a fixed w_0, $\frac{p(z)}{q(z)} = w_0$ for some $z = z_0$ if and only if $w_0 q(z_0) - p(z_0) = 0$. If $\deg(p(z)) > 1$, $wq(z) - p(z)$ is a polynomial of degree > 1 for all but at most one value w and so f is not one-to-one. Similarly the degree of $q(z)$ must be 1. So $f(z) = \frac{az+b}{cz+d}$, c, d not both 0. If $ad-bc = 0$, then if $d \neq 0$, $f(z) = b/d$ and if $c \neq 0$, $f(z) = a/c$; and so f is not onto.

Finally, clearly $f(z) = \frac{az+b}{cz+d}$, $ad-bc \neq 0$ is meromorphic in $\mathbb{C}_\infty$; to prove it is onto:

If $c \neq 0$, $f: -d/c \to \infty$; if $c = 0$ then $a \neq 0$ and $f: \infty \to \infty$.

If $\alpha \neq a/c$, let $w = \frac{-d\alpha+b}{c\alpha-a}$, then $f: w \to \alpha$.

If $\alpha = a/c$, $c \neq 0$, then $f: \infty \to a/c$.

That f is one-to-one is clear.

Theorem 2.2: The set of all non-singular linear fractional transformations forms a group under composition.

Proof: An obvious verification. Note that if

$$\phi(z) = \frac{az+b}{cz+d}\ , \text{ the inverse } \ \phi^{-1}(z) = \frac{-dz+b}{cz-a}\ .$$

Theorem 2.3: A non-singular linear fractional transformation is conformal at all z except $z = -d/c$ (∞ if $c = 0$), where it has a pole.

Proof: If $\phi(z) = \frac{az+b}{cz+d}$, then $\phi'(z) = \frac{ad-bc}{(cz+d)^2}$.

We now briefly turn our attention to circles (and straight lines, which are degenerate circles).

Definition 2.2. Two distinct points of $\mathbb{C}$, p and q are called inverse with respect to the circle $C(z_0,\rho)$, $\rho > 0$, if they are collinear with z_0, lie on the same side of z_0, and $|p-z_0|\ |q-z_0| = \rho^2$.

Suppose p and q are distinct points inverse with respect to $C(z_0,\rho)$, and $p = z_0 + me^{i\mu}$, $m > 0$. Then collinearity implies $\arg(q-z_0) = \arg(p-z_0) = \mu$, and we also have $|q-z_0| = \rho^2/m$; so $q = z_0 + \frac{\rho^2}{m} e^{i\mu}$. If $z \in C(z_0,\rho)$ then $z = z_0 + \rho e^{i\theta}$ for some θ, and

$$\left|\frac{z-p}{z-q}\right| = \left|\frac{\rho e^{i\theta} - me^{i\mu}}{\rho e^{i\theta} - \frac{\rho^2}{m} e^{i\mu}}\right| = \frac{m}{\rho}\left|\frac{\rho e^{i\theta} - me^{i\mu}}{me^{i\theta} - \rho e^{i\mu}}\right| = \frac{m}{\rho}\ ,$$

since $\overline{\rho e^{i\theta} - me^{i\mu}} = \rho e^{-i\theta} - me^{-i\mu} = -e^{-i(\theta+\mu)}(me^{i\theta} - \rho e^{i\mu})$.

Furthermore, since p does not equal q, $m/\rho \neq 1$. Thus we have proved

Theorem 2.4: The circle $C(z_0,\rho)$ is a subset of

$$\left\{z: \left|\frac{z-p}{z-q}\right| = \frac{m}{\rho} \neq 1\right\}\ , \quad \text{where } m = |p-z_0| \neq 0$$

and p and q are distinct points inverse with respect to $C(z_0,\rho)$. The converse of this theorem is also true:

Theorem 2.5: The equation $\frac{z-p}{z-q} = k$, $k \neq 1, \neq 0$ always represents a circle with respect to which p and q are inverse points.

