STUDENT MATHEMATICAL LIBRARY
Volume 9

Inversion Theory and Conformal Mapping

David E. Blair

AMS
AMERICAN MATHEMATICAL SOCIETY

STUDENT MATHEMATICAL LIBRARY
Volume 9

Inversion Theory and Conformal Mapping

David E. Blair

AMERICAN MATHEMATICAL SOCIETY

2000 *Mathematics Subject Classification.* Primary 26B10, 30–01, 30C99, 51–01, 51B10, 51M04, 52A10, 53A99.

About the cover: The picture on the cover is M. C. Escher's "Hand with Reflecting Sphere", a self-portrait of the artist. Reflection (inversion) in a sphere is a conformal map; thus, while distances are distorted, angles are preserved and the image is recognizable. It is for this reason that we commonly use spherical mirrors, e.g. the right hand rear view mirror on an automobile.

Library of Congress Cataloging-in-Publication Data

Blair, David E., 1940–
Inversion theory and conformal mapping / David E. Blair.
p. cm. — (Student mathematical library, ISSN 1520-9121 ; v. 9)
Includes bibliographical references and index.
ISBN 0-8218-2636-0 (softcover : alk. paper)
1. Inversions (Geometry) 2. Conformal mapping. I. Title. II. Series.
QA473.B57 2000
516.3′6—dc21 00-033137

The paper used in this book is acid-free and falls within the guidelines established to ensure permanence and durability.
Visit the AMS home page at URL: http://www.ams.org/

10 9 8 7 6 5 4 3 2 1 05 04 03 02 01 00

To Marie and Matthew

Contents

Preface

It is rarely taught in an undergraduate, or even graduate, curriculum that the only conformal maps in Euclidean space of dimension greater than 2 are those generated by similarities and inversions (reflections) in spheres. This contrasts with the abundance of conformal maps in the plane, a fact which is taught in most complex analysis courses. The principal aim of this text is to give a treatment of this paucity of conformal maps in higher dimensions. The result was proved in 1850 in dimension 3 by J. Liouville [**22**]. In Chapter 5 of the present text we give a proof in general dimension due to R. Nevanlinna [**26**] and in Chapter 6 give a differential geometric proof in dimension 3 which is often regarded as the classical proof, though it is not Liouville's proof. For completeness, in Chapter 4 we develop enough complex analysis to prove the abundance of conformal maps in the plane.

In addition this book develops inversion theory as a subject along with the auxiliary theme of "circle preserving maps".

The text as presented here is at the advanced undergraduate level and is suitable for a "capstone course", topics course, senior seminar, independent study, etc. The author has successfully used this material for capstone courses at Michigan State University. One particular feature is the inclusion of the paper on circle preserving transformations by C. Carathéodory [**6**]. This paper divides itself up nicely into small sections, and students were asked to present the paper to the class. This turned out to be an enjoyable and profitable experience for the students. When there were more than enough students in the class for this exercise, some of the students presented Section 2.8.

The author expresses his appreciation to Dr. Edward Dunne and the production staff of the American Mathematical Society for their kind assistance in producing this book.

Chapter 1

Classical Inversion Theory in the Plane

1.1. Definition and basic properties

Let $\mathcal{C}$ be a circle centered at a point O with radius r. If P is any point other than O, the *inverse* of P with respect to $\mathcal{C}$ is the point P' on the ray $\overrightarrow{OP}$ such that the product of the distances of P and P' from O is equal to r^2. Inversion in a circle is sometimes referred to as "reflection" in a circle; some reasons for this will become apparent as we progress.

Clearly if P' is the inverse of P, then P is the inverse of P'. Note also that if P is in the interior of $\mathcal{C}$, P' is exterior to $\mathcal{C}$, and vice-versa. So the interior of $\mathcal{C}$ except for O is mapped to the exterior and the exterior to the interior. $\mathcal{C}$ itself is left pointwise fixed. O has no image, and no point of the plane is mapped to O. However, points close to O are mapped to points far from O and points far from O map to points close to O. Thus adjoining one "ideal point", or "point at infinity", to the Euclidean plane, we can include O in the domain and range of inversion. We will treat this point at infinity in detail in Section 2.2.

We denote by $\overline{PQ}$ the length of the line segment PQ. The similarity and congruence of triangles will be denoted by $\sim$ and $\cong$ respectively.

Given $\mathcal{C}$, note the ease with which we can construct the inverse of a point P. If P is interior to $\mathcal{C}$, construct the perpendicular to OP at P meeting the circle at T; the tangent to $\mathcal{C}$ at T then meets $\overrightarrow{OP}$ at the inverse point P' (Figure 1.1). To see this, simply observe that $\triangle OPT \sim \triangle OTP'$ and hence

$$\frac{\overline{OP}}{\overline{OT}} = \frac{\overline{OT}}{\overline{OP'}}.$$

Therefore $\overline{OP} \cdot \overline{OP'} = \overline{OT}^2 = r^2$. If P is exterior to $\mathcal{C}$, construct a tangent to $\mathcal{C}$ from P meeting $\mathcal{C}$ at T; the perpendicular from T to $\overrightarrow{OP}$ meets $\overrightarrow{OP}$ at the inverse point by virtue of the same argument.

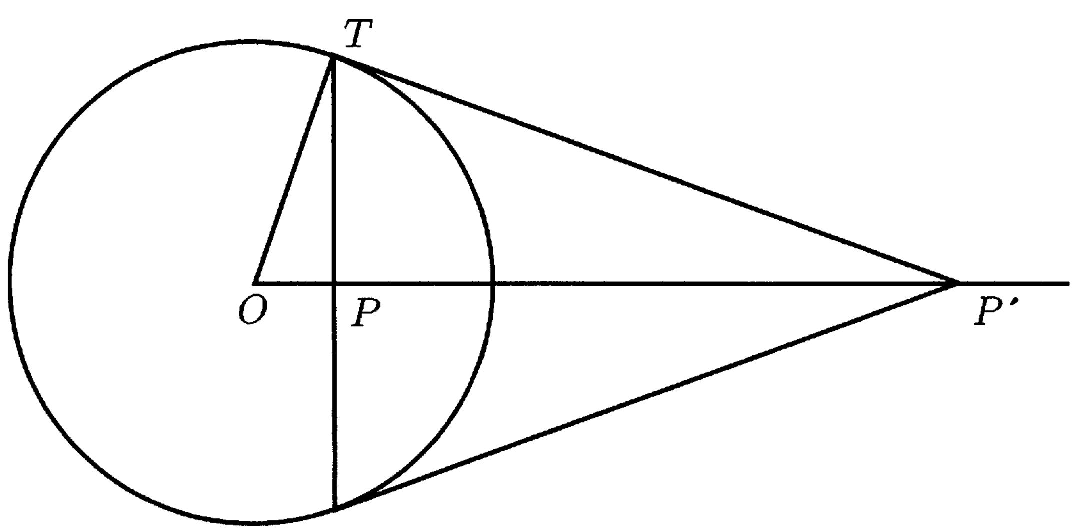

Figure 1.1

A common alternate construction of the inverse point is the following. Construct the diameter of $\mathcal{C}$ perpendicular to OP at O meeting $\mathcal{C}$ at points N and S. Draw $\overrightarrow{NP}$ meeting $\mathcal{C}$ at Q and draw $\overrightarrow{SQ}$ meeting $\overrightarrow{OP}$ at P' (Figure 1.2). Then $\triangle NOP \sim \triangle NQS \sim \triangle P'OS$ and hence $\overline{OP}/\overline{ON} = \overline{OS}/\overline{OP'}$, giving $\overline{OP} \cdot \overline{OP'} = \overline{ON} \cdot \overline{OS} = r^2$. Therefore P' is the inverse of P. Here P may be interior or exterior to $\mathcal{C}$.

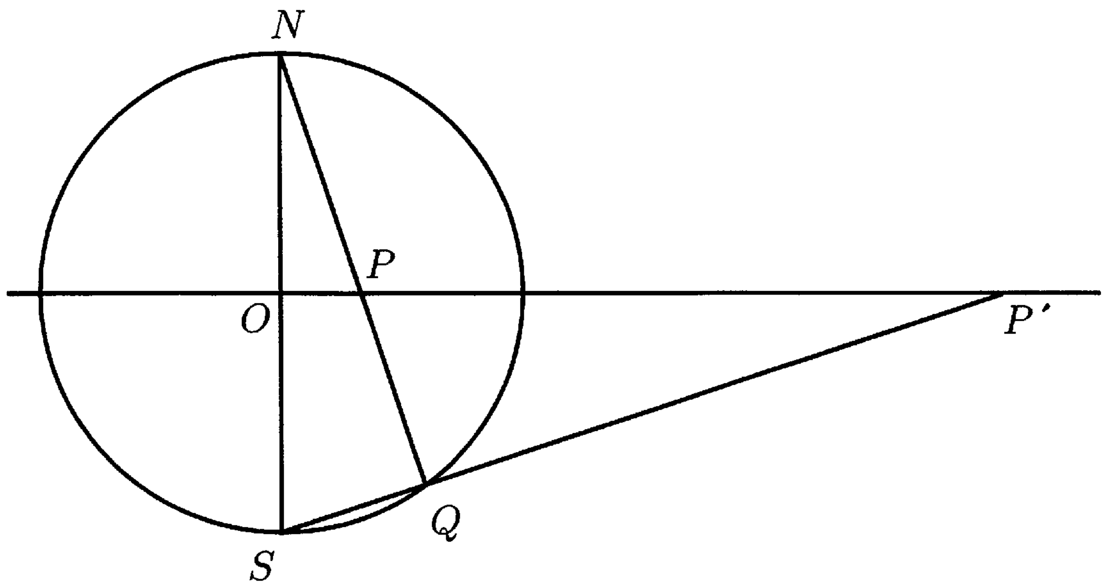

Figure 1.2

Yet another construction of the inverse of a point is given by the following exercise; though more complicated, we will make use of this construction in Chapter 2.

EXERCISE

Construct a radius of $\mathcal{C}$ perpendicular to OP at O meeting the circle at N, and construct the circle $\mathcal{D}$ with diameter ON. Draw NP meeting $\mathcal{D}$ at Q. Draw the parallel to ON through Q meeting $\mathcal{D}$ at Q'. Show that the ray $\overrightarrow{NQ'}$ meets $\overrightarrow{OP}$ at the inverse point P'.

The first basic property of inversion that we will prove is that lines and circles as a class are mapped to lines and circles.

Theorem 1.1. a) *The inverse of a line through the center of inversion is the line itself.*

b) *The inverse of a line not passing through the center of inversion is a circle passing through the center of inversion.*

c) *The inverse of a circle through the center of inversion is a line not passing through the center of inversion.*

d) *The inverse of a circle not passing through the center of inversion is a circle not passing through the center of inversion.*

Proof. Let $\mathcal{C}$ be the circle of inversion with center O and radius r. Since O is collinear with any pair of inverse points, a) is clear. For b), drop the perpendicular from O to the line meeting at P and let P' be the inverse of P (Figure 1.3).

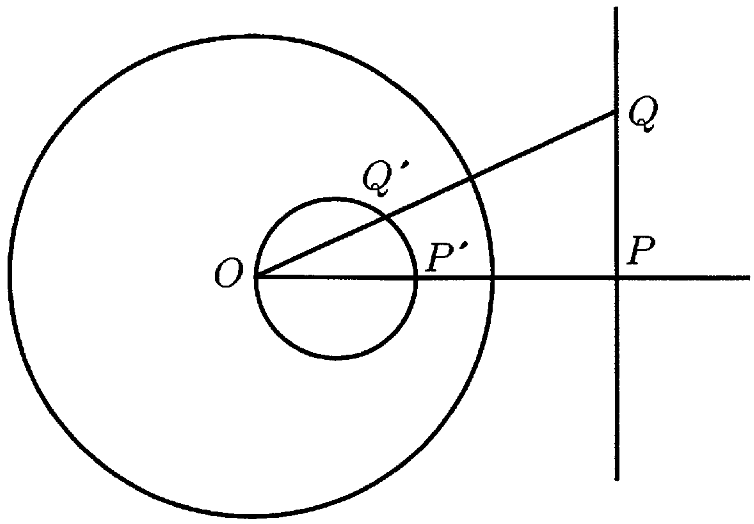

Figure 1.3

Let Q be any other point on the line and Q' its inverse. Then $\overline{OP} \cdot \overline{OP'} = \overline{OQ} \cdot \overline{OQ'} = r^2$ or $\overline{OP}/\overline{OQ} = \overline{OQ'}/\overline{OP'}$, and hence $\triangle OPQ \sim \triangle OQ'P'$. Therefore $\angle OQ'P'$ is a right angle, and hence Q' is on the circle $\mathcal{A}$ of diameter OP'. Thus the image of the line lies in the point set of $\mathcal{A}$; now reverse the argument to show that any point $Q' \neq O$ on $\mathcal{A}$ is the image of some point on the line.

To prove c), let P be the point on the given circle diametrically opposite to O and extend, if necessary, this diameter to the inverse point P' of P (in Figure 1.3 reverse the roles of P and P' and Q and Q'). Let Q be any other point on the circle and Q' its inverse. Again $\triangle OPQ \sim \triangle OQ'P'$. Therefore $\angle OP'Q'$ is a right angle and hence Q' is on the perpendicular to $\overrightarrow{OP}$ at P'; the result then follows as before.

Finally to prove d), let $\mathcal{A}$ be the given circle with center A. If $O = A$ the result is immediate, so assume $O \neq A$. Draw the line through O and A cutting $\mathcal{A}$ at P and Q, and let P' and Q' be the inverse points of P and Q respectively. Let R be any other point on $\mathcal{A}$, and R' its inverse (Figure 1.4). Then $\overline{OP} \cdot \overline{OP'} = \overline{OR} \cdot \overline{OR'} = r^2$ and hence $\triangle OPR \sim \triangle OR'P'$. Similarly $\triangle OQR \sim \triangle OR'Q'$. Thus $\angle OPR \cong \angle OR'P'$ and $\angle OQR \cong \angle OR'Q'$, but $\angle PRQ$ is a right angle and therefore $\angle P'R'Q'$ is a right angle. Thus as the point R moves on $\mathcal{A}$, R' moves on the circle $\mathcal{A}'$ with diameter $P'Q'$ and any point on $\mathcal{A}'$ is the image of a point of $\mathcal{A}$. □

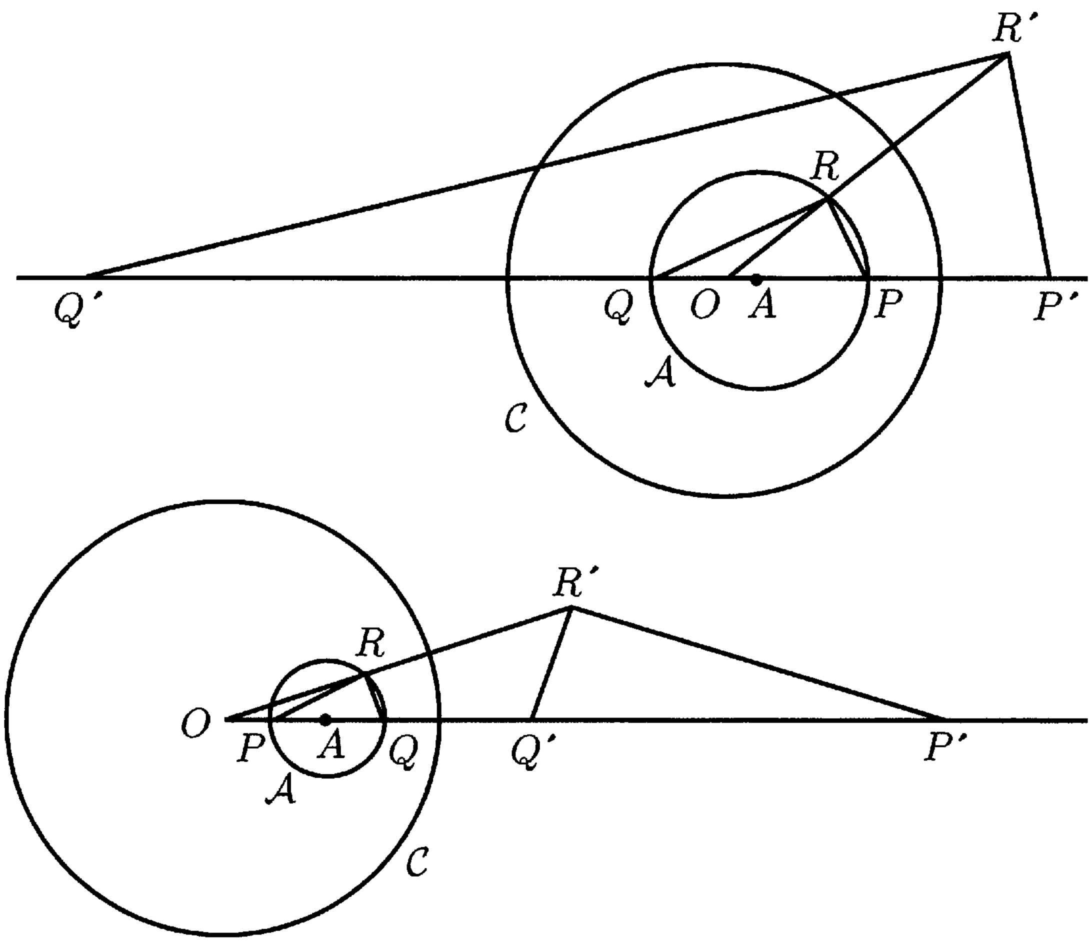

Figure 1.4

The second basic property of inversion is that a circle orthogonal to the circle of inversion inverts to itself.

Theorem 1.2. *Any circle through a pair of inverse points is orthogonal to the circle of inversion; and, conversely, any circle cutting the circle of inversion orthogonally and passing through a point P, passes through its inverse P'.*

This theorem is an immediate consequence of the well known theorem in Euclidean geometry that a tangent to a circle from an external point is the mean proportional between the segments of any secant from the point. To see this, consider the segment PT of a tangent to a circle from an external point P making contact at T.

Let R and S be the intersection points of a secant from P (Figure 1.5). Then $\triangle PRT \sim \triangle PTS$, and hence $\overline{PR}/\overline{PT} = \overline{PT}/\overline{PS}$.