Proof: A somewhat wearisome computation which is left to the reader. The circle turns out to be

$$\left\{ z: \left| z - \frac{p-k^2q}{1-k^2} \right| = \frac{|k|\,|p-q|}{|1-k^2|} \right\}.$$

Note 2.1: The geometric definition of circles by using inverse points goes back to Apollonius (c. 255-170 B.C.). Suppose $p \in B(z_0,\rho)$, $p \neq z_0$, then one can easily give a geometric construction of the point q which is inverse to p with respect to $C(z_0,\rho)$ as follows:

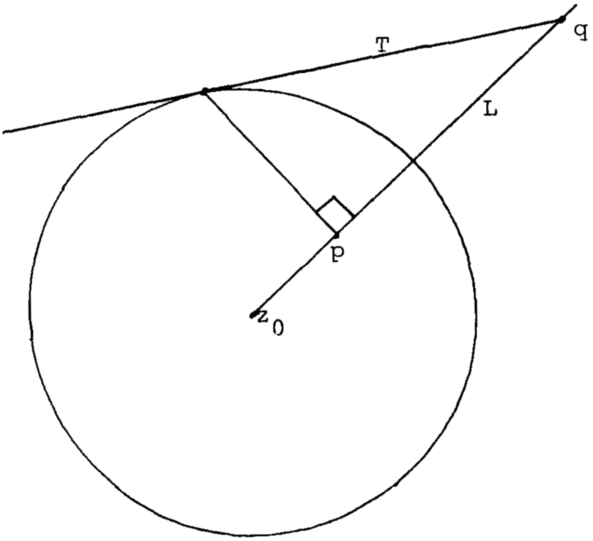

Diagram I.3 (Description on next page.)

If L is the line passing through z_0 and p, construct the perpendicular to L at p, and at the point where this line intersects $C(z_0,\rho)$ draw the tangent T to the circle. The intersection of L and T is q. This construction clearly can be reversed so that given a point in the complement of $\overline{B(z_0,\rho)}$, the point inverse to it with respect to $C(z_0,\rho)$ can be found.

Note 2.2: As $p \to z_0$ along L in Diagram I.3, $q \to \infty$ (this may also be seen directly from Definition 2.2). We therefore make the definition: z_0 and ∞ are inverse with respect to $C(z_0,\rho)$.

Note 2.3: $S = \{z: |\frac{z-p}{z-q}| = 1\}$ is a straight line, for $z \in S$ means z lies equidistant from p and q, hence lies on perpendicular bisector of line joining them, and this is the set of all points equidistant from p and q.

Because of this last note, it is conventional to consider straight lines as "degenerate" circles (they may also be viewed as circles passing through ∞.) Therefore, for the rest of this section, the word "circle" will mean "circle or straight line".

Theorem 2.6: Under a non-singular linear fractional transformation, circles go into circles and inverse points with respect to the original circle into inverse points with respect to the image circle (inverse points with respect to a straight line meaning points symmetrical about the line and equidistant from it).

Proof: Let $K = \{z: \left|\frac{z-p}{z-q}\right| = k, k \neq 0\}$, where k is fixed. Let $f(z) = \frac{az+b}{cz+d}$, $ad - bc \neq 0$. Since f is one-to-one, $w \in f(K)$ if and only if $f^{-1}(w) \in K$, that is if and only if

$\frac{-dw + b}{cw - a} \in K$, that is if and only if

$$\left| \frac{\frac{-dw + b}{cw - a} - p}{\frac{-dw + b}{cw - a} - q} \right| = k, \text{ that is if and only if}$$

$$\left| \frac{w - \frac{ap + b}{cp + d}}{w - \frac{aq + b}{cq + d}} \right| = k \left| \frac{cq + d}{cp + d} \right| ,$$

assuming $cq + d \neq 0$, $cp + d \neq 0$, which proves the result under this assumption. If $cq + d = 0$, $f: q \to \infty$, and the equation for $f(K)$ becomes

$$\left| w - \frac{ap + b}{cp + d} \right| = \frac{k}{|c|} \frac{|ad - bc|}{|cp + d|} ,$$

and $\frac{ap + b}{cp + d}$ and ∞ are inverse with respect to $f(K)$ by Note 2.2. Similarly, if $cp + d = 0$, $f: p \to \infty$ and the equation for $f(K)$ becomes

$$\left| w - \frac{aq + b}{cq + d} \right| = \frac{|ad - bc|}{k|c| |cq+d|} .$$

We now give some examples of the use of linear fractional transformations.

Example 2.1: Problem: Find all non-singular linear fractional transformations $f: z \to w = f(z)$ of the closed upper half-plane $\{z: \text{Im } z \geq 0\}$ onto the closed unit disk $\{w: |w| \leq 1\}$ with the real line mapping onto the unit circle.

Solution: Suppose $f(z) = \frac{az + b}{cz + d}$ is such a transformation. If $a = 0$, then $f: \infty \to 0$ contradicting the real line mapping onto the unit circle, so $a \neq 0$. If $c = 0$, $f: \infty \to \infty$ which is impossible. So also $c \neq 0$, and $f: {}^{-b}/a \to 0$, $f: {}^{-d}/c \to \infty$.

Now since f is non-singular linear fractional so is f^{-1}, and 0 and ∞ are inverse points with respect to $C(0,1)$. Hence by Theorem 2.6, $^{-b}/a$ and $^{-d}/c$ must be inverse points with respect to the real line $\{z\colon \operatorname{Im} z = 0\}$. Hence we can write $-b/a = \alpha$, $-d/c = \overline{\alpha}$ where $\operatorname{Im} \alpha \neq 0$.

So $f(z) = \dfrac{az+b}{cz+d} = \dfrac{a}{c}\left(\dfrac{z-\alpha}{z-\overline{\alpha}}\right)$.

Now $|f(0)| = 1$ (since the real axis maps onto the unit circle), and so $\left|\left(\dfrac{a}{c}\right)\left(\dfrac{-\alpha}{-\overline{\alpha}}\right)\right| = 1$, or $|a| = |c|$, whence $a = ce^{i\lambda}$, λ real, and $f(z) = e^{i\lambda}\left(\dfrac{z-\alpha}{z-\overline{\alpha}}\right)$. Finally, since $f(\alpha) = 0 \in B(0,1)$, we have $\operatorname{Im} \alpha > 0$.

It remains to verify that all transformations of the form

$$f(z) = e^{i\lambda}\left(\frac{z-\alpha}{z-\overline{\alpha}}\right), \quad \lambda \text{ real}, \quad \operatorname{Im} \alpha > 0$$

do in fact map the closed upper half-plane onto the closed unit disk in the desired fashion. But it is an easy computation that, if $\operatorname{Im} z = 0$ then $|f(z)| = 1$, while if $\operatorname{Im} z > 0$, then $|f(z)| < 1$.

It is worth noting that since there are three arbitrary constants: λ, $\operatorname{Im} \alpha$, $\operatorname{Re} \alpha$ in the solution, one can make three given points on the real axis correspond to three points on the circle.

The reader, if he wishes, can invent any number of similar exercises for himself mapping specified circles or half-planes onto other circles or half-planes.

Example 2.2: Problem: Suppose $f(z)$ is analytic in $B(0,1)$ and $\operatorname{Im} f(z) > 0$. Find an inequality on $|f'(0)|$.

Solution: If one could find a linear fractional transformation ϕ of $\mathbb{C}$ onto $\mathbb{C}$ such that $\phi\colon f(z) \to g(z) = \phi(f(z))$ with the properties:

(i) $\phi: \{z: \operatorname{Im} f(z) > 0\} \xrightarrow{\text{onto}} \{z: |g(z)| < 1\}$

(ii) $\phi: f(0) \to 0$,

then $|g(z)| < 1$ for $\operatorname{Im} f(z) > 0$, that is for $z \in B(0,1)$, and also $g(0) = 0$, whereupon Schwarz' Lemma leads to the desired inequality. Let $f(0) = \xi$. By Example 2.1, (i) will hold if g is of the form

$$g(z) = \phi(f(z)) = e^{i\lambda}\left(\frac{f(z)-\alpha}{f(z)-\bar{\alpha}}\right), \quad \operatorname{Im} \alpha > 0,$$

and property (ii) requires $\alpha = \xi$.

So $g(z) = e^{i\lambda}\left(\dfrac{f(z)-\xi}{f(z)-\bar{\xi}}\right)$, whence

$$f(z) = \frac{\bar{\xi}\, g(z)-\xi\, e^{i\lambda}}{g(z)-e^{i\lambda}}, \quad \text{and} \quad f'(z) = \frac{e^{i\lambda} g'(z)(\xi-\bar{\xi})}{(g(z)-e^{i\lambda})^2} .$$

Since $g(0) = 0$, $|f'(0)| = 2 \operatorname{Im} \xi |g'(0)|$. By Schwarz' Lemma, $|g'(0)| \leq 1$, and so we get the result

$$|f'(0)| \leq 2 \operatorname{Im} f(0).$$

Linear fractional transformations are also closely related to the concept of cross-ratio, perhaps familiar to some readers from projective geometry. Many standard complex variable texts contain discussion of this connection. Thron's complex variable text (John Wiley, 1953) also contains an elementary analysis of various subclasses of linear fractional transformations.