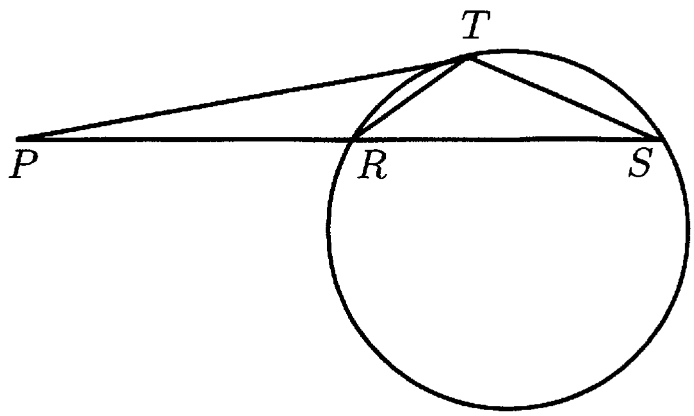

Figure 1.5

Corollary 1.1. *A circle orthogonal to the circle of inversion inverts to itself.*

Corollary 1.2. *Through two points P and Q in the interior of a circle C and not on the same diameter, there exists one and only one circle orthogonal to C.*

Remark. This last corollary is important for the Poincaré model of the hyperbolic plane. Consider a geometry whose points are the interior points of a circle C and whose lines are the diameters of C and arcs of circles orthogonal to C. The corollary assures the existence and uniqueness of a line through two given points. It is also easy to see that the parallel postulate of Euclidean geometry does not hold in this geometry. With some effort one can show that the other axioms of Euclidean geometry do hold in this geometry (see e.g. Greenberg [**16**]) and hence that the parallel postulate is independent of the other axioms of Euclidean geometry.

The third basic property of inversion that we consider is its conformality. Let C_1 and C_2 be two differentiable curves meeting at a point P with tangent lines at P. (Recall that if a plane curve is given parametrically by $x = x(t)$, $y = y(t)$ with not both $x'(t_0)$ and $y'(t_0)$ equal to zero, then the curve has a *tangent vector* or *velocity vector* at $(x(t_0), y(t_0))$, namely $x'(t_0)\mathbf{i}+y'(t_0)\mathbf{j}$ in classical vector notation.) By the *angle* between two curves we mean the undirected angle between their tangent vectors at P. A transformation T mapping a subset of the plane into the plane is said to be *conformal* at P if it preserves the angle between any two curves at P. T is said to be *conformal* if it is conformal at each point of its domain. Some authors require that the sense of angle be preserved as well as the magnitude, but here we define conformality in the wider sense; in fact, inversion reverses the sense of angles.

Our proof of conformality will use a formula found in many calculus texts, but since it is often omitted in first year courses, we briefly derive it here. Let (ρ, θ) be polar coordinates in the plane, and consider a differentiable curve $\rho = f(\theta)$. Let α be the angle of inclination of the tangent lines and $\psi = \alpha - \theta$ (Figure 1.6); then

$$\cot \psi = \frac{1}{\rho}\frac{d\rho}{d\theta} \text{ or } \frac{f'(\theta)}{f(\theta)}.$$

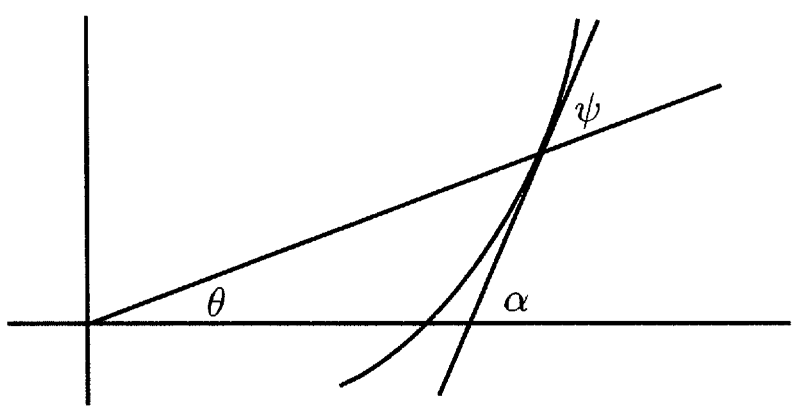

Figure 1.6

In cartesian coordinates the curve is given as $x = \rho\cos\theta$, $y = \rho\sin\theta$. Then

$$\tan\alpha = \frac{dy}{dx} = \frac{dy/d\theta}{dx/d\theta} = \frac{(d\rho/d\theta)\sin\theta + \rho\cos\theta}{(d\rho/d\theta)\cos\theta - \rho\sin\theta}.$$

Substituting this into

$$\tan\psi = \frac{\tan\alpha - \tan\theta}{1 + \tan\alpha\tan\theta}$$

and simplifying gives the desired formula.

Theorem 1.3. *Inversion in a circle is a conformal map.*

Proof. Let (ρ, θ) be polar coordinates in the plane with the origin at the center of inversion and let r be the radius of the circle of inversion. Suppose that $\rho = f_i(\theta)$, $i = 1, 2$, are two differentiable curves meeting at a point P. Let $\rho = g_i(\theta)$, $i = 1, 2$, be the images of the two curves under inversion; then $g_i(\theta) = \frac{r^2}{f_i(\theta)}$. Let ψ_i and ϕ_i denote the undirected angle between the ray corresponding to θ and tangent to $\rho = f_i(\theta)$ and $\rho = g_i(\theta)$ respectively. Let $\beta = \psi_2 - \psi_1$ and $\beta' = \phi_2 - \phi_1$ at P; we shall show that $\beta = \beta'$ to within sign. Since $g_i(\theta) = r^2/f_i(\theta)$, we get

$$g_i'(\theta) = -\frac{r^2 f_i'(\theta)}{(f_i(\theta))^2},$$

and hence

$$\cot\phi_i = \frac{g_i'(\theta)}{g_i(\theta)} = -\frac{f_i'(\theta)}{f_i(\theta)} = -\cot\psi_i.$$

Therefore

$$\cot\beta' = \frac{\cot\phi_2\cot\phi_1 + 1}{\cot\phi_1 - \cot\phi_2} = -\cot\beta = \cot(-\beta).$$

□

Exercises

1. Let O be a point on a circle with center C and suppose the inverse of this circle with respect to O as center of inversion intersects $\overrightarrow{OC}$ at the point A'. If C' is the inverse of C, show that $\overline{OA'} = \overline{A'C'}$.

2. Find the equation of the circle that is the inverse of the line $ax + by = c, c \neq 0$, under inversion in the circle $x^2 + y^2 = 1$.

3. Let P, P' and Q, Q' be two pairs of points, inverse with respect to a circle $\mathcal{C}$. Show that a circle passing through three of these points passes through the fourth.

4. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two circles intersecting in two points P and Q. If $\mathcal{C}_1$ and $\mathcal{C}_2$ are both orthogonal to a third circle $\mathcal{C}_3$ with center O, show that O, P and Q are collinear.

5. Given three collinear points O, P, P' with O not between P and P', construct a circle centered at O with respect to which P and P' are inverse points.

6. Let P and Q be inverse points with respect to a circle $\mathcal{A}$. Prove that inversion in a circle with center $O \neq P, Q$ nor on $\mathcal{A}$ maps P and Q to points P' and Q' which are inverse with respect to the image circle $\mathcal{A}'$. In particular, inversion is an inversive invariant. Hint: Consider two circles $\mathcal{B}$ and $\mathcal{D}$ passing through both P and Q but neither passing through O.

7. Discuss the meaning of Exercise 6 when O is on $\mathcal{A}$ and when O is the point P or Q.

8. Show that the inverse of the center of a circle $\mathcal{A}$ orthogonal to circle of inversion $\mathcal{C}$ is the midpoint of the common chord. More generally, show that the inverse of the center A of a circle $\mathcal{A}$ not through the center of inversion O is the inverse of O in the circle which is the inverse of $\mathcal{A}$. (If O is exterior to $\mathcal{A}$ this is fairly easy using a tangent from O to $\mathcal{A}$ and its inverse. For a clever proof, use Exercise 7. Another proof can be given using the ideas of the next section.)

9. Let $\mathcal{C}$ be the circle $x^2 + y^2 = 1$ in the xy-plane; find the equation of the circle through $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$ orthogonal to $\mathcal{C}$.

10. Given circle $\mathcal{C}$ with center O, point $P \neq O$ interior to $\mathcal{C}$, and line l through P but not through O, construct the circle through P, tangent to l and orthogonal to $\mathcal{C}$.

11. Compass Construction of the Inverse: Given a circle $\mathcal{C}$ with center O and point P exterior to $\mathcal{C}$, draw the circle centered at P and passing through O meeting $\mathcal{C}$ at R and S. Draw circles centered at R and S and passing through O; let P' be the other point of intersection of these circles. Prove that P' is the inverse of P in the circle $\mathcal{C}$.

1.2. Cross ratio

Let $\vec{AB}$ denote the *directed distance* from A to B along a line l; that is, we designate a positive direction or orientation on l, and $\vec{AB} = \overline{AB}$ if the ray with initial point A containing B has the positive direction of the orientation and $\vec{AB} = -\overline{AB}$ if the ray has the opposite direction. Clearly $\vec{AB} = -\vec{BA}$.

Lemma 1.1. *If A, B and C are collinear, then* $\vec{AB} + \vec{BC} + \vec{CA} = 0$.

Proof. The proof is by cases. If C is between A and B, then $\vec{AB} = \vec{AC} + \vec{CB}$ or $\vec{AB} - \vec{AC} - \vec{CB} = 0$. Therefore $\vec{AB} + \vec{BC} + \vec{CA} = 0$. The proofs of the other cases are similar. □

Lemma 1.2. *Let AB be a segment of a line l and O any point of l. Then* $\vec{AB} = \vec{OB} - \vec{OA}$.

Proof. $\vec{AB} + \vec{BO} + \vec{OA} = 0$ by Lemma 1.1, and hence $\vec{AB} = \vec{OB} - \vec{OA}$. □

Let AB be a segment of a line l and $P \in l$. P is said to *divide* AB in the ratio $\vec{AP}/\vec{PB}$. This ratio has several basic properties, which we now present.

(1) The ratio is independent of the orientation of l.

(2) The ratio is positive if P is between A and B, and negative if P is exterior to AB.

(3) If $\vec{AP}/\vec{PB} = \vec{AP'}/\vec{P'B}$, then $P = P'$.
Proof. We have

$$\frac{\vec{AP} + \vec{PB}}{\vec{PB}} = \frac{\vec{AP'} + \vec{P'B}}{\vec{P'B}},$$

so by Lemma 1.1

$$\frac{\vec{AB}}{\vec{PB}} = \frac{\vec{AB}}{\vec{P'B}}.$$

Therefore $\vec{PB} = \vec{P'B}$ or $\vec{BP} = \vec{BP'}$, and hence $P = P'$. □

(4) If $r \neq -1$, there exists a point P such that $\vec{AP}/\vec{PB} = r$.
Proof. Consider the equation

$$r = \frac{\vec{AP}}{\vec{AB} - \vec{AP}};$$

solving gives $\vec{AP} = \frac{r}{1+r}\vec{AB}$. Then, given $r \neq -1$, we can find the point P. □

(5) $\displaystyle\lim_{P\to\infty} \frac{\vec{AP}}{\vec{PB}} = -1.$

Proof. Indeed,

$$r = \frac{\vec{AP}}{\vec{AB} - \vec{AP}} = \frac{1}{\frac{\vec{AB}}{\vec{AP}} - 1} \longrightarrow -1.$$

□

Suppose now that A, B, C and D are four distinct points on an oriented line l; we define their *cross ratio* (AB, CD) by

$$(AB, CD) = \frac{\vec{AC}/\vec{CB}}{\vec{AD}/\vec{DB}}.$$

Note that the cross ratio is positive if both C and D are between A and B or if neither C nor D is between A and B, whereas the cross ratio is negative if the pairs $\{A, B\}$ and $\{C, D\}$ separate each other.

Given three distinct points A, B and C on l and a real number $\mu \neq 0, 1, -\frac{\vec{AC}}{\vec{CB}}$, let D be the unique point dividing the segment AB in the ratio

$$\frac{1}{\mu}\frac{\vec{AC}}{\vec{CB}};$$

thus there exists a unique fourth point D such that $(AB, CD) = \mu$.

We say that four points on a line, A, B, C and D form a *harmonic set* (Figure 1.7) if

$$(AB, CD) = -1.$$

We denote a harmonic set of points by $H(AB, CD)$ and we say that C and D are *harmonic conjugates* with respect to A and B.

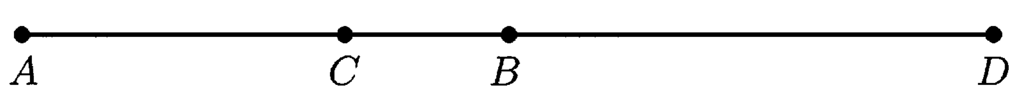

Figure 1.7

In Exercise 3 below one sees that $(AB, DC) = \frac{1}{(AB,CD)}$ and hence that the notion of harmonic conjugate is well defined: If D is the harmonic conjugate of C with respect to AB, C is the harmonic conjugate of D with respect to AB.

If $H(AB, CD)$, then the lengths of segments $\overline{AC}$, $\overline{AB}$, $\overline{AD}$ in this order form a harmonic progression. For if $(AB, CD) = -1$, then

$\frac{\vec{CB}}{\vec{AC}} = \frac{\vec{BD}}{\vec{AD}}$; but $\vec{CB} = \vec{AB} - \vec{AC}$ and $\vec{BD} = \vec{AD} - \vec{AB}$, and hence

$$\frac{\vec{AB} - \vec{AC}}{\vec{AB} \cdot \vec{AC}} = \frac{\vec{AD} - \vec{AB}}{\vec{AD} \cdot \vec{AB}}.$$

Thus

$$\frac{1}{\vec{AB}} = \frac{\frac{1}{\vec{AC}} + \frac{1}{\vec{AD}}}{2};$$

that is, $\frac{1}{\vec{AB}}$ is the arithemetic mean of $\frac{1}{\vec{AC}}$ and $\frac{1}{\vec{AD}}$.

Theorem 1.4. *Let $\mathcal{C}$ be a circle with center O, and C and D a pair of points inverse with respect to $\mathcal{C}$. Let A and B be the endpoints of the diameter through C and D. Then $(AB, CD) = -1$. Conversely, if A and B are the endpoints of a diameter and $(AB, CD) = -1$, then C and D are inverse points.*

Proof. $\frac{\vec{AC}}{\vec{CB}} = -\frac{\vec{AD}}{\vec{DB}}$ is equivalent to

$$\frac{\vec{OC} - \vec{OA}}{\vec{OB} - \vec{OC}} = -\frac{\vec{OD} - \vec{OA}}{\vec{OB} - \vec{OD}};$$

but $\vec{OA} = -\vec{OB}$, so that

$$(\vec{OC} + \vec{OB})(\vec{OB} - \vec{OD}) = -(\vec{OD} + \vec{OB})(\vec{OB} - \vec{OC})$$

or $\vec{OC} \cdot \vec{OD} = \vec{OB}^2$. □

The following lemma is used here and in later applications.

Lemma 1.3. *Let $\mathcal{C}$ be a circle of inversion with center O and radius r. If P, P' and Q, Q' are pairs of inverse points, then*

$$\overline{P'Q'} = r^2 \frac{\overline{PQ}}{\overline{OP} \cdot \overline{OQ}}.$$

Proof. We give the proof here in the case when O, P and Q are collinear; the non-collinear case is left to the reader in Exercise 1 below. $\vec{OP} \cdot \vec{OP'} = \vec{OQ} \cdot \vec{OQ'}$ but $\vec{OP} = \vec{OQ} + \vec{QP}$ and $\vec{OQ'} = \vec{OP'} + \vec{P'Q'}$, giving $\vec{QP} \cdot \vec{OP'} = \vec{P'Q'} \cdot \vec{OQ}$. Therefore

$$\overline{P'Q'} = \frac{\overline{QP} \cdot \overline{OP'} \cdot \overline{OP}}{\overline{OP} \cdot \overline{OQ}} = r^2 \frac{\overline{PQ}}{\overline{OP} \cdot \overline{OQ}}.$$

□

Theorem 1.5. *Let A, B, C and D be four points collinear with the center of inversion O. Let A', B', C' and D' be their respective inverse points. Then $(AB, CD) = (A'B', C'D')$.*

Proof. First note that inversion preserves the separation or non-separation of the pairs A, B and C, D, and hence it suffices to show that $|(A'B', C'D')| = |(AB, CD)|$. But this follows from Lemma 1.3:

$$\frac{\overline{A'C'} \cdot \overline{D'B'}}{\overline{C'B'} \cdot \overline{A'D'}} = \frac{r^2 \frac{\overline{AC}}{\overline{OA} \cdot \overline{OC}} r^2 \frac{\overline{DB}}{\overline{OD} \cdot \overline{OB}}}{r^2 \frac{\overline{CB}}{\overline{OC} \cdot \overline{OB}} r^2 \frac{\overline{AD}}{\overline{OA} \cdot \overline{OD}}} = \frac{\overline{AC} \cdot \overline{DB}}{\overline{CB} \cdot \overline{AD}}.$$

□

Exercises

1. Prove Lemma 1.3 in the case when O, P and Q are not collinear.

2. Let A, B, C and D be four concircular points and define their *cyclic cross ratio* (AB, CD) by

$$(AB, CD) = \pm \frac{\overline{AC} \cdot \overline{DB}}{\overline{CB} \cdot \overline{AD}},$$

choosing the sign according as the pair C, D does not or does separate the pair A, B. Let A, B, C and D be four points on a line not through the center of inversion. Show that their cross ratio is equal to the cyclic cross ratio of their inverse points. Similarly, if A, B, C and D lie on a circle not through the center of inversion, show that inversion preserves the cyclic cross ratio.

3. Let A, B, C and D be four collinear points with $(AB, CD) = \lambda$. Show that the cross ratios of these points in the 24 different orders are equal by fours to the six numbers λ, $\frac{1}{\lambda}$, $1 - \lambda$, $\frac{1}{1-\lambda}$, $\frac{\lambda}{\lambda-1}$, $\frac{\lambda-1}{\lambda}$. Note trivially that $(AB, CD) = (BA, DC) = (CD, AB) = (DC, BA)$.