1.3 Univalent Functions

Definition 3.1: A function f is univalent (simple, schlicht) in a region D if it is analytic (and single-valued) in D and does not take any value more than once in D.

Note 3.1: The definition says that $f: D \to f(D)$ is one-to-one, but points in $f(D)$ may have other points outside D as preimages under f.

The terminology "univalent" for the concept of Definition 3.1 seems to have won out in English. The original German word "schlicht" was used in English as a loan word for many years, and, for a while, "simple" had a vogue. Readers will find both terms used in other books and papers they may consult.

Theorem 3.1: If f is univalent in a region D, then $f'(z) \neq 0$ for all $z \in D$.

Proof: If for some $z_0 \in D$, $f'(z_0) = 0$, then $g(z) = f(z) - f(z_0)$ has a zero of order $n \geq 2$ at z_0. Since f is univalent, there is a closed neighborhood K of z_0, $K \subset D$ such that $g(z) \neq 0$ on the boundary of K (f is non-constant) and the only zero of f' in the interior of K is at z_0 (since the zeros of non-constant analytic functions are isolated).

Suppose $K = \overline{B(z_0,\delta)}$. Then $\inf\limits_{z \in C(z_0,\delta)} |g(z)| = m$ is > 0. Suppose a is any complex number with $0 < |a| < m$. Then by Rouché's Theorem $g(z) - a$ has as many zeros as $g(z)$ does in $B(z_0,\delta)$, and hence at least two zeros there. But since $\frac{d}{dz}(g(z)-a) = g'(z) = f'(z) \neq 0$ for $z \in B(z_0,\delta) - \{z_0\}$, $g(z) - a$ does not have a multiple zero in $B(z_0,\delta)$. Hence $g(z)$ takes the value a for two distinct values of $z \in B(z_0,\delta) \subset D$; hence $f(z)$ takes on $f(z_0) + a$ for these two values, contradicting f univalent.

Theorem 3.2: A univalent function of a univalent function is univalent.

Proof: Clear.

Theorem 3.3: The inverse of a univalent function f is a univalent function ϕ with domain ϕ = range f.

Proof: Clear.

Theorem 3.4: If f is univalent and maps $\overline{B(0,1)}$ onto itself with $f(0) = 0$ and some point on the unit circle remaining fixed, then $f(z) = z$.

Proof: By Schwarz' Lemma since $f(0) = 0$ and $|f(z)| \leq 1$ for $|z| \leq 1$, $|f(z)| \leq |z|$ for $|z| \leq 1$. The same argument applied to the inverse function $\phi(w)$ defined by $\phi(f(z)) = z$ gives $|z| \leq |f(z)|$. So $\left|\frac{f(z)}{z}\right| = 1$ for $|z| \leq 1$, whence $\frac{f(z)}{z}$ is constant by the maximum modulus principle.

So $f(z) = e^{i\theta}z$ for some fixed $\theta, 0 \leq \theta \leq 2\pi$. But f fixes some point $e^{i\psi}$, ψ real; hence $\theta = 0$.

Theorem 3.5: If f is univalent and maps the closed unit disk onto itself, then f is a non-singular fractional transformation.

Proof: Clearly non-singular linear fractional transformations are univalent.

Suppose $f(0) = \zeta$. Then there exists a non-singular linear fractional transformation L such that L maps the closed unit disk onto itself and such that $L(\zeta) = 0$. (In fact, since 0 and ∞ are inverse with respect to $C(0,1)$, it is easy to see by Theorem 2.6 that

$$e^{i\lambda}\left(\frac{z-\zeta}{\overline{\zeta}z-1}\right) \quad \text{is such a mapping.)}$$

Let $\psi(z) = L(f(z))$; then ψ maps the closed unit disk onto itself, is univalent, and $\psi(0) = 0$.

Hence as in the preceding proof $\psi(z) = az$, where $|a| = 1$. So $f(z) = L^{-1}(\psi(z)) = L^{-1}(az)$ is a non-singular linear fractional transformation.

Theorem 3.6: Suppose $f_n(z)$ is univalent in a region D for each $n = 1,2,\ldots$ and suppose as $n \to \infty$, $f_n(z) \to f(z)$ uniformly on compact subsets of D. Then f is either univalent or a constant.