4. Let P and P' be two fixed points. Consider the locus of points Q such that the ratio $\overline{QP}/\overline{QP'}$ is a constant λ. Clearly when $\lambda = 1$, Q lies on the perpendicular bisector of PP'. When $\lambda \neq 1$, show that Q lies on a circle with respect to which P and P' are inverse points. Draw the figure for $\lambda = \frac{1}{3}, \frac{1}{2}, 1, 2, 3$.

1.3. Applications

We conclude our discussion of planar inversion with some of the classical applications of inversion theory. For more on such applications see e.g. [**11**]. The idea is that a theorem or configuration which is somewhat awkward may invert to a more manageable one. For example, consider the following theorem of Pappus.

Theorem 1.6. *Let $\mathcal{C}$ be a semicircle with diameter AB, and $\mathcal{C}'$ and $\mathcal{C}_0$ semicircles on the same side of AB with diameters AC and CB respectively (Figure 1.8). Let $\mathcal{C}_1, \mathcal{C}_2, \ldots$ be a sequence of circles tangent to $\mathcal{C}$ and $\mathcal{C}'$ and such that $\mathcal{C}_n$ is tangent to $\mathcal{C}_{n-1}$. Let r_n be the radius of $\mathcal{C}_n$ and d_n the distance of the center of $\mathcal{C}_n$ from AB. Then $d_n = 2nr_n$.*

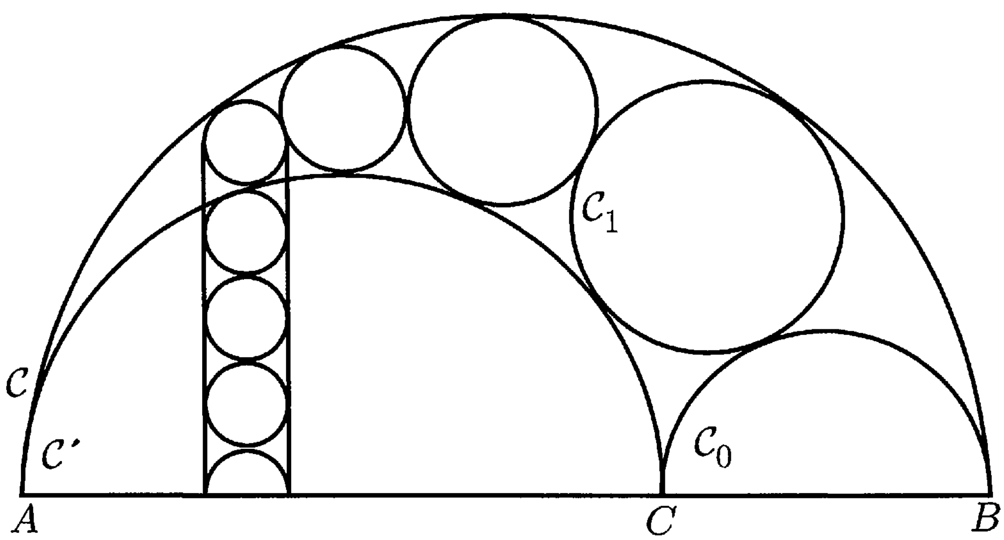

Figure 1.8

Proof. Let a_n be the length of the tangent to $\mathcal{C}_n$ from A, and invert the figure in the circle $\mathcal{A}_n$ with center A and radius a_n. Then $\mathcal{C}_n$ inverts to itself. On the other hand, $\mathcal{C}$ and $\mathcal{C}'$ pass through A and are orthogonal to AB; thus they invert to a pair of parallel lines perpendicular to AB. Since $\mathcal{C}_n$ inverts to itself and is tangent to $\mathcal{C}$ and $\mathcal{C}'$, $\mathcal{C}_n$ is tangent to these parallel lines. Finally, $\mathcal{C}_0, \ldots, \mathcal{C}_{n-1}$ will also invert to circles tangent to the parallel lines, and $d_n = 2nr_n$ follows immediately. □

Consider a circle of inversion $\mathcal{C}$ with center O, and let P and P' be a pair of inverse points. Lines through P' invert to a family of circles through O and P including one line corresponding to the line OP. Circles concentric at P' are orthogonal to the radial lines from P' and hence, by the conformality, invert to a family of circles orthogonal to the first family including one line l corresponding to the circle centered at P' and passing through O (Figure 1.9). Note that l is the perpendicular bisector of the line segment OP (cf. Exercise 1 at the end of Section 1.1). If R is any point on l, R is the center of a circle belonging to the first family. Thus tangents from R to the circles of the second family are equal, and l is called the *radical axis* of the second family.

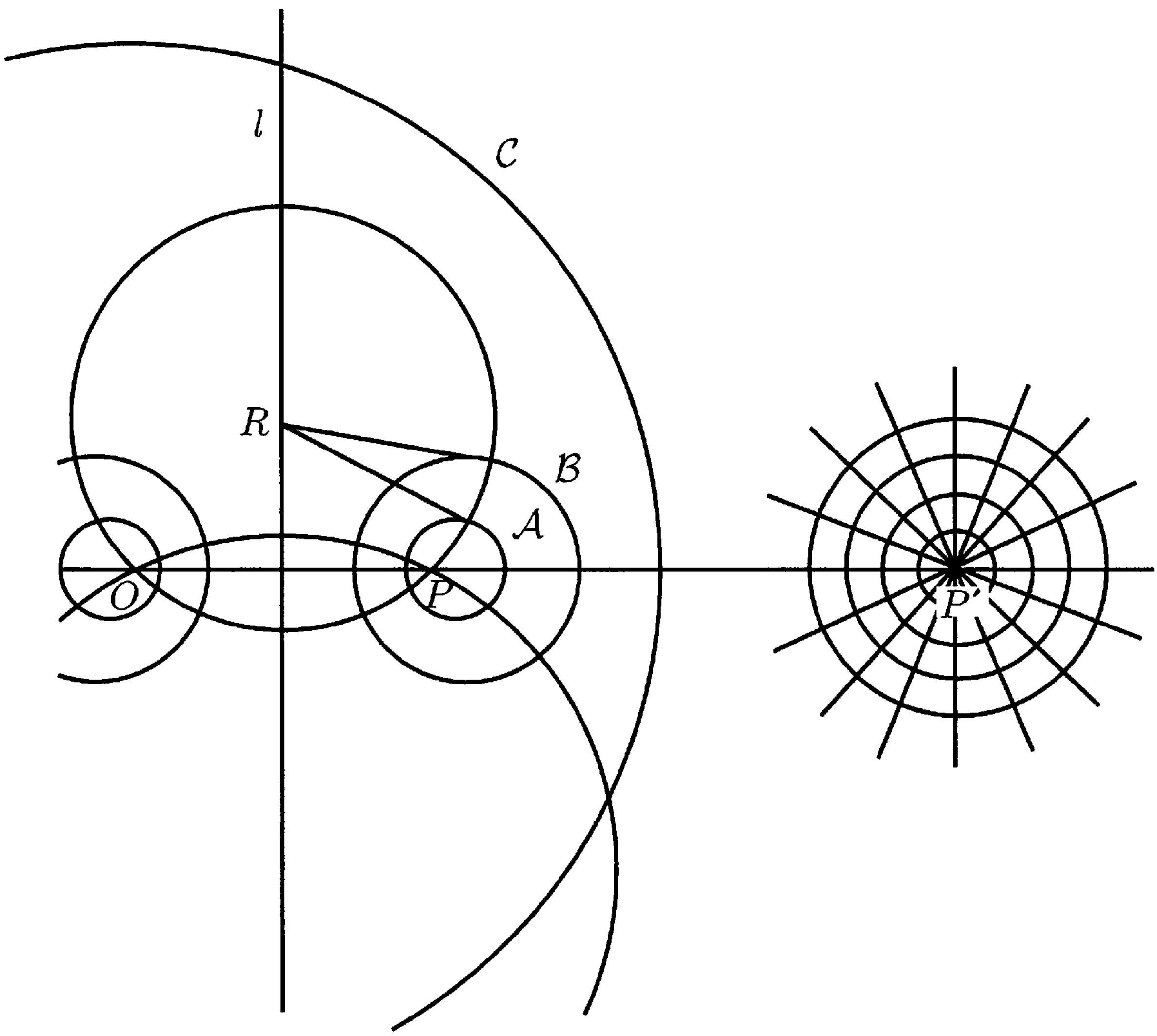

Figure 1.9

Given two non-concentric, non-intersecting circles, we can find their radical axis and hence the rest of the family. Let $\mathcal{A}$ and $\mathcal{B}$ be the circles with centers A and B, and let $\mathcal{C}$ be a circle meeting $\mathcal{A}$ and

$\mathcal{B}$ at points A', A'' and B', B'' respectively. If the lines $A'A''$ and $B'B''$ are not parallel, they meet at a point on the radical axis. Thus the radical axis is the perpendicular through this point to the line of centers AB.

Theorem 1.7. *Two non-concentric, non-intersecting circles can be inverted into two concentric circles.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be the circles as above and l their radical axis. Choosing R on l, draw the circle with center R and radius equal to the tangent length from R to $\mathcal{A}$ and $\mathcal{B}$. This circle cuts AB in points O and P. Using O or P as center of inversion, $\mathcal{A}$ and $\mathcal{B}$ invert to concentric circles. □

The following theorem of Steiner is an immediate corollary of this result.

Theorem 1.8. *Let $\mathcal{A}$ be a circle lying in the interior of a circle $\mathcal{B}$. Suppose there exists a sequence of n circles $\mathcal{C}_1, ..., \mathcal{C}_n$ in the region between $\mathcal{A}$ and $\mathcal{B}$, tangent to both $\mathcal{A}$ and $\mathcal{B}$ with $\mathcal{C}_i$ tangent to $\mathcal{C}_{i-1}$ and with $\mathcal{C}_n$ tangent to $\mathcal{C}_1$. Then there exist infinitely many such sequences, and any circle between $\mathcal{A}$ and $\mathcal{B}$ and tangent to $\mathcal{A}$ and $\mathcal{B}$ belongs to such a sequence.*

Another simple application of inversion theory is the following theorem of Ptolemy.

Theorem 1.9. *Let $ABCD$ be a convex quadrilateral inscribed in a circle. Then the product of the diagonals is equal to the sum of the products of the two pairs of opposite sides.*

Proof. Invert the configuration in the circle with center A and radius r and let B', C' and D' be the inverse points of B, C and D respectively. Then B', C' and D' are collinear, and since $ABCD$ is convex, C' is between B' and D'; and therefore $\overline{B'C'} + \overline{C'D'} = \overline{B'D'}$. Using Lemma 1.3, we get

$$r^2 \frac{\overline{BC}}{\overline{AB} \cdot \overline{AC}} + r^2 \frac{\overline{CD}}{\overline{AC} \cdot \overline{AD}} = r^2 \frac{\overline{BD}}{\overline{AB} \cdot \overline{AD}},$$

and so $\overline{BC} \cdot \overline{AD} + \overline{AB} \cdot \overline{CD} = \overline{BD} \cdot \overline{AC}$. □

Exercises

1. Prove the following generalization of Ptolemy's theorem. In a convex quadrilateral $ABCD$ we have

$$\overline{BC} \cdot \overline{AD} + \overline{AB} \cdot \overline{CD} \geq \overline{BD} \cdot \overline{AC},$$

with equality holding if and only if the quadrilateral can be inscribed in a circle.

2. Given two circles $\mathcal{C}_1$ and $\mathcal{C}_2$ intersecting at two points A and B, let C_1 and C_2 be the other points of intersection of $\mathcal{C}_1$ and $\mathcal{C}_2$ with the diameters of $\mathcal{C}_2$ and $\mathcal{C}_1$ through B respectively. Show that A and B are collinear with the center of the circle through B, C_1 and C_2.

3. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two circles intersecting at A and B, and let s and t be lines tangent to $\mathcal{C}_1$ and $\mathcal{C}_2$ at S_1, S_2 and T_1, T_2 respectively. Show that the circles through S_1, S_2 and A and T_1, T_2 and A are tangent.

4. Let L be an interior point of a circle $\mathcal{C}$ other than the center O. Prove that the tangents to $\mathcal{C}$ at the endpoints of chords through L intersect on a line l.

1.4. Miquel's Theorem

Our next application is a theorem of A. Miquel [**25**]. This theorem is deserving of its own section; not only is it a wonderful application of inversion, but from the axiomatic or foundational point of view of inversive geometry, it plays an important role and is taken as an axiom. As the axiomatic point of view is not an integral part of our discussion, we will only present the axioms and a couple of exercises at the end of this section. For more detail the interested reader is referred to Ewald's book [**12**]. We begin with another theorem of Miquel [**24**], referring to it as Miquel's "Little" Theorem, and then turn to Miquel's "Big" Theorem [**25**].

Theorem 1.10. *In $\triangle ABC$ let D, E, F be points on the sides opposite A, B, C respectively. Then the circles AEF, BDF and CDE pass through a common point M.*

Proof. Suppose BDF and CDE intersect again at M. We will give the proof when M lies within the triangle (Figure 1.10) and invite the reader to supply the proof when M is outside the triangle. The theorem is also true for the case $M = D$ on the triangle, the circles BDF and CDE being tangent at D; again we invite the reader to supply the proof. Denote the supplement of an angle by $s(\angle \ldots)$. Join M to D, E and F and recall that a quadrilateral can be inscribed in a circle if and only if its opposite angles are supplementary. The proof then is

$$\begin{aligned}\angle MFA = s(\angle MFB) &= \angle MDB = s(\angle MDC)\\ &= \angle MEC = s(\angle MEA),\end{aligned}$$

and we conclude that the vertices of the quadrilateral $AFME$ lie on a circle. □

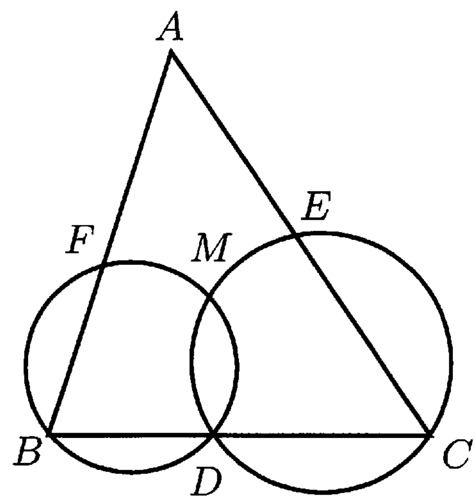

Figure 1.10

We now prove Miquel's "Big" Theorem (Figure 1.11).

Theorem 1.11. *Let $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$, $\mathcal{C}_4$ be four circles, no three with a point in common. Suppose $\mathcal{C}_1$ and $\mathcal{C}_2$ intersect at P and P'; $\mathcal{C}_2$ and $\mathcal{C}_3$ intersect at Q and Q'; $\mathcal{C}_3$ and $\mathcal{C}_4$ intersect at R and R'; $\mathcal{C}_4$ and $\mathcal{C}_1$ intersect at S and S'. Then P, Q, R and S are concircular or collinear if and only if the same is true of P', Q', R' and S'.*

Proof. Suppose P, Q, R and S lie on circle or line $\mathcal{A}$, and invert the configuration with S as center of inversion. Then $\mathcal{C}_1$, $\mathcal{C}_4$ and $\mathcal{A}$ invert to lines forming $\triangle PRS'$, where we use the same letters to denote the points, etc., after the inversion; C_2 and C_3 invert to circles. Miquel's "Little" Theorem now applies to $\triangle PRS'$, and inverting back gives the result. □

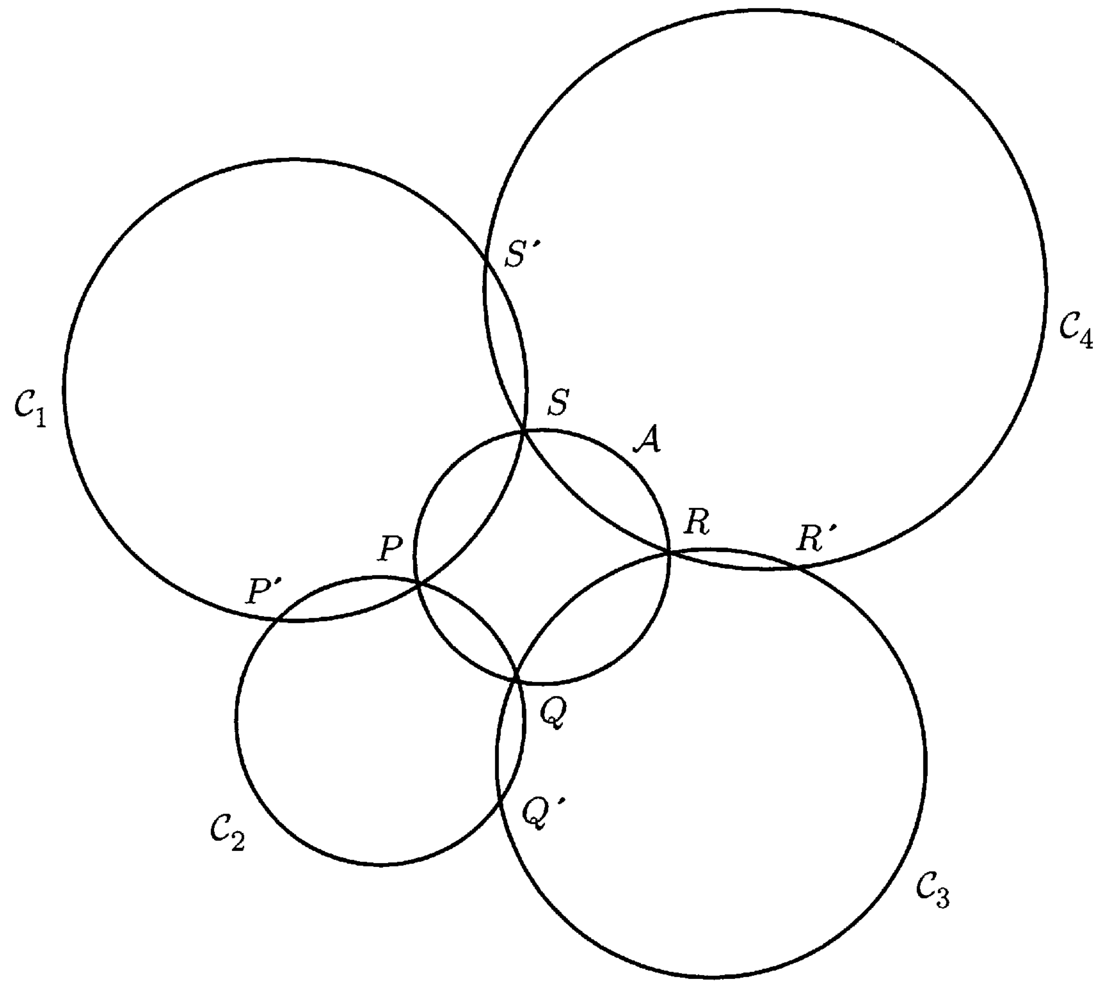

Figure 1.11

We remark that the theorem also holds if one pair of the given circles are tangent, e.g. P and P' coincide and the other pairs are distinct. Miquel's Theorem also contains the following slight generalization.