Proof: $f(z)$ is analytic in D (by the "Weierstrass Double-series Theorem"). Suppose it were not univalent. Then there are $z_1, z_2 \in D$; $z_1 \neq z_2$ and a complex number ζ such that

$$f(z_1) = f(z_2) = \zeta,$$

Let K_1, K_2 be two closed disks with centers z_1, z_2 respectively such that $K_1 \subset D$, $K_2 \subset D$, $K_1 \cap K_2 = \phi$. Suppose f is non-constant. Then we can also choose K_1 and K_2 so that $f(z) \neq \zeta$ for z on the boundary of $K_1 \cup K_2$. Let

$$\inf_{z \in Bd(K_1 \cup K_2)} |f(z)-\zeta| = m.$$

Then $m > 0$, and we can find an n_0 so large that for $n \geq n_0$, $|f_n(z)-f(z)| < m$ for $z \in Bd(K_1 \cup K_2)$. So, since $f_n(z) - \zeta = (f_n(z)-f(z)) + (f(z)-\zeta)$, by Rouché's Theorem $f_n(z) - \zeta$ has as many zeros in $K_1 \cup K_2$ as $f(z) - \zeta$, namely, at least two, contradicting the univalence of $f_n(z)$.

Note 3.1: The case of a constant limit in Theorem 3.6 can actually occur as the example $f_n(z) = z/n$ shows.

Theorem 3.7: If f is analytic at $z = 0$ and $f'(0) \neq 0$, then there is a $\rho > 0$ such that f is univalent in $\overline{B(0,\rho)}$.

Proof: $f(z) = \sum_{n=0}^{\infty} a_n z^n$, $a_1 \neq 0$, for all $z \in B(0,R)$ for some R. If $f(z_1) = f(z_2)$ for some $z_1, z_2 \in B(0,R)$ then

$$\sum_{n=1}^{\infty} a_n(z_1^n - z_2^n) = 0$$ and so

$$0 = |z_1-z_2|\ |a_1 + \sum_{n=2}^{\infty} a_n(z_1^{n-1} + z_1^{n-2}z_2+\ldots+z_2^{n-1})|$$

$$\geq |z_1-z_2| \left(|a_1| - \left|\sum_{n=2}^{\infty} a_n(z_1^{n-1} + z_1^{n-2}z_2+\ldots+z_2^{n-1})\right|\right)$$

$$\geq |z_1-z_2| \left(|a_1| - \sum_{n=2}^{\infty} |a_n| nR^{n-1}\right) .$$

Now as $R \to 0$, $\sum_{n=2}^{\infty} n|a_n|R^{n-1} \to 0$, since a power series is uniformly convergent in its disk of convergence. Hence for all R sufficiently small, $\sum_{n=2}^{\infty} n|a_n|R^{n-1} < |a_1|$, and so $z_1 = z_2$.

A proof of Theorem 3.7 can also be given using Rouché's Theorem in a by now familiar way. We close this section with a well-known result which provides yet another indication of the influence of boundary behavior on the values of an analytic function.

Theorem 3.8: (Darboux) Let C be a simple closed rectifiable (oriented) curve which is the boundary of a region D. Suppose f is analytic on $D \cup C$ and one-to-one on C, then f is univalent in D.

Proof: $f(C) = \Gamma$ is clearly a simple closed rectifiable (oriented) curve. Let Δ = the Jordan interior of Γ. Let $z_0 \in D$ be such that $f(z) - f(z_0) \neq 0$ for $z \in C$. Then letting $w_0 = f(z_0)$ the number of zeros of $f(z) - f(z_0)$ in D is

$$\frac{1}{2\pi i} \int_C \frac{f'(z)}{f(z)-f(z_0)} dz = \frac{1}{2\pi i} \int_\Gamma \frac{1}{w-w_0} dw,$$

on making the substitution $w = f(z)$. The left side of this equation is an integer ≥ 1 (since $z_0 \in D$), the right side is ± 1 or 0 according as $w_0 \in \Delta$ or $w_0 \notin \Delta$, and depending on the direction in which Γ is described. Hence the only possibility is 1, $w_0 \in \Delta$, and f takes the value w_0 just once (at z_0) in D.

Note 3.2: This theorem of Darboux is susceptible to an extensive topological generalization. For example, among other related results, Meisters and Olech (Duke Mathematical Journal (1963), 63-80) have proved the following

Theorem: Let X be a compact subset of n-dimensional Euclidean space $\mathbb{R}_n$ $(n \geq 2)$ such that Bd X is an irreducible separating set of $\mathbb{R}_n$. Let E be a possible "exceptional set" with the properties: (i) $E \cap \text{Int } X$ is discrete, (ii) $\text{Bd } X - E \neq \phi$. Suppose f is a continuous mapping of X into $\mathbb{R}_n$ which is locally one-to-one on $X - E$.