Theorem 1.12. *Let C_1, C_2, C_3, C_4 be four lines and circles, no three with a point in common. Suppose succesive pairs meet in points P and P', etc. as before, where for two succesive lines the intersection points are the Euclidean point of intersection and the ideal point at infinity (which lies on all lines, see Chapter 2). Then the conclusion of Miquel's Theorem holds.*

Proof. Invert the configuration with respect to some point not on the configuration, giving the original Miquelean configuration. □

The incidence axioms of an *inversive plane* are as follows, *point* and *circle* being undefined terms.

A.1: Three distinct points lie on one and only one circle.

A.2: If P is a point on circle $\mathcal{A}$ and Q a point not on $\mathcal{A}$, there exists a unique circle through P and Q having only P in common with $\mathcal{A}$ (*tangent* to $\mathcal{A}$).

A.3: On each circle there exist at least three points.

A.4: There exist a point and a circle that are not incident.

An inversive plane is said to be *odd* (see [**12**]) if it satisfies

A.5: There exist four circles each two of which are tangent and such that the six points of tangency are distinct; no other circle through one of these points is tangent to three of the circles.

The final axiom is:

A.6: Miquel's "Big" Theorem.

With this axiom the inversive plane is said to be *Miquelian.* Miquel's "Big" Theorem is a closure theorem for inversive geometry and plays the role in the axiomatic foundation similar to the role played by Pappus' Theorem in the axiomatic foundation of projective geometry.

EXERCISES

1. Prove the "Bundle Theorem": Let $P, P', Q, Q', R, R', S, S'$ be distinct points not lying on a common circle. If

$$\{P, P', Q, Q'\},\ \{Q, Q', R, R'\},\ \{R, R', S, S'\},$$
$$\{P, P', R, R'\},\ \{Q, Q', S, S'\}$$

are concircular quadruples of points, then so is $\{P, P', S, S'\}$. Hint: Invert with one of the eight points as center of inversion. We remark that Miquel's Theorem is stronger than the Bundle Theorem; for an example of a non-Miquelian inversive plane in which the Bundle Theorem holds, see [**12**].

2. Consider a geometry whose "points" are the points $(x, y) \in \mathbb{R}^2$ together with an ideal point ∞ belonging to all lines and whose "circles" are the lines in $\mathbb{R}^2$ and the family of curves of the form $(x-h)^4 + (y-k)^4 = a^4$. Show that this is a non-Miquelian inversive plane, i.e. axioms A.1-4 hold but Miquel's theorem is false. One might approach this problem more generally. Consider a smooth closed convex curve $\mathcal{C}$ whose tangent lines have only one point in common with $\mathcal{C}$. Define a geometry whose "circles" are either lines or images of $\mathcal{C}$ under translations or homotheties of positive ratio, and show that axioms A.1-4 hold. (This is not a routine exercise, and could be used as a student project.)

3. If a point A is removed from an inversive plane (axioms A.1-4), prove that the circles through A (with A removed) satisfy the axioms of an affine plane: 1. Two distinct points lie on one and only one line. 2. Every line has at least two points. 3. There exist three noncollinear points. 4. Given a line and a point not on it, there exists a unique line through the point parallel the given line.

1.5. Feuerbach's Theorem

One of the most famous applications of classical inversion theory is the theorem of Feuerbach which states that the nine-point circle of a triangle is tangent to the incircle and each of the excircles of the triangle. We begin by describing the nine-point circle of a triangle by means of the following theorem. Recall that the altitudes of a triangle are concurrent at a point called the *orthocenter* of the triangle.

Theorem 1.13. *In a triangle, the midpoints of the sides, the feet of the altitudes and the midpoints of the segments from the orthocenter to the vertices are concircular.*

Proof. Let $\triangle ABC$ be the triangle; A', B', C' the midpoints of the sides opposite the respective vertices; D, E, F the feet of the altitudes from A, B, C; H the orthocenter; and U, V, W the midpoints of the segments AH, BH, CH respectively (Figure 1.12). Consider the circle $\mathcal{C}$ through A', B', C'; we shall show that D, E, F and U, V, W lie on $\mathcal{C}$.

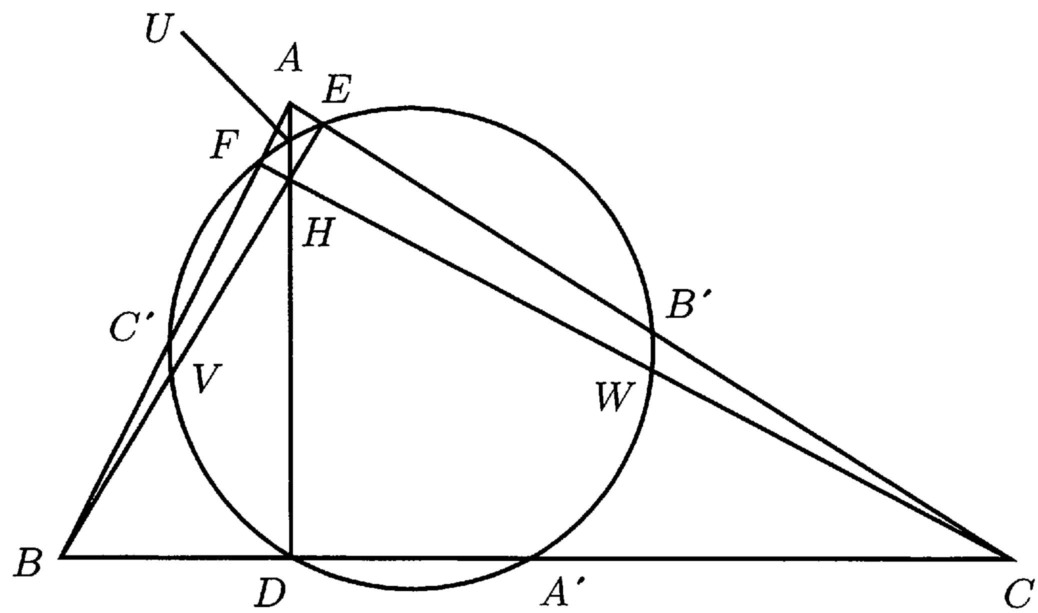

Figure 1.12

First note that in the right triangle $\triangle ABD$, C' is the midpoint of the hypotenuse and hence $C'D \cong C'B$. Also, the quadrilateral $C'B'A'B$ is a parallelogram since in any triangle a line joining the midpoints of two sides is parallel to the third side. Thus $C'B'A'D$ is an isosceles trapezoid; but a trapezoid can be inscribed in a circle if and only if it is isosceles. Therefore the circle $\mathcal{C}$ through A', B' and C' passes through D. In like manner, E and F lie on $\mathcal{C}$.

To show that U, and similarly V and W, lie on $\mathcal{C}$, we first observe that $C'U$ is parallel to BE and therefore perpendicular to $A'C'$. Similarly $B'U$ is perpendicular to $A'B'$. Thus, since $\angle UC'A'$ and $\angle UB'A'$ are right angles, B' and C' lie on the circle with diameter $A'U$, which is therefore $\mathcal{C}$. □

The circle of Theorem 1.13 is called the *nine-point circle* of the triangle.

As is well known, the angle bisectors of a triangle are concurrent at a point which is equidistant from the sides and hence is the center of a circle tangent to the sides called the *incircle*. Also the bisector of one angle and the bisectors of the exterior angles at the other two vertices are concurrent at points equidistant from the sides, giving rise to three *excircles* of the triangle (Figure 1.13).

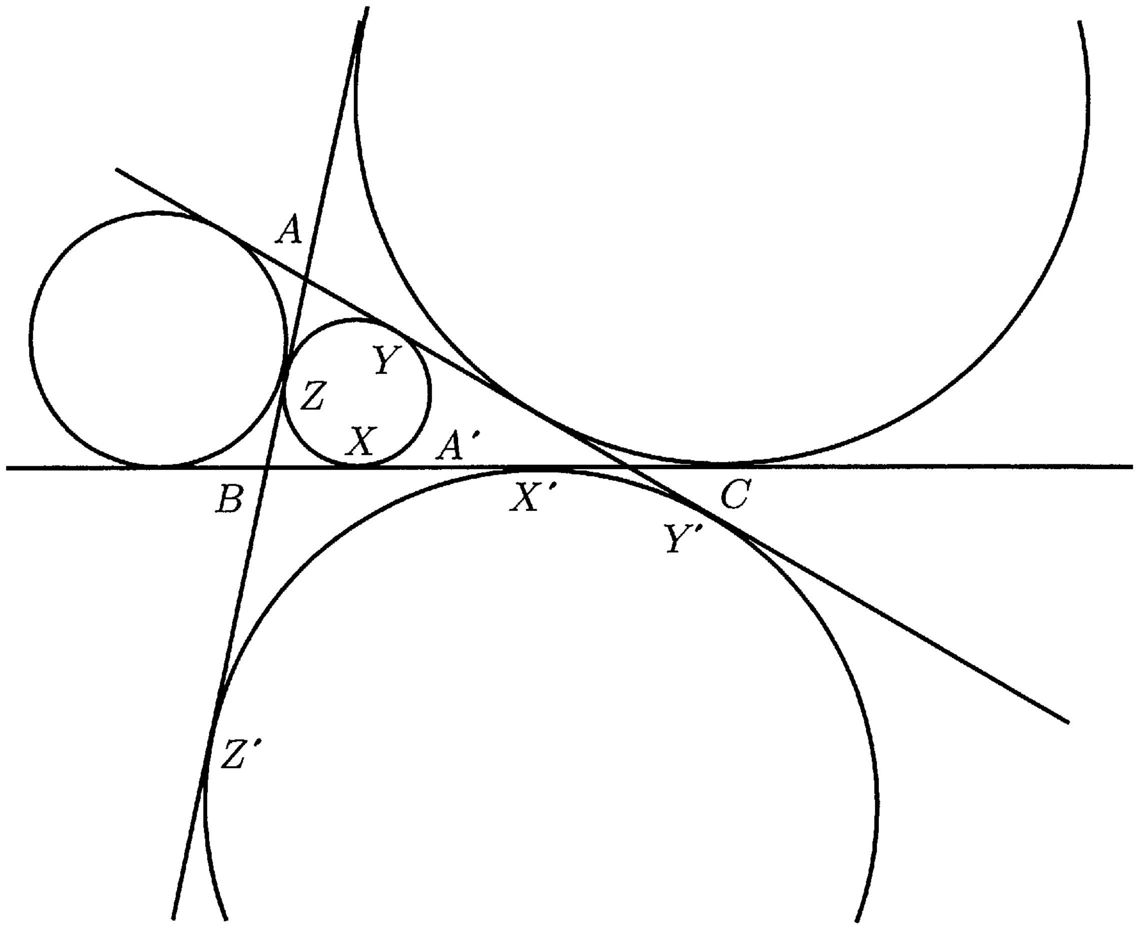

Figure 1.13

In $\triangle ABC$ let X, Y and Z be the point of contact of the incircle with the sides opposite A, B and C respectively, and let X', Y' and Z' be the points of contact of the excircle opposite A with the same respective sides. We now prove, as a lemma, the fact that the midpoint of a side is also the midpoint of the segment determined by these contact points on that side.

Lemma 1.4. $\overline{XA'} = \overline{A'X'}$.

Proof. Since $\overline{BX'} = \overline{BZ'}$ and $\overline{CX'} = \overline{CY'}$, the perimeter of the triangle is $\overline{AZ'} + \overline{AY'} = 2\overline{AY'}$. Now $\overline{BX} = \overline{BZ} = \overline{AB} - \overline{AZ} = \overline{AB} - \overline{AY}$ and $\overline{BX} = \overline{BC} - \overline{CX} = \overline{BC} - \overline{CY}$, and hence $2\overline{BX} = \overline{AB} + \overline{BC} - \overline{AC} = \text{perimeter} - 2\overline{AC}$. We also have $\overline{CX'} = \overline{AY'} - \overline{AC}$ or $2\overline{CX'} = \text{perimeter} - 2\overline{AC}$. Therefore $\overline{BX} = \overline{CX'}$, which gives the result. □

We can now state and prove Feuerbach's Theorem.

Theorem 1.14. *The nine-point circle of a triangle is tangent to the incircle and to each of the three excircles.*

Proof. Let $\triangle ABC$ be the triangle with the special points involved denoted as above. Let I be the center of the incircle $\mathcal{I}$ and E the center of the excircle $\mathcal{E}$ opposite A. $\mathcal{I}$ and $\mathcal{E}$ have common tangents AB, AC, CB and a fourth one RS meeting BC at G, R being on AC and S on AB (Figure 1.14). Note also that A, I, G, E are collinear.

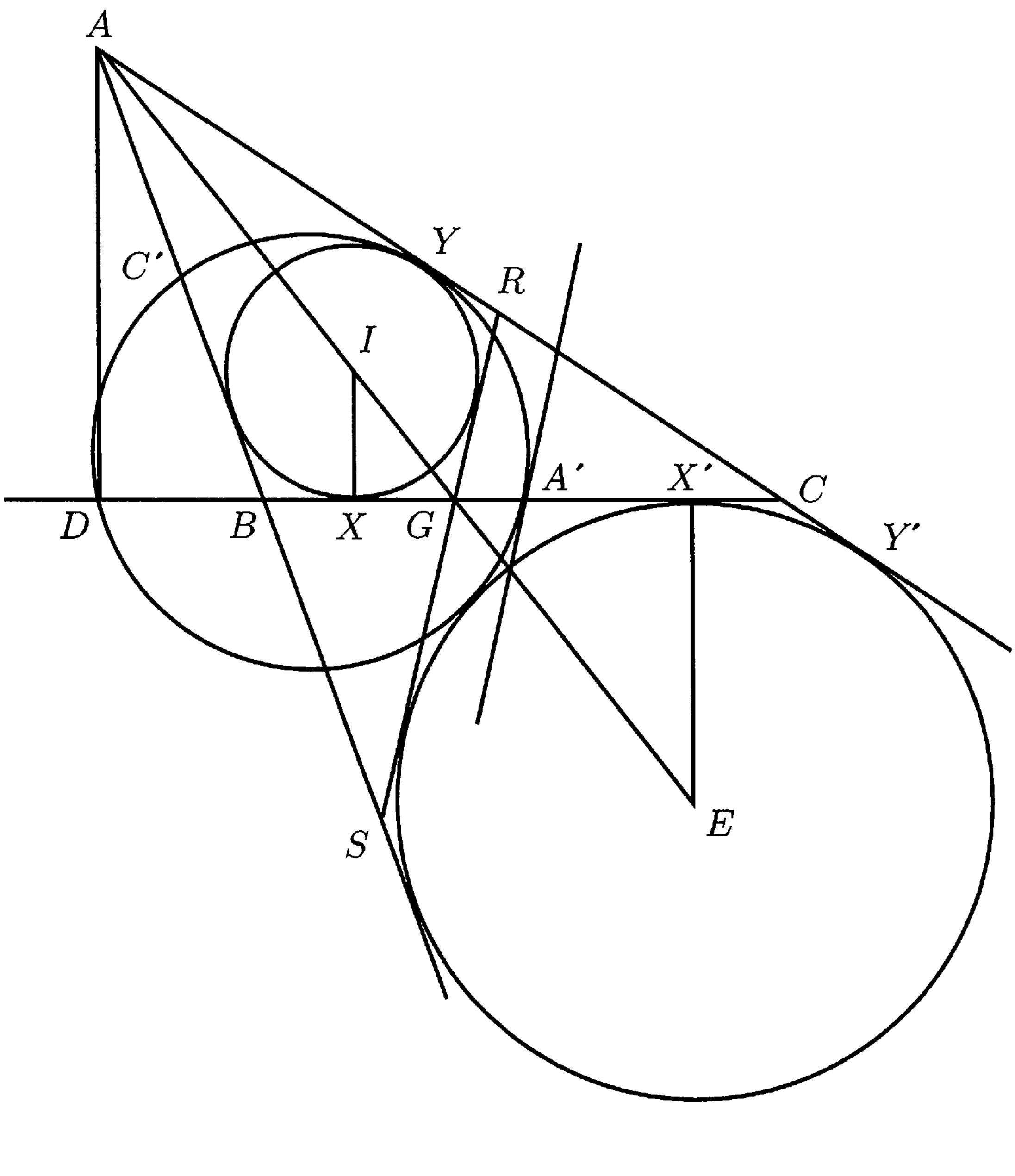

Figure 1.14

Now $\triangle AYI \sim \triangle AY'E$ and $\triangle GXI \sim \triangle GX'E$. Therefore

$$\frac{\overline{AI}}{\overline{AE}} = \frac{\overline{IY}}{\overline{EY'}} = \frac{\overline{IX}}{\overline{EX'}} = \frac{\overline{IG}}{\overline{EG}},$$

and so

$$\frac{\vec{AI}}{\vec{IG}} = -\frac{\vec{AE}}{\vec{EG}}.$$

Thus, since AD, IX and EX' are parallel, we have

$$\frac{\vec{DX}}{\vec{XG}} = -\frac{\vec{DX'}}{\vec{X'G}},$$

i.e. the cross ratio $(DG, XX') = -1$. Therefore $(XX', DG) = -1$.