Then if $f|\text{Bd } X$ is one-to-one, f is a homeomorphism of X onto $f(X)$.

That an analytic function in the complex plane is a locally one-to-one mapping on its domain of definition, except for a discrete set E, can be seen by computing its Jacobian.

We shall return to univalent functions in more detail in later chapters. Their study is one of the most active areas of present research in one complex variable.

I.4 Normal Families

Definition 4.1: A family F of complex-valued functions f defined on a region D of the complex plane is called normal if every sequence $\{f_n\}$ of functions in F either contains a subsequence $\{f_{n_k}\}$ such that $\{f_{n_k}\}$ converges uniformly, or a subsequence $\{f_{n_k}\}$ which tends uniformly to ∞ (i.e. given

$\delta > 0$, $|f_{n_k}(z)| \geq \delta$ for $k \geq k_0$ and all z), on every compact subset of D.

Note 4.1: If $\mathcal{F}$ is a normal family of analytic functions then the limit function F, say, of $\{f_{n_k}\}$ is either an analytic function or identically ∞.

Note 4.2: F need not belong to $\mathcal{F}$. (This is the only difference between "normal" and "sequentially compact.")

Note 4.3: If $\mathcal{F}$ is a normal family of analytic functions and $\mathcal{F}' = \{f' : f \in \mathcal{F}\}$, then $\mathcal{F}'$ need not be normal. For, consider the family $\mathcal{F}$ of all functions $f_n(z) = nz^2 - n^2$ defined on the whole plane. $\mathcal{F}$ is a normal family since $f_n \to \infty$ uniformly on every compact subset of the plane. But $f_n'(z) = 2nz$, and $\mathcal{F}'$ is not normal since $f_n'(z) \to \infty$ for $z \neq 0$ but $\to 0$ for $z = 0$.

The existence of "good subsequences" is clearly a useful property for a family of functions to have, but no definition is much use unless it is satisfied in some way not immediately obvious, but nevertheless useful. In this connection there is a famous theorem due in various versions to Ascoli, Arzela (and others, though these are the names usually prefaced to it even in more generalized versions). The theorem has proved of use in several areas of analysis.

We recall first

Definition 4.2: A family of complex-valued functions $\mathcal{F}$ is said to be equicontinuous on a subset E of the complex plane if and only if for each $\varepsilon > 0$, there is a $\delta > 0$ such that whenever $z, z_0 \in E$ and $|z-z_0| < \delta$, then $|f(z)-f(z_0)| < \varepsilon$ simultaneously for all $f \in \mathcal{F}$.

Theorem 4.1: (Arzela-Ascoli): A family $\mathcal{F}$ of continuous complex-valued functions defined on a region D of the complex plane is normal if

(i) $\mathcal{F}$ is equicontinuous on every compact subset of D;
and
(ii) For each $z \in D$, $\{f(z): f \in \mathcal{F}\}$ lies in a compact subset of the plane.

Proof: D contains an everywhere dense countable set of points (e.g. the points with rational coordinates contained in it). Let $\{\zeta_k\}$ denote these points in some fixed ordering. For a fixed value of k, say $k = 1$ the sequence $\{f_n(\zeta_1)\}$ lies in a compact subset by (ii), and so some subsequence of it converges. We now repeat the process starting with this subsequence and $k = 2$. Repeating this process for each successive fixed value of k we thus find an array of subscripts

$$\begin{array}{lllllll}
n_{1,1} & < & n_{1,2} & < \cdots < & n_{1,i} & < & \cdots \\
n_{2,1} & & \cdots\cdots & < & n_{2,i} & < & \cdots \\
\cdot & & & & & & \\
\cdot & & & & & & \\
\cdot & & & & & & \\
n_{k,1} & & \cdots\cdots & < & n_{k,i} & < & \\
\cdot & & & & & & \\
\cdot & & & & & & \\
\cdot & & & & & &
\end{array}$$

such that

(a) Each row is a subsequence of the preceding row, and

(b) $\lim_{i\to\infty} f_{n_{k,i}}(\zeta_k)$ exists for each k.

Consider the diagonal sequence $n_{i,i}$; it is strictly increasing and ultimately a subsequence of each row of the