Now invert in the circle with center A' and radius $\overline{XA'}$, which is equal to $\overline{A'X'}$ by the lemma. $\mathcal{I}$ and $\mathcal{E}$ are orthogonal to the circle of inversion, and hence invert to themselves. The nine-point circle passes through A' and D and hence inverts to a line l through G, the inverse of D since $(XX', DG) = -1$. We shall show that l is the line of RS and hence tangent to $\mathcal{I}$ and $\mathcal{E}$, proving the theorem. To do this it suffices to show that RS is parallel to the tangent to the nine-point circle at A' (the inverse of a circle through the center of inversion A' is a line perpendicular to the diameter of the circle through A'; see Section 1). Now the angle between this tangent and $A'B'$ is congruent to $\angle B'C'A'$, since both these angles have measure equal to half the measure of the arc $\widehat{A'B'}$. Now $\angle B'C'A'$ is congruent to $\angle A'CB'$, since $B'C'A'C$ is a parallelogram. Finally, by reflection across the line of IE we see that $\angle A'CB'$ is congruent to $\angle RSA$. Thus, since AS is parallel to $A'B'$, it follows that RS is parallel to the tangent at A' and hence must be l. □

Exercises

1. Let O be the center of the circumcircle of a triangle and H its orthocenter. Show that the center N of the nine-point circle is the midpoint of OH.

2. Find the center of the nine-point circle of $\triangle A'B'C'$.

Chapter 2

Linear Fractional Transformations

2.1. Complex numbers

We denote by $\mathbb{C}$ the set of all complex numbers,

$$\mathbb{C} = \{x + iy \mid x, y \,\text{real}; i^2 = -1\}.$$

Indentifying complex numbers with ordered pairs of real numbers, (x, y), we also refer to $\mathbb{C}$ as the *complex plane.* The numbers x and y separately are called the *real* and *imaginary parts* of $z = x + iy$, also denoted $\Re z$ and $\Im z$. Recall the addition, multiplication and division of complex numbers:

$$(a + ib) + (c + id) = (a + c) + i(b + d),$$

$$(a + ib)(c + id) = (ac - bd) + i(ad + bc),$$

$$\frac{a + ib}{c + id} = \frac{ac + bd}{c^2 + d^2} + i\frac{bc - ad}{c^2 + d^2}.$$

The complex number $x - iy$ is called the *complex conjugate* of $z = x + iy$ and is denoted by $\bar{z}$. The non-negative number $|z| = \sqrt{x^2 + y^2}$ is called the *modulus* of $z = x + iy$. Note that $z\bar{z} = |z|^2$ and $\frac{1}{z} = \frac{\bar{z}}{|z|^2}$. Geometrically $|z|$ is the distance of z from the origin and $|z_1 - z_2|$ is the distance between two points z_1 and z_2.

An angle θ from the positive real axis to the ray from the origin to a complex number z is called an *argument* of z, and we write $\theta = \arg z$. This is determined only up to multiples of 2π; so when necessary to avoid ambiguity one chooses a *principal determination* of the argument by requiring that $-\pi < \theta \leq \pi$ and writes $\theta = \operatorname{Arg} z$. If $z = x + iy$ and $r = |z|$, then $x = r\cos\theta$, $y = r\sin\theta$ and

$$z = r(\cos\theta + i\sin\theta), \qquad \bar{z} = r(\cos\theta - i\sin\theta).$$

The *complex exponential* e^z is defined by $e^z = e^x(\cos y + i\sin y)$; in particular $e^{iy} = \cos y + i\sin y$. Thus a given complex number z may be written $re^{i\theta}$, explicitly showing its modulus and argument.

For two complex numbers z_1 and z_2 their product is given by

$$\begin{aligned} z_1 z_2 &= r_1 r_2((\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + i(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2)) \\ &= r_1 r_2(\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)). \end{aligned}$$

Therefore $\arg z_1 z_2 = \arg z_1 + \arg z_2$ up to the ambiguity mod 2π. Clearly $|z_1 z_2| = r_1 r_2$, so in Figure 2.1, $\triangle(0,1,z_1) \sim \triangle(0,z_2,z_1z_2)$.

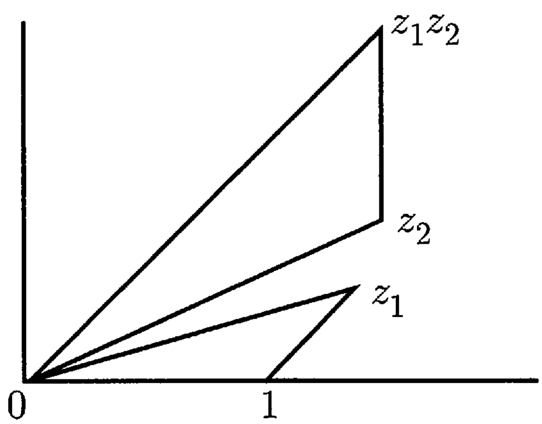

Figure 2.1

Consider now

$$\frac{1}{z} = \frac{\bar{z}}{|z|^2} = \frac{1}{r}(\cos\theta - i\sin\theta);$$

then $|\frac{1}{z}| = \frac{1}{|z|}$ and $\arg\frac{1}{z} = -\arg z$. Note also that $\frac{1}{\bar{z}}$ and z are collinear with the origin and that the product of their distances from the origin is equal to 1. Thus the map $z \longrightarrow \frac{1}{\bar{z}}$ is inversion in the unit circle. In Figure 2.2 $\triangle(0,1,z) \sim \triangle(0,\frac{1}{z},1)$, and therefore $\frac{1}{z}$ can be obtained from z by two geometric transformations, namely, inversion in the unit circle and reflection in the real axis:

$$z \underbrace{\longrightarrow}_{\text{inversion}} \frac{1}{\bar{z}} \underbrace{\longrightarrow}_{\text{reflection}} \frac{1}{z}$$

or

$$z \underbrace{\longrightarrow}_{\text{reflection}} \bar{z} \underbrace{\longrightarrow}_{\text{inversion}} \frac{1}{z}$$

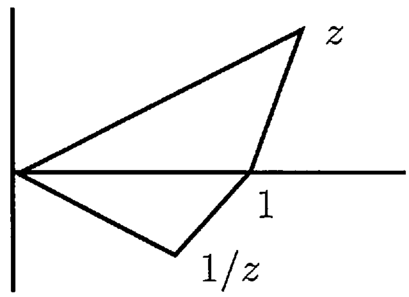

Figure 2.2

2.2. The extended complex plane and stereographic projection

As with inversion, the transformation $w = \frac{1}{z}$ maps the complex plane onto itself with two exceptions: $z = 0$ has no image and $w = 0$ has no preimage. To remove this difficulty we form the *extended complex plane* $\bar{\mathbb{C}}$, sometimes called the *inversive plane*, by adjoining to $\mathbb{C}$ an ideal point, often called the *point at infinity* or the complex number *infinity* and denoted by ∞. Let $a \in \mathbb{C}$; the rules governing ∞ are: $\frac{a}{\infty} = 0$; $\frac{a}{0} = \infty, a \neq 0$; $\infty \pm a = \infty$; $\infty \cdot a = \infty, a \neq 0$; $\infty \cdot \infty = \infty$. The following are left undefined: $\infty \pm \infty$, $\frac{\infty}{\infty}$, $\frac{0}{0}$, $\infty \cdot 0$. For any $R > 0$, the set $\{z \big| |z| > R\}$ is thought of as a neighborhood of ∞; it is mapped onto a neighborhood of the origin $|w| < \frac{1}{R}$ by $w = \frac{1}{z}$.

To visualize $\bar{\mathbb{C}}$ we use a sphere rather than a plane. Consider the sphere $x^2 + y^2 + u^2 = u$ tangent to $\mathbb{C}$ at the origin and of radius $\frac{1}{2}$. The line joining any point P in the plane to $N = (0, 0, 1)$ intersects the sphere at a point Q, and, conversely, the line joining N to any point Q on the sphere meets the plane in a point P (Figure 2.3). Thus the points of the extended complex plane may be identified with the points of a sphere, the ideal point ∞ being identified with the point N. This mapping of $\bar{\mathbb{C}}$ to the sphere is called *stereographic projection*, and generally the inverse mapping of the sphere onto $\bar{\mathbb{C}}$ is also referred to as stereographic projection. The extended complex plane $\bar{\mathbb{C}}$ is sometimes referred to as the *complex sphere* or as the *Riemann sphere.*

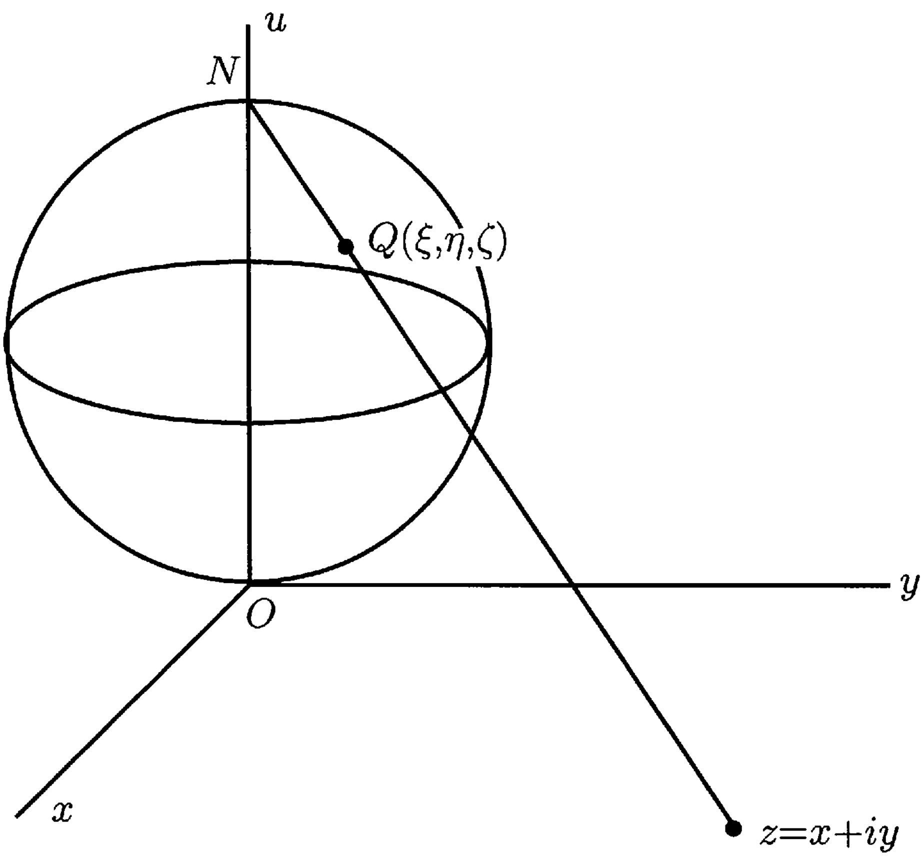

Figure 2.3

Note that the image of the unit circle about the origin O is mapped to the horizontal equator on the sphere. Now in the exercise on page 3 rotate the circle $\mathcal{D}$ about OP so that the plane of $\mathcal{D}$ is perpendicular to the plane of $\mathcal{C}$. Thus we see that inversion in the unit circle is just reflection in the horizontal equator of the complex sphere.

We now obtain a coordinate expression for stereographic projection. Let $z \in \mathbb{C}$ represent P and let (ξ, η, ζ) be coordinates of the point Q and $r = |z|$. Now $\triangle OPN \sim \triangle QON$, and hence

$$\frac{r^2}{1+r^2} = \frac{\xi^2+\eta^2+\zeta^2}{1} = \zeta.$$

Projecting to the xu- and yu-planes, we have

$$\frac{x}{\xi} = \frac{1}{1-\zeta}, \qquad \frac{y}{\eta} = \frac{1}{1-\zeta}.$$

Therefore

$$x = \frac{\xi}{1-\zeta}, \quad y = \frac{\eta}{1-\zeta}, \quad r^2 = \frac{\zeta}{1-\zeta}$$

and

$$\xi = \frac{x}{1+r^2}, \quad \eta = \frac{y}{1+r^2}, \quad \zeta = \frac{r^2}{1+r^2}.$$

Theorem 2.1. *Stereographic projection maps lines and circles in the plane to circles on the sphere, and, conversely, circles on the sphere map stereographically to lines and circles in the plane.*

Proof. A circle on the sphere is the intersection of the sphere with a plane, say $Ax + By + Cu = D$; if (ξ, η, ζ) is on this circle we have

$$A\frac{x}{1+r^2} + B\frac{y}{1+r^2} + C\frac{r^2}{1+r^2} = D$$

or

$$(C - D)(x^2 + y^2) + Ax + By = D, \tag{2.1}$$

which is a line or a circle in the plane according to whether or not $C = D$. Conversely, a line or circle in the plane is given by an equation of the form (2.1). Substitution for x and y in terms of ξ, η, ζ gives

$$(C - D)\frac{\zeta}{1-\zeta} + A\frac{\xi}{1-\zeta} + B\frac{\eta}{1-\zeta} = D \tag{2.2}$$

or

$$A\xi + B\eta + C\zeta = D,$$

so (ξ, η, ζ) lies on a plane determining a circle on the sphere. □

We shall now prove the conformality of stereographic projection; that is, if two differentiable plane curves meet at a point P, the angle between the image curves is equal to the angle between the curves.

Lemma 2.1. *The angle between two lines is equal to the angle between their image circles under stereographic projection.*

Proof. Let $A_1x + B_1y = D_1$ and $A_2x + B_2y = D_2$ be two lines in the plane intersecting at P. The angle between them is determined by their slopes $-\frac{A_1}{B_1}$ and $-\frac{A_2}{B_2}$. By equation (2.2) the images of the lines lie in the planes $A_1x + B_1y = D_1(1-u)$ and $A_2x + B_2y = D_2(1-u)$. Therefore the tangents to the corresponding circles at N are the intersections of these planes with the plane $u = 1$, and hence their equations are $A_1x + B_1y = 0$, $A_2x + B_2y = 0$, $u = 1$, which intersect at the same angle determined by the ratios $-\frac{A_1}{B_1}$ and $-\frac{A_2}{B_2}$. Let Q be the image of point P, and consider reflection in the plane which is the perpendicular bisector of chord QN. Then we have that the angle at Q is equal to the angle at N, giving the result. □

Let (ξ_0, η_0, ζ_0) be a point on the circle $Ax + By + Cu = D$, $x^2 + y^2 + u^2 - u = 0$, and let $\mathbf{i}, \mathbf{j}, \mathbf{k}$ be unit vectors in the directions of the positive x-, y- and u-axes. Recall that the vector $A\mathbf{i} + B\mathbf{j} + C\mathbf{k}$ is normal to the plane and that $\xi_0\mathbf{i} + \eta_0\mathbf{j} + (\zeta_0 - \frac{1}{2})\mathbf{k}$ is normal to the sphere (radial vector from the center $(0, 0, \frac{1}{2})$). Consequently the vector

$$(A\mathbf{i} + B\mathbf{j} + C\mathbf{k}) \times (\xi_0\mathbf{i} + \eta_0\mathbf{j} + (\zeta_0 - \tfrac{1}{2})\mathbf{k})$$

is tangent to the circle at (ξ_0, η_0, ζ_0) giving the following lemma.

Lemma 2.2. *The tangent line to the circle $Ax + By + Cu = D$, $x^2 + y^2 + u^2 - u = 0$ at (ξ_0, η_0, ζ_0) is*

$$\frac{x - \xi_0}{B(\zeta_0 - \frac{1}{2}) - C\eta_0} = \frac{y - \eta_0}{C\xi_0 - A(\zeta_0 - \frac{1}{2})} = \frac{u - \zeta_0}{A\eta_0 - B\xi_0}.$$

Theorem 2.2. *Stereographic projection is a conformal map.*

Proof. In view of Lemma 1.1 it remains only to prove that the image of the tangent line to a differentiable curve is tangent to the image of the curve. Let $x = x(t), y = y(t)$ be a differentiable curve in the plane and P the point corresponding to $t = t_0$, and let $x_0 = x(t_0)$ and $y_0 = y(t_0)$. Then

$$x = \xi(t) = \frac{x(t)}{1 + x(t)^2 + y(t)^2}, \quad y = \eta(t) = \frac{y(t)}{1 + x(t)^2 + y(t)^2},$$

$$u = \zeta(t) = \frac{x(t)^2 + y(t)^2}{1 + x(t)^2 + y(t)^2}$$

is the image of the curve on the sphere. Let $\xi_0 = \xi(t_0)$, $\eta_0 = \eta(t_0)$ and $\zeta_0 = \zeta(t_0)$. The tangent line to the image curve at the image point Q is

$$\frac{x-\xi_0}{\xi'(t_0)} = \frac{y-\eta_0}{\eta'(t_0)} = \frac{u-\zeta_0}{\zeta'(t_0)}. \tag{2.3}$$

By differentiation we have

$$\xi'(t_0) = \frac{x'(t_0)(1-x_0^2+y_0^2) - 2x_0y_0y'(t_0)}{(1+x_0^2+y_0^2)^2},$$

$$\eta'(t_0) = \frac{y'(t_0)(1+x_0^2-y_0^2) - 2x_0y_0x'(t_0)}{(1+x_0^2+y_0^2)^2},$$

$$\zeta'(t_0) = \frac{2x_0x'(t_0) + 2y_0y'(t_0)}{(1+x_0^2+y_0^2)^2},$$

and (2.3) becomes

$$\begin{aligned} &\frac{x-\xi_0}{x'(t_0)(1-x_0^2+y_0^2) - 2x_0y_0y'(t_0)} \\ &\qquad = \frac{y-\eta_0}{y'(t_0)(1+x_0^2-y_0^2) - 2x_0y_0x'(t_0)} \\ &\qquad = \frac{u-\zeta_0}{2(x_0x'(t_0)+y_0y'(t_0))}. \end{aligned} \tag{2.4}$$

Now the tangent line to the original curve at P is

$$y'(t_0)x - x'(t_0)y = y'(t_0)x_0 - x'(t_0)y_0. \tag{2.5}$$

Setting $A = y'(t_0)$, $B = -x'(t_0)$ and $C = D = y'(t_0)x_0 - x'(t_0)y_0$, we see by Lemma 2.2 that the tangent to the image of (2.5) is the same line as equation (2.4), for if we multiply the equation of Lemma 2.2 by $\frac{1}{2(1+r^2)}$, the first denominator becomes

$$\begin{aligned} &2(1+r^2)(B(\zeta_0 - \tfrac{1}{2}) - C\eta_0) \\ &\quad = -x'(t_0)(x_0^2+y_0^2-1) - 2(y'(t_0)x_0 - x'(t_0)y_0)y_0 \\ &\quad = x'(t_0)(1-x_0^2+y_0^2) - 2x_0y_0y'(t_0), \end{aligned}$$

which is the first denominator of (2.4). Proceed similarly for the other denominators. □

EXERCISES

1. Let α, β be complex numbers. Show that $\left|\frac{\alpha - \beta}{1 - \bar{\beta}\alpha}\right| = 1$ if and only if at least one of α and β has modulus 1. Show that if $|\alpha| < 1$ and $|\beta| < 1$, then $\left|\frac{\alpha - \beta}{1 - \bar{\beta}\alpha}\right| < 1$.

2. Let α, β be complex numbers and $\lambda \neq 1$ a positive constant. Show that $\left|\frac{z - \alpha}{z - \beta}\right| = \lambda$ is a circle. What is the locus for $\lambda = 1$? Compare this with Exercise 4 in Section 1.2.

3. Let $z_1, z_2 \in \mathbb{C}$, and let Z_1, Z_2 be their images under stereographic projection. Let $d(Z_1, Z_2)$ be the length of the chord joining Z_1 and Z_2. Show that

$$d(Z_1, Z_2) = \frac{|z_1 - z_2|}{\sqrt{1 + |z_1|^2}\sqrt{1 + |z_2|^2}}.$$

4. Consider the antipodal map of $x^2 + y^2 + u^2 = u$. What is the corresponding map of $\mathbb{C}$ under stereographic projection?

2.3. Linear fractional transformations

We now study in some detail a special class of transformations of the extended complex plane. By a *linear fractional transformation* (sometimes called a *homography*, also sometimes called a *Möbius transformation*) we mean a transformation of the form

$$w = \frac{az + b}{cz + d}, \quad ad - bc \neq 0,$$

where $a, b, c, d \in \mathbb{C}$. If $ad - bc = 0$ and $c \neq 0$, then $w = \frac{az+b}{cz+d}$ reduces to $w = \frac{a}{c}$, as will become apparent below. Let us begin with some simple special cases.

- $a = d = 1, c = 0$, i.e. $w = z + b$: This is just the addition of a fixed vector b to the vector z. Thus each point of the plane undergoes a translation by b, and any figure is mapped to a congruent figure.
- $b = c = 0, d = 1$, i.e. $w = az$: This is first a rotation through $\arg a$ and then a magnification by $|a|$. If $|a| = 1$, then a is of

the form $e^{i\theta} = \cos\theta + i\sin\theta$ and $w = az$ is simply a rotation through the angle θ, which is a congruence. If a is real, this is a homothety. Thus the rotation-stretching $w = az$ maps plane figures to similar ones and so, in particular, is angle preserving.

The special case $a = d = 0, b = c = 1$, viz. $w = \frac{1}{z}$, has already been discussed, so we turn to the general case. If $c = 0$, then $w = \frac{a}{d}z + \frac{b}{d}$, which is a combination of the two special cases above. If $c \neq 0$, then

$$w = -\frac{ad - bc}{c}\frac{1}{cz + d} + \frac{a}{c},$$

and hence the mapping is a rotation-stretching followed by a translation, a reflection in the real axis, an inversion, a rotation-stretching and a translation. Note also that the point $-\frac{d}{c}$ is mapped to ∞ and ∞ is mapped to $\frac{a}{c}$. Thus we have the following theorem.

Theorem 2.3. *Linear fractional transformations map the extended plane $\bar{\mathbb{C}}$ onto itself, map lines and circles to lines and circles, and are angle preserving.*

Now let T_1 and T_2 be two linear fractional transformations given by

$$w = \frac{az + b}{cz + d} \quad \text{and} \quad \zeta = \frac{\alpha w + \beta}{\gamma w + \delta},$$

respectively. Then their composition T_2T_1 is also a linear fractional transformation, for

$$\zeta = \frac{\alpha\frac{az+b}{cz+d} + \beta}{\gamma\frac{az+b}{cz+d} + \delta} = \frac{(\alpha a + \beta c)z + (\alpha b + \beta d)}{(\gamma a + \delta c)z + (\gamma b + \delta d)}.$$

Note that the coefficients can be obtained by multiplying the matrices of the respective transformations:

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \alpha a + \beta c & \alpha b + \beta d \\ \gamma a + \delta c & \gamma b + \delta d \end{pmatrix}.$$

Taking determinants, we see that

$$(\alpha a + \beta c)(\gamma b + \delta d) - (\alpha b + \beta d)(\gamma a + \delta c) \neq 0.$$

Note that if each of a, b, c, d is multiplied by a non-zero complex number, then the matrix is multiplied by this number, but the transformation remains the same.

If T_1, T_2, T_3 are three linear fractional transformations, then

$$T_3(T_2T_1) = (T_3T_2)T_1,$$

since the corresponding multiplication of matrices is associative. The linear fractional transformation T_0 given by $w = z$ is the identity transformation, and for any linear fractional transformation T we have $TT_0 = T_0T = T$. Finally, if T is given by $w = \frac{az+b}{cz+d}$, then the transformation T^{-1} given by $z = \frac{-dw+b}{cw-a}$ is the inverse of T, i.e. $TT^{-1} = T^{-1}T = T_0$. Thus we see that the linear fractional transformations form a group.

We now discuss fixed points of a linear fractional transformation. Let $w = \frac{az+b}{cz+d}$. Then ∞ is a fixed point if and only if $c = 0$. In particular, if $c = 0$ and $\frac{a}{d} \neq 1$, then ∞ and $\frac{b}{d-a}$ are the only fixed points. If $c = 0$ and $\frac{a}{d} = 1$, then the transformation is a translation and ∞ is the only fixed point. In general, setting $z = \frac{az+b}{cz+d}$, we have $cz^2 - (a-d)z - b = 0$; if $c \neq 0$ both fixed points (which may coincide) are finite. Thus we have the following theorem.

Theorem 2.4. *A linear fractional transformation which is not the identity has at most two fixed points. In particular, if a linear fractional transformation leaves three points fixed, it is the identity.*

Theorem 2.5. *Three given distinct points z_1, z_2, z_3 can be mapped respectively onto three given distinct points w_1, w_2, w_3 by one and only one linear fractional transformation.*

Proof. Let T_1 and T_2 be linear fractional transformations given by

$$T_1(z) = \frac{z - z_1}{z - z_3}\frac{z_2 - z_3}{z_2 - z_1} \quad \text{and} \quad T_2(z) = \frac{z - w_1}{z - w_3}\frac{w_2 - w_3}{w_2 - w_1}.$$

Now T_1 maps z_1, z_2, z_3 onto $0, 1, \infty$ respectively if $z_1, z_2, z_3 \neq \infty$, and similarly T_2 maps w_1, w_2, w_3 onto $0, 1, \infty$. If $z_1 = \infty$, let $T_1(z) = \frac{z_2-z_3}{z-z_3}$; if $z_2 = \infty$, let $T_1(z) = \frac{z-z_1}{z-z_3}$; and if $z_3 = \infty$, let $T_1(z) = \frac{z-z_1}{z_2-z_1}$; and similarly for T_2. Then $T = T_2^{-1}T_1$ maps z_1, z_2, z_3 to w_1, w_2, w_3 respectively. Suppose S is any linear fractional transformation mapping z_1, z_2, z_3 to w_1, w_2, w_3. Then $S^{-1}T$ has three fixed points, viz. z_1, z_2, z_3, and therefore $S^{-1}T$ is the identity. Consequently $S = T$. $\square$

EXERCISES

1. Find the linear fractional transformations that map

a) $0, 1, i$ onto $\frac{1}{2}, \frac{2}{3}, \frac{3+i}{6}$ respectively,

b) $\infty, i, 1$ onto $-1, 0, i$ respectively,

c) $-1, \infty, i$ onto $\infty, i, 1$ respectively.

2. Let C and C' be circles or lines. Is there a linear fractional transformation which maps C onto C', and if so, is it unique? Find the following.

a) Find a linear fractional transformation that maps the real axis onto the unit circle.

b) Find a linear fractional transformation that maps the unit circle $|z| = 1$ onto the circle $|z - i| = 1$.

c) Is there a linear fractional transformation that maps the unit circle $|z| = 1$ onto the circle $|z - i| = 1$ leaving the points of intersection of these two circles fixed?

3. Find the fixed points of $w = \frac{z-1}{z+1}$. What is the image of the disk $|z| \leq 1$?

2.4. Cross ratio

Let z_1, z_2, z_3, z_4 be four points in $\bar{\mathbb{C}}$. Their *cross ratio* (z_1, z_2, z_3, z_4) is defined by

$$(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)},$$

where if ∞ is one of the points, the two factors containing it should be canceled (cf. the proof of Theorem 2.5). Note that if z_1, z_2, z_3 and z_4 are all on the real axis this is the same as the cross ratio of four collinear points defined in the last chapter.

Theorem 2.6. *The cross ratio of four points is invariant under a linear fractional transformation.*

Proof. Let $w = \frac{az+b}{cz+d}$ be the transformation and w_i the image of $z_i, i = 1, \ldots, 4$. Then

$$w_i - w_j = \frac{az_i + b}{cz_i + d} - \frac{az_j + b}{cz_j + d} = \frac{(ad - bc)(z_i - z_j)}{(cz_i + d)(cz_j + d)},$$

from which we readily obtain

$$(w_1, w_2, w_3, w_4) = (z_1, z_2, z_3, z_4).$$

□

Notice, however, the effect of inversion.

Theorem 2.7. *Let z_i' be the inverse of points $z_i, i = 1, \ldots, 4$, under inversion in a circle $\mathcal{C}$. Then*

$$(z_1', z_2', z_3', z_4') = \overline{(z_1, z_2, z_3, z_4)}.$$

Proof. Let z_0 be the center of $\mathcal{C}$ and r its radius. Then inversion in $\mathcal{C}$ is given by

$$z' = z_0 + \frac{r^2}{\bar{z} - \bar{z}_0}.$$

Therefore

$$z_i' - z_j' = \frac{r^2}{\bar{z}_i - \bar{z}_0} - \frac{r^2}{\bar{z}_j - \bar{z}_0} = \frac{r^2(\bar{z}_j - \bar{z}_i)}{(\bar{z}_i - \bar{z}_0)(\bar{z}_j - \bar{z}_0)},$$

from which the result follows. □

The basic geometric property of the cross ratio is the following theorem.

Theorem 2.8. *Four distinct points are concircular or collinear if and only if their cross ratio is real.*

Proof. Let z_1, z_2, z_3, z_4 be the points and T the linear fractional transformation mapping z_1, z_2, z_3 to $\infty, 0, 1$ respectively, i.e.

$$T(z) = \frac{(z_1 - z_3)(z_2 - z)}{(z_1 - z)(z_2 - z_3)}.$$

Then $(z_1, z_2, z_3, z_4) = T(z_4) = (\infty, 0, 1, T(z_4))$. But T maps lines and circles to lines and circles and $\infty, 0, 1$ are collinear, so the real axis is the image of the line or circle through z_1, z_2, z_3. Therefore z_4 is on this line or circle if and only if $T(z_4)$ is real. □

One final basic property is that, given three points z_1, z_2, z_3 and a number $\lambda \in \mathbb{C}$, $\lambda \neq 0, 1, \frac{z_1-z_3}{z_2-z_3}$, there exists a unique point z_4 such that $(z_1, z_2, z_3, z_4) = \lambda$. For if

$$\frac{(z_1 - z_3)(z_2 - z_4)}{(z_1 - z_4)(z_2 - z_3)} = \lambda,$$

then we may solve for z_4. Setting $\mu = \frac{(z_1-z_3)}{(z_2-z_3)}$, we get

$$z_4 = \frac{\mu z_2 - \lambda z_1}{\mu - \lambda}.$$

EXERCISES

1. We remarked that if z_1, z_2, z_3, z_4 are on the real axis, the cross ratio agrees with that of Chapter 1. Show that if z_1, z_2, z_3, z_4 lie on a circle, the cross ratio agress with the cyclic cross ratio of Exercise 2 of Section 1.2.

2. Show that the linear fractional transformation mapping z_1, z_2, z_3 onto w_1, w_2, w_3 is determined by

$$(z_1, z_2, z_3, z) = (w_1, w_2, w_3, w).$$

3. Find the equation of the circle through z_1, z_2, z_3 in the form z equal to a complex valued function of a real parameter.

4. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be two circles which intersect at z_0 and z_0*, and let z_1 and z_2 be points on the shorter arcs $\widehat{z_0 z_0*}$ of $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. Show that the angle between the arcs is equal to $\arg(z_2, z_1, z_0, z_0*)$.

5. Give a cross ratio theoretic proof of Miquel's "Big" Theorem. (This could be used as a small project.)

2.5. Some special linear fractional transformations

Let us now consider some special linear fractional transformations. First we will find the most general linear fractional transformation mapping the upper half plane $\{z | \Im z \geq 0\}$ onto the unit disk $\{w | |w| \leq 1\}$. Since linear fractional transformations map lines to lines or circles, the image of the real axis must be a circle. Moreover, since the real axis is the boundary of the upper half plane and linear fractional transformations are continuous, the image of the real axis must be the unit circle $|w| = 1$. In particular, if $T(z) = \frac{az+b}{cz+d}$ is the transformation, then $|T(0)| = |T(\infty)| = 1$. Therefore we must have $|b| = |d|$ and $|a| = |c|$. From the second condition we may write $\frac{a}{c} = e^{i\theta}$ for some angle θ, and T becomes

$$T(z) = e^{i\theta}\frac{z+\frac{b}{a}}{z+\frac{d}{c}};$$

but since we also have $|b| = |d|$, we get $|\frac{b}{a}| = |\frac{d}{c}|$. If now x is any other point on the real axis, then $|T(x)| = 1$, and hence $|x + \frac{b}{a}| = |x + \frac{d}{c}|$. Therefore

$$\left(x+\frac{b}{a}\right)\left(x+\frac{\bar{b}}{\bar{a}}\right) = \left(x+\frac{d}{c}\right)\left(x+\frac{\bar{d}}{\bar{c}}\right)$$

or

$$\left(\frac{b}{a}+\frac{\bar{b}}{\bar{a}}\right)x = \left(\frac{d}{c}+\frac{\bar{d}}{\bar{c}}\right)x.$$

Now $\frac{b}{a} + \frac{\bar{b}}{\bar{a}} = 2\Re(\frac{b}{a})$ and similarly for $\frac{d}{c}$; thus, $\frac{b}{a}$ and $\frac{d}{c}$ have the same real part and the same modulus. Therefore either $\frac{b}{a} = \frac{d}{c}$ or $\frac{b}{a} = \frac{\bar{d}}{\bar{c}}$. The first alternative gives $T(z) = e^{i\theta} =$ const., and consequently, setting $z_0 = -\frac{b}{a}$,

$$T(z) = e^{i\theta}\frac{z-z_0}{z-\bar{z}_0}.$$

This is the desired linear fractional transformation provided that z_0 is in the upper half plane; if z_0 is in the lower half plane, the upper half plane is mapped to the exterior of the unit circle.

Using the above, we can find the most general linear fractional transformation mapping the unit disk onto itself. First note that T^{-1} above maps the unit disk onto the upper half plane. Taking $\{z \big| |z| \le 1\}$ as the unit disk and $\{\zeta \big| \Im\zeta \ge 0\}$ as the upper half plane, we obtain

$$\zeta = T^{-1}(z) = \frac{\bar{z}_0 z - e^{i\theta}z_0}{z - e^{i\theta}}.$$

Consider now a fixed such transformation, for example

$$\zeta = S(z) = -i\frac{z-1}{z+1}.$$

Then $U(z) = T(\zeta) = TS(z)$ maps the unit disk onto the unit disk. Since $T = US^{-1}$ is the general linear fractional transformation mapping the upper half plane onto the unit disk, U is the general linear fractional transformation mapping the unit disk onto itself. Computing U explicitly, we have

$$U(z) = e^{i\theta}\frac{-i\frac{z-1}{z+1} - z_0}{-i\frac{z-1}{z+1} - \bar{z}_0} = e^{i\theta}\frac{-(z_0+i)z - (z_0 - i)}{-(\bar{z}_0+i)z - (\bar{z}_0 - i)}$$

$$= e^{i\theta}\frac{(z_0+i)(-1)}{(1+\bar{z}_0 i)(-i)}\,\frac{z + \frac{z_0-i}{z_0+i}}{(-\frac{\bar{z}_0+i}{\bar{z}_0-i}z - 1)}.$$

Note that $\left|\frac{z_0-i}{z_0+i}\right| < 1$, since z_0 is in the upper half plane. Thus, setting $z_1 = -\frac{z_0-i}{z_0+i}$ and using Exercise 1 of Section 2.2, we have

$$U(z) = e^{i\phi}\frac{z - z_1}{\bar{z}_1 z - 1},$$

where ϕ is real and z_1 is in the open unit disk, as the most general linear fractional transformation mapping the unit disk onto itself.[1]

The question of mapping the upper half plane onto itself is also easily treated. If ∞ is a fixed point, then $w = az + b$ with $a > 0$ and b real maps the upper half plane onto itself. More generally, since the real axis must be mapped to the real axis, let $0, 1, \infty$ be the images of the real numbers x_1, x_2, x_3 respectively. Now, equating the cross ratios $(w, 1, 0, \infty) = (z, x_2, x_1, x_3)$, we have

$$w = \frac{z - x_1}{z - x_3}\,\frac{x_2 - x_3}{x_2 - x_1}.$$

[1]The reader with some experience in complex variable theory can easily see that $U(z)$ is in fact the most general one-to-one holomorphic mapping of the unit disk onto itself. First recall the Schwarz Lemma (see e.g. [1] or [**13**]): Suppose that f is a holomorphic function on the open unit disk with $f(0) = 0$ and $|f(z)| < 1$. Then either $|f(z)| < |z|$ for all z with $0 < |z| < 1$, or $f(z) = e^{i\phi}z$, where ϕ is a real constant.

Theorem 2.9. *If f is a holomorphic function mapping the unit disk one-to-one onto itself, then $f(z)$ if of the form $U(z)$.*

Proof. Let z_1 be the point in the disk mapped to 0 by f and let $T(z) = \dfrac{z - z_1}{\bar{z}_1 z - 1}$. Then by the Schwarz Lemma we have $|(f \circ T^{-1})(\zeta)| \le |\zeta|$ and $|(T \circ f^{-1})(w)| \le |w|$, giving $|(f \circ T^{-1})(\zeta)| = |\zeta|$ and hence $f \circ T^{-1}(\zeta) = e^{i\phi}\zeta$; that is, $f(z) = e^{i\phi}T(z) = U(z)$. $\square$

Since the mapping must take the upper half plane onto itself, the imaginary part of the image of i must be positive, giving

$$\frac{x_2 - x_3}{x_2 - x_1}\,\frac{x_1 - x_3}{x_3^2 + 1} > 0.$$

Thus, writing the mapping in the form $w = \frac{az+b}{cz+d}$, the determinant $ad - bc$ of the mapping is positive. In particular we can write the mapping in the form $w = \frac{az+b}{cz+d}$ with $ad - bc > 0$ and a, b, c, d real.

Now clearly the linear fractional transformation

$$w = \frac{a\mu z + b\mu}{c\mu z + d\mu}, \qquad \mu \neq 0,\ \mu \in \mathbb{C}$$

is geometrically the same as $w = \frac{az+b}{cz+d}$. Writing $\mu = Re^{i\theta}$, its two square roots $\pm\sqrt{\mu}$ are $\pm\sqrt{R}e^{i\theta/2}$. In particular, taking $\mu = \frac{1}{\sqrt{ad-bc}}$, we may normalize any linear fractional transformation so that $ad - bc = 1$. Thus, if a linear fractional transformation $T(z)$ has been normalized so that $ad - bc = 1$, then T maps the upper half plane onto itself if and only if a, b, c, d are real.

Exercises

1. If $w = \frac{z-1}{z}$, what is the image of the positive real axis?

2. Let $U(z)$ be the general linear fractional transformation mapping the unit disk onto itself. Show that $U(z)$ may be written in the following forms:

$$U(z) = e^{i\phi}\frac{e^{i\theta}z - R}{Re^{i\theta}z - 1}, \qquad 0 \le R < 1, \quad \theta, \phi \text{ real},$$

$$U(z) = \frac{az + b}{\bar{b}z + \bar{a}}, \qquad |a|^2 - |b|^2 > 0.$$

3. The transformation

$$U(z) = e^{i\phi}\frac{z - z_1}{\bar{z}_1 z - 1}$$

maps the unit disk onto the unit disk. Find z_1 and ϕ such that U maps the segment $0 \le x \le a, a < 1$, onto a segment of the form $-b \le x \le b, b < 1$.

2.6. Extended Möbius transformations

We have already seen that inversion in the unit circle is given by $w = 1/\bar{z}$ or in a circle of radius r and center z_0 by

$$w = z_0 + \frac{r^2}{\bar{z} - \bar{z}_0} = \frac{z_0\bar{z} + (r^2 - |z_0|^2)}{\bar{z} - \bar{z}_0};$$

thus inversion is not a linear fractional transformation, but is of a type we now discuss. Consider a mapping of $\bar{\mathbb{C}}$ onto itself given by

$$w = \frac{a\bar{z} + b}{c\bar{z} + d},$$

again with $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$. Such a transformation is sometimes called an *anti-homography.* As with linear fractional transformations, if $c = 0$, then $w = \frac{a}{d}\bar{z} + \frac{b}{d}$, which is a reflection in the real axis followed by a similarity. If $c \neq 0$, then

$$w = -\frac{ad - bc}{c}\frac{1}{c\bar{z} + d} + \frac{a}{c},$$

which is again a composition of similarities, reflection in the real axis and an inversion. Thus we have the following theorem.

Theorem 2.10. *Anti-homographies map lines and circles to lines and circles, and are conformal.*

Moreover, the cross ratio of the images of four points under an anti-homography is the conjugate of the cross ratio of the points. In particular, the cross ratio of four collinear or concircular points is preserved.

Now considering compositions of anti-homographies with homographies (linear fractional transformations) or anti-homographies, we see that the set of all homographies and anti-homographies forms a group. We call this group the group of *extended Möbius transformations.*

Since extended Möbius transformations map lines and circles to lines and circles, one might ask if this property characterizes such transformations. It is a beautiful result of Carathéodory [**6**] that even locally this is true. That is, not only is a 1-1 circle-preserving map of $\bar{\mathbb{C}}$ onto itself an extended Möbius transformation, but a 1-1 map of a plane region R onto a set R' such that the image of every circle lying in R is a line or circle in R' is such a transformation. Not even the continuity of the transformation is assumed. The reader is encouraged to work through Carathéodory's proof, drawing an appropriate diagram for each step of the proof; we reprint the paper below (published in 1937 in the Bulletin of the American Mathematical Society). For definitions of some concepts from analysis and topology, such as open set, region, etc., see the beginning of the next chapter.

Exercises

1. Show that an extended Möbius transformation which leaves four non-collinear, non-concircular points fixed is the identity.

2. Prove that if a transformation of the plane onto itself preserves cross ratios, then it is a linear fractional transformation.

3. Let $\mathcal{B}$ be a circle and $\mathcal{A}$ a circle in the interior of $\mathcal{B}$. Show that there is a linear fractional transformation mapping $\mathcal{A}$ and $\mathcal{B}$ onto concentric circles. Use the results of Chapter 1, or prove this directly using ideas from this chapter.

4. One may have a transformation mapping lines to lines but not circles to circles. Show that $x' = x + y$, $y' = y$ maps lines to lines but not circles to circles, and that it is not angle preserving.

5. Show that a 1-1 mapping of the plane onto itself which maps lines to lines and circles to circles is a similarity. You may want to use the Carathéodory result.

THE MOST GENERAL TRANSFORMATIONS OF PLANE REGIONS WHICH TRANSFORM CIRCLES INTO CIRCLES

BY CONSTANTIN CARATHÉODORY

1. *Introduction.* If we consider two plane regions R and R' which are mapped conformally the one upon the other, there corresponds to every circle c contained in R an analytic closed curve c' contained in R'. If an arc α lying on the circle c has a circular image the curve c' must be itself a circle.

Suppose that the interior of the circle c belongs to the region R. Then both circular discs bounded by c and c' respectively are represented conformally the one upon the other. It is a well known fact that in this case the transformation of these circles into one another is given by a transformation of Möbius by which, furthermore, every circle of R is transformed into a circle of R'.

If we drop now the condition that the one to one correspondence of our regions must be conformal, the assumption that one single circle of R has a circular image is no longer sufficient to characterize the transformations of Möbius. On the other hand, if we make the stronger assumption that *every* closed circle contained in R is transformed into a circle, we shall see in the course of this paper that a theorem analogous to the one stated above holds under surprisingly general conditions. To prove that the transformation which we consider is a transformation of Möbius it is no longer necessary to assume from the outset (as is the case for the analogous theorem concerning collinear transformations) that this transformation is continuous, or that it is measurable in the sense of Lebesgue, or even that the point set R' is itself a region.

The condition that circles lying in a region R have circular images characterizes the group of Möbius transformations among all the one to one arbitrary correspondences between the points of R and the points of a quite arbitrary point set of the plane.

2. *Statement of a Preliminary Theorem.* We consider a circular open disc which we shall call C and suppose that there is a one

to one correspondence between the points P of this disc and the points P' of a *bounded* point set C'. We do not suppose this correspondence to be continuous or even to be measurable in the sense of Lebesgue. But we assume that every closed circular line c which is contained in the domain C is transformed by the above correspondence into a line c' of the same kind whose points lie on C'. We shall prove that every transformation which possesses these properties is a Möbius transformation and therefore analytic.

It is very important to make the following rather obvious remark: if two circles c_1' and c_2' contained in C' are the images of two circles of C, say of c_1 and c_2, then the number of points common to the pair of circles c_1' and c_2' is the same as the number of points common to c_1 and c_2. In particular, if c_1 and c_2 are tangent to one another the same is the case for c_1' and c_2' and conversely.

3. *Reduction of the Transformation to a Normal Type.* In using transformations of M bius we plot the domain C and a circle D' which contains in its interior the bounded point set C' on the unit circles of two planes with coordinates x, y and x', y' respectively. In doing this we can always assume that the origins of both planes correspond to points of C and of C' which are images of one another by the given transformation.

Finally we make two inversions respectively on the unit circles of both these planes.

4. *Preservation of Parallelism.* To the original transformation of the domain C into the point set C' corresponds now a one to one transformation of the exterior E of the unit circle in the plane with coordinates x, y into some point set E' lying outside of the unit circle of the second plane. To every closed circular line lying in E corresponds a closed circle of E'. Finally every straight line contained entirely in E is transformed into a straight line lying in E'.

Likewise every pair of parallel straight lines of E is transformed into a pair of parallel straight lines of E'.

5. *Preservation of Orthogonality.* Take now two such pairs of parallel straight lines lying in E and cutting at right angles. Suppose furthermore that there is a circle contained entirely in E and circumscribed to the rectangle formed by these lines.

The whole figure is transformed into a circle of E' in which a parallelogram is inscribed; this parallelogram must therefore be a rectangle.

We infer herefrom without difficulty that two straight lines of E orthogonal to one another are always transformed into lines of E' having the same property.

6. *The Image of C is a Domain.* We take two circles of E which are concentric to the unit circle and consider two or more rectangles inscribed in the one of these circles and having two opposite sides tangent to the other. As the corresponding figure lying on E' shows exactly the same disposition, the circles contained in these figures and which are images of the two circles considered above must also be concentric to one another.

Furthermore let c be an arbitrary circle of E concentric to the unit circle. To every point P' lying on the image c' of c corresponds a point P of c. To the tangent to the circle c passing through the point P corresponds the tangent to c' passing through P'. Consequently every point of the plane with coordinates x', y' exterior to c' must belong to the point set E'. We conclude from this that E' is an open set consisting of all the points which are exterior to a certain circle. The point set C' which we have considered above must therefore be a circular disc and we could have taken the circle D' coinciding with the boundary of C'.

If we do this the point set E' coincides with the exterior of the unit circle of our plane with coordinates x', y', and the circles of E concentric to the unit circle are transformed into circles of E' having the same property.

7. *Preservation of the Centers of Circles.* Take now a circle γ lying in E whose center is at a distance from the origin greater than $2^{1/2}$. Under these conditions there are at least two diameters of γ orthogonal to one another which belong to straight lines lying in E. There are also at least two tangents to the same circle cutting at right angles and parallel to the diameters which we have been considering and entirely contained in E.

This figure transforms into a figure contained in E' from whose inspection we conclude that if we call γ' the image of γ, the center of γ' is the image of the center of γ.

8. *Preservation of Angles at the Origin.* We take now the positive number

$$r > 2^{1/2}, \tag{1}$$

and consider the circle

$$x^2 + y^2 = r^2 \tag{2}$$

and the circle

$$x'^2 + y'^2 = r'^2 \tag{3}$$

into which the circle (2) is transformed. We will show that every regular polygon inscribed in (2) and whose sides have a length less than $(r-1)$ is transformed into a regular polygon inscribed in (3) and having the same number of vertices.

Two adjoining sides of the former polygon must in fact, by the result of the last paragraph, be transformed into contiguous segments of equal length inscribed in (3). The second polygon is therefore either a regular polygon similar to the first or it is starshaped. But in this latter case at least two sides of it would cross at a point interior to the circle (3) and such a point must necessarily be the image of a point exterior to the circle (2). But this is readily shown to be impossible by using a reasoning like that of §6. Thus it is proved that the two polygons are similar to one another.

We apply this to all such polygons inscribed in (2) which have a common vertex at the point $x=r$, $y=0$.

Any point of the circle (2) which satisfies the condition

$$x + iy = r \cdot e^{2\pi i p/q} \tag{4}$$

where p and q are arbitrary integers and $q \neq 0$, can be considered as a vertex of one at least of these polygons. And we can always choose the coordinates x' and y' in such a way that the images of all the points (4) must be calculated by the formulas

$$\frac{x'}{r'} = \frac{x}{r}, \qquad \frac{y'}{r'} = \frac{y}{r}. \tag{5}$$

9. *Families of Circles which Are Transformed by a Similarity.* Take now all the tangents to the circle (2) whose points of contact coincide with the point set (4) and call A the set of points of intersection of each two of these tangents. Each point of the

set A will then be transformed into a certain point contained in the exterior of the circle (3) whose coordinates must be calculated by (5).

Call A^* the family of circles, each of which is contained in E and passes through three points at least of the set A. The image of every one of these circles is necessarily connected with the circle itself by the transformation (5).

We remark finally that the point set A is everywhere dense in the exterior of the circle (2) and that therefore every circle lying in E and possessing points outside of (2) can be indefinitely approximated by circles belonging to the family A^*.

10. *Proof of the Continuity of the Transformation.* Our next step is to show that our transformation is necessarily continuous along the circle (2). If this were not the case there would exist on this circle at least one point P_0 whose image on the circle (3) we obtain by a rotation around the origin through an angle θ lying between zero and 2π. We could then construct a point set B analogous to our former point set A and a set of circles B^* analogous to A^* and possessing the following property: To obtain the image of a circle belonging to B^* we must combine the amplification of §9 according to the cases either with a rotation around the origin through the angle θ or with a reflection on some diameter of the circle (3).

In both these cases it is always possible to find two circles, say c_1 and c_2, both lying in the region

$$x^2 + y^2 > r^2, \tag{6}$$

cutting one another, and such that if we transform c_1 by the amplification of §9 and c_2 by one of the transformations which have just been described, we obtain as result of these transformations circles which have no point in common.

We now approximate c_1 by a circle c_1^* belonging to A^* and c_2 by a circle c_2^* belonging to B^*. If these approximations are close enough the circles c_1^* and c_2^* will cut one another and will be transformed by the transformation we study into circles having no common point. But this is in contradiction with the general principle laid down at the end of §2.

We must therefore assume that all the points of the circle (2) are transformed into the points of (3) by the transformation

(5) and the same is then the case for all the points of the region (6).

11. *The Theorem for the Circle.* Take now a point P of E lying between the unit circle and the circle (2). Consider two circles of E passing through P and possessing points inside of the region (6). By hypothesis these circles are transformed into circles and besides they contain arcs whose images we know.

The result is that the formulas (5) must hold for every point of the region E. As this region must by assumption be transformed into E' we find that we must have

$$r' = r. \tag{7}$$

In short we have proved the following theorem:

THEOREM 1. *Every arbitrary one to one correspondence between the points of a circular disc C and a bounded point set C' by which circles lying completely in C are transformed into circles lying in C' must always be either a direct or an inverse transformation of Möbius.*

12. *The General Theorem.* We now consider a general region R and suppose that there exists a one to one correspondence between the points of R and some point set R'. We assume furthermore that by this correspondence circles of R are transformed into circles lying on R'. We call c and c' two such corresponding circles and suppose that every point interior to c belongs to R. Through every pair of points P and Q lying in the interior of c there passes at least one circle c_1 having no points in common with c. To this circle corresponds a circle c_1' which has no point in common with c'. Therefore both points P' and Q' which correspond to P and Q must lie on the same side of c'. It follows that the interior of c must correspond to a point set which lies on one side of c'. This point set is either itself bounded, or it can be transformed by an inversion on the circle c' into a bounded set. In both cases we can apply the previous theorem. As the circle c can be taken at random we have finally the following theorem.

THEOREM 2. *Suppose that to every point P of a region R corresponds a point P' of some point set R' and that to two different points P and Q of R correspond two points P' and Q' of R' differ-*

ent from one another. Suppose that to every circle c contained with its interior in the region R corresponds a point set of R' which consists of all the points of a closed circle c'. Then the point set R' is itself a region and the transformation of R into R' is analytic and either a direct or an inverse transformation of Möbius.

It is not difficult to generalize this result by restricting the class of circles c belonging to R which are supposed to be transformed into circles. Take for instance a continuous positive function $\phi(P)$ defined everywhere in the region R. Then the theorem holds if we suppose that every circle of center P and whose radius is less than $\phi(P)$ is transformed into a circle lying on R'.

The following generalization of Theorem 2 is nearly self evident if we note that three circles in space cutting one another at six different points must lie on the same sphere.

THEOREM 3. *Supposing that a plane region R is transformed by a one to one correspondence into an arbitrary point set R' lying in n-space ($n \geqq 3$) under the same assumptions as before, then R' must be a two dimensional sphere or a plane and the transformation is a transformation of Möbius.*

Other generalizations can also be imagined which, however, are outside the scope of this note.

UNIVERSITY OF MUNICH

2.7. The Poincaré models of hyperbolic geometry

We have touched very briefly on the Poincaré disk model of the hyperbolic plane in the Remark in Section 1.1. Here it seems appropriate to expand this discussion, though the reader interested in our main topics of inversion, circle preserving maps and conformal maps may skip this and the next section.

The points of the Poincaré disk model of the hyperbolic plane H^2 are the points interior to the unit disk, and the lines are diameters and arcs of circles orthogonal to the boundary, $|z| = 1$. We begin by defining the distance between two points in this geometry. Given two points z_1 and z_2, we constructed the hyperbolic or Poincaré line joining them in Section 1.1. Let ζ_1 and ζ_2 be the endpoints of the diameter or arc determining the Poincaré line (Figure 2.4).

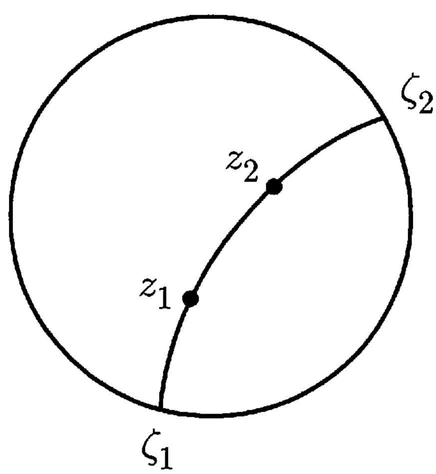

Figure 2.4

The distance between z_1 and z_2 is then defined by

$$d(z_1, z_2) = |\log(z_1, z_2, \zeta_1, \zeta_2)|.$$

Note that the cross ratio is real and positive so that the logarithm is defined; note also that the definition is independant of which endpoint is designated as ζ_1. The distance function d enjoys the following properties: $d(z_1, z_2) \geq 0$ with equality if and only if $z_1 = z_2$; $d(z_1, z_2) = d(z_2, z_1)$; and if z_2 is between z_1 and z_3 on the Poincaré line, then $d(z_1, z_2) + d(z_2, z_3) = d(z_1, z_3)$. To see this last property we multiply the cross ratios:

$$\left(\frac{(z_1-\zeta_1)(z_2-\zeta_2)}{(z_1-\zeta_2)(z_2-\zeta_1)}\right)\left(\frac{(z_2-\zeta_1)(z_3-\zeta_2)}{(z_2-\zeta_2)(z_3-\zeta_1)}\right)=\left(\frac{(z_1-\zeta_1)(z_3-\zeta_2)}{(z_1-\zeta_2)(z_3-\zeta_1)}\right).$$

Taking logarithms, we have

$$\log(z_1,z_2,\zeta_1,\zeta_2)+\log(z_2,z_3,\zeta_1,\zeta_2)=\log(z_1,z_3,\zeta_1,\zeta_2).$$

Each of these logarithms has the same sign, and we have $d(z_1,z_2)+d(z_2,z_3)=d(z_1,z_3)$.

With this notion of distance, one can say that two Poincaré segments are congruent if they have the same length. Angles in the Poincaré model are measured in the Euclidean way, i.e. the measure of the angle between two Poincaré lines is the Euclidean measure of the angle between their tangents. Now with these notions of congruence and angle measure, one can verify that all the axioms of Euclidean geometry hold in the Poincaré model except for the axiom of parallels (see e.g. Greenberg [**16**]). Indeed, as we noted in Section 1.1, through a point not on a given Poincaré line, there are many parallels.

We have seen that the linear fractional transformations form a group. If now U and V are linear fractional transformations mapping the unit disk onto itself, then so is their composition; also U^{-1} is such a mapping. Therefore the set of all linear fractional transformations mapping the unit disk onto itself is a subgroup of the group of all linear fractional transformations. Also U composed with the map $z \longrightarrow \bar{z}$ maps the unit disk onto itself, and the general anti-homography mapping the unit disk onto itself is of this form. We can therefore form the subgroup G of the group of extended Möbius transformations mapping the unit disk onto itself.

Now let $U \in G$ and $w_i = U(z_i)$, $i = 1, 2$. Since the transformations map lines and circles to lines and circles and are conformal, U maps the Poincaré line through z_1 and z_2 to a line or circle through w_1 and w_2 which, since $|z| = 1$ is mapped to itself, is orthogonal to $|z| = 1$. Therefore U maps the Poincaré line through z_1 and z_2 to the Poincaré line through w_1 and w_2.

Since transformations in G preserve the cross ratio of four collinear or concircular points, they preserve the hyperbolic distance between two points in the Poincaré disk. Thus elements of G are isometries of the Poincaré disk model of the hyperbolic plane. We will see below that a hyperbolic circle in this model is also a Euclidean circle. Thus an isometry of the Poincaré disk is a circle-preserving map and hence an element of G. Therefore G is the isometry group of the Poincaré disk model of hyperbolic geometry. The most basic isometry is reflection in a line, which here is just inversion in the circle whose arc in the unit disk is the Poincaré line. We state without proof the fact that if $z_1, z_2 \in H^2$, the shortest curve in the Poincaré metric[2] joining them is the Poincaré line. Also, given a Poincaré line and a point not on it, the distance from the point to the line is given by the unique perpendicular from the point to the line.

The group G acts *transitively* on the open unit disk; i.e., given z_1 and z_2 in the open unit disk, there exists $U \in G$ such that $U(z_1) = z_2$. To see this, note that $U_1(z) = e^{i\phi}(z - z_1)/(\bar{z}_1 z - 1)$ maps z_1 to 0. Let U_2 be a similar mapping taking z_2 to 0. Then $U_2^{-1}U_1$ maps z_1 to z_2. In fact we have shown that the subgroup of G consisting of the linear fractional transformations mapping the unit disk onto itself also acts transitively on the unit disk.

Returning to the multitude of parallels to a given line through a given point, we observe that two of these are special. For a Poincaré line in H^2 let ζ_1 and ζ_2 be the points on the boundary $|z| = 1$, and z a point in H^2 not on the line. Then the Poincaré lines through z and the boundary points ζ_1 and ζ_2 are parallel to the given Poincaré line, since $\zeta_1, \zeta_2 \notin H^2$; the rays from z to ζ_1 and ζ_2 are called "limiting" or "asymptotic" rays (Figure 2.5). The perpendicular from z to the line bisects the angle between these rays (Exercise 1, at the end of this section), and the angle α between one of these rays and the

[2]The length of a smooth curve between two points is found by integrating the element of arc length ds. For the Poincaré disk model of hyperbolic geometry the element of arc length is given by

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}$$

(see e.g. [**23**], p. 242).

perpendicular is called the *angle of parallelism* at z with respect to the Poincaré line.

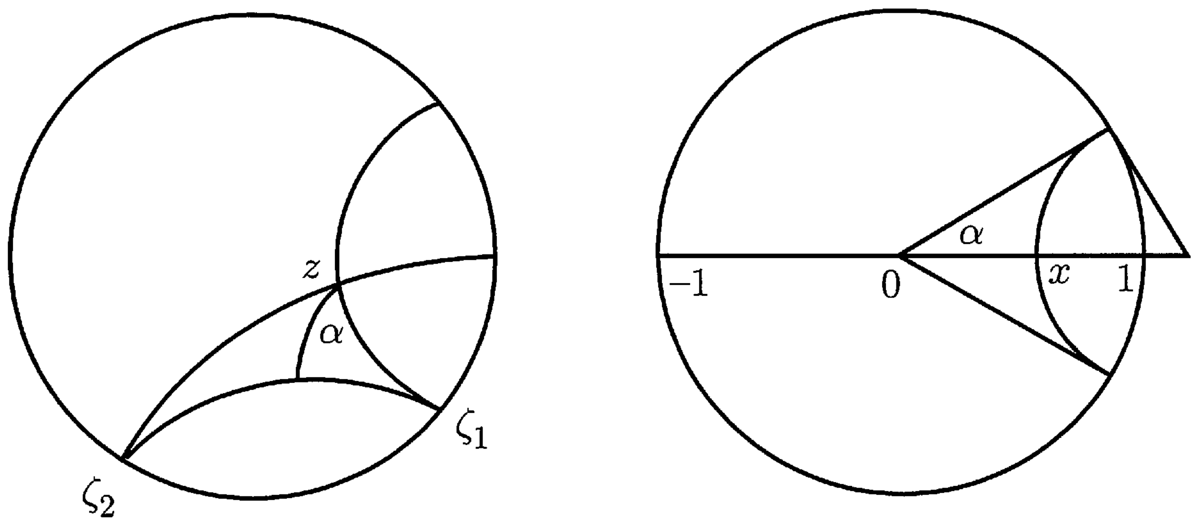

Figure 2.5

We now derive a formula for the angle of parallelism. Let d be the distance from z to the Poincaré line; then

$$e^{-d} = \tan\frac{\alpha}{2}.$$

To see this, first, if necessary, map z to the origin and rotate about the origin so that the perpendicular is along the positive real axis (Figure 2.5). Let x be the foot of the perpendicular. Now $d = |\log(0, x, -1, 1)| = \log\frac{1+x}{1-x}$, and note that $x = \sec\alpha - \tan\alpha$. We then make the following computation:

$$\begin{aligned} e^{-d} = \frac{1-x}{1+x} &= \frac{\cos\alpha - 1 + \sin\alpha}{\cos\alpha + 1 - \sin\alpha} = \frac{\cos^2\alpha + 2\cos\alpha\sin\alpha + \sin^2\alpha - 1}{\cos^2\alpha + 2\cos\alpha + 1 - \sin^2\alpha} \\ &= \frac{\sin\alpha}{\cos\alpha + 1} = \frac{2\cos\frac{\alpha}{2}\sin\frac{\alpha}{2}}{2\cos^2\frac{\alpha}{2}} = \tan\frac{\alpha}{2} \end{aligned}$$

From this formula we make the important observation that the distance from one asymptotic ray to another tends to zero. In the above formalism let the point z move along the asymptotic ray to the boundary point ζ_1; as it does, the angle of parallelism tends to $\frac{\pi}{2}$ and hence the distance $d \longrightarrow 0$.

On the other hand, for two parallels which are not limiting, the distance between points on one Poincaré line and points on the other goes to infinity. Notice that, in fact, as a point z moves to the boundary along one ray, the foot of the perpendicular from z to the other Poincaré line approaches a finite point, i.e. a point still in the interior of the disk.

A second common model of the hyperbolic plane is the Poincaré upper half plane model. Here the points of the geometry are the points z in the upper half plane, i.e. $\Im z > 0$, and the lines are either vertical rays from points on the real axis or semicircles with diameter on the real axis. Given two points z_1 and z_2 in the upper half plane, the Euclidean perpendicular bisector of the Euclidean segment joining them meets the real axis at the center of a semicircle through the two points. If we let ζ_1 and ζ_2 be the endpoints of the semicircle, we can define distance in this model as in the disk model, viz. $d(z_1, z_2) = |\log(z_1, z_2, \zeta_1, \zeta_2)|$.

Since we have studied linear fractional transformations mapping the unit disk onto the upper half plane, we immediately have a one-to-one mapping of one model onto the other which maps the lines of one model to the lines of the other, preserves distances, and maps the isometry group of one model to the isometry group of the other. Thus the two models are isometric, and some properties are easier to prove in one model than in the other.

In our discussion of parallels we saw that the distance from one asymptotic ray to another went to zero, but for non-asymptotic parallels the distance between them goes to infinity. Thus in neither case are parallel lines equidistant sets. This raises the question of what a set of points equidistant from a line on one side of the line might be. Let ζ_1 and ζ_2 be the boundary points of a Poincaré line in either model, and consider an arc of another Euclidean circle through ζ_1 and ζ_2 (Figure 2.6). We will show that the Poincaré line and this arc are equidistant sets. First map the disk model to the upper half plane model, or apply an isometry of the upper half plane model, so that the image of the Poincaré line is a vertical line ($\zeta_1 \longrightarrow 0$, $\zeta_2 \longrightarrow \infty$) (cf. Section 2.5 and the exercises below). Then the arc will be mapped to an oblique Euclidean ray from the origin (see Figure 2.6). Let ia be a point on the imaginary axis, and consider the semicircle centered at the origin of radius a. Let $b + ic$ be the point of intersection of the oblique Euclidean ray with the semicircle, and note that the semicircle is perpendicular to both Euclidean rays. The distance between ia and $b + ic$ is the logarithm of the cross ratio $(ia, b + ic, -a, a) = \frac{c}{a+b}$. Now if we start with the point $i\mu a$, its distance to the point $\mu b + i\mu c$ on the other ray is the same, since $(\mu ia, \mu b + i\mu c, -\mu a, \mu a) = (ia, b + ic, -a, a)$. In particular we also note that the Euclidean homothety (stretching or shrinking) $w = \mu z$ is an isometry of the upper half plane.

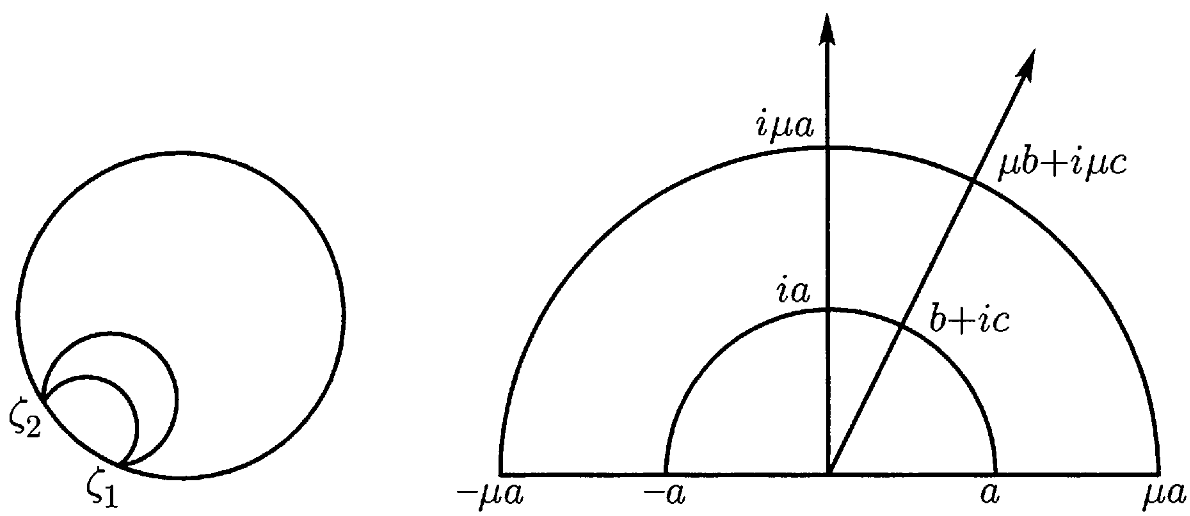

Figure 2.6

We just noted the role played by a Euclidean circle with an arc in the Poincaré disk. A circle lying in the disk except for a point of tangency with the boundary is called a *horocycle* in hyperbolic geometry. In the upper half plane model a horocycle is either a circle tangent to the real axis or a Euclidean line parallel to the real axis. Any two horocycles are congruent, i.e. there is an isometry mapping one onto the other. To see this, first use a rotation of the unit disk to position the two horocycles so that they have the same boundary point. Now map the unit disk to the upper half plane with the common boundary point being mapped to ∞. The horocycles are then mapped to a pair of Euclidean lines parallel to the real axis. Let $y = a$ and $y = b$ be these lines. Then $w = \frac{b}{a}z$ maps $y = a$ to $y = b$ and is an isometry of the upper half plane. Noting also that $(x + ia, x + ib, x, \infty) = \frac{a}{b}$, we see that two horocycles with the same boundary point are equidistant sets.

We have discussed arcs of Euclidean circles lying in the disk and Euclidean circles tangent to the boundary, so now consider a Euclidean circle which lies entirely in the open unit disk. We have seen that two non-concentric circles can be inverted to concentric circles. Following this by a homothety and translation, as necessary, we have an isometry of the disk mapping the given circle to one concentric with the origin. Rotation about the origin is both a Euclidean and a hyperbolic isometry; thus a Euclidean circle within the disk centered at the origin is also a hyperbolic circle. Therefore, mapping this circle back to its original position by the inverse isometry, we see that a Euclidean circle lying in the disk is also a hyperbolic circle, but the Euclidean center is not the hyperbolic center except when the center is the origin (Figure 2.7).

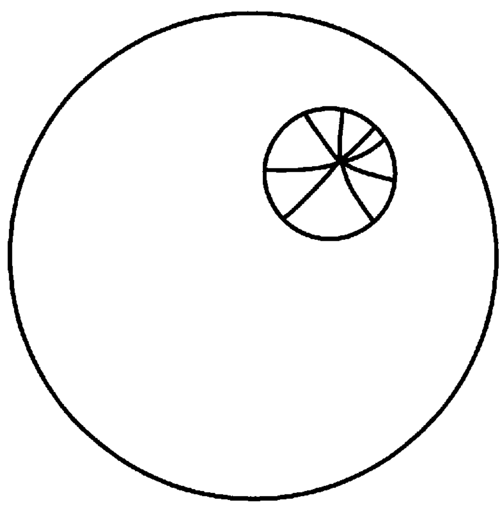

Figure 2.7

Finally, and perhaps most dramatically, we note that the sum of the angles of a triangle in hyperbolic geometry is less than 180°. We won't discuss this in detail, but refer to [**16**] for an extensive discussion of the angle-sum question. Suffice it to observe that one can construct triangles with angle sum as small as one likes, e.g. less than 1°. In the disk model, simply take a triangle whose vertices are near, in the Euclidean sense, to the boundary; then all three angles will be very small. There are other interesting phenomena in hyperbolic geometry, e.g. that AAA is a congruence! and hence there is no notion of similarity in hyperbolic geometry, but we leave these for a more thorough treatment as in [**16**].

EXERCISES

1. For a given Poincaré line and point z not on it, prove that the perpendicular from z to the line bisects the angle between the limiting rays from z to the Poincaré line.

2. Show that the linear fractional transformations with $ad - bc = 1$ and a, b, c, d real form a group H. Here H is the group of orientation preserving isometries of the upper half plane model of the hyperbolic plane. Show that H acts transitively on the upper half plane. Show also that, given a boundary point ζ on the real axis, there is an element of H mapping $\zeta \longrightarrow \infty$.

2.8. A distortion theorem

Before considering our topic we give, by means of exercises, some further geometry of the circle.

EXERCISES

1. In Chapter 1 we saw that through two points in the interior of a circle and not collinear with the center, there exists a unique circle orthogonal to the given circle. Show that through two point in the interior of a circle C there exist two circles tangent to C. Hint: Consider tangents to C from the point of intersection of the line through the two points and the line through the points where the circle orthogonal to C through the two points meets C.

2. Let P be a point in the interior of a circle C. Find the locus of points equidistant from P and C. Using this, give another solution to Exercise 1.

3. Let C be the unit circle, z_1 and z_2 two points in the interior of C, and C_1 and C_2 the circles through z_1 and z_2 tangent to C found in the previous exercises. Let t_1 and t_2 be the points of tangency of C_1 and C_2 with C, respectively. Given a negative number μ, let a_1 and a_2 be the points on C_1 and C_2 respectively such that $(z_1, z_2, a_i, t_i) = \mu, i = 1, 2$. Show that there exists a circle $\mathcal{A}$ tangent to C_1 and C_2 at a_1 and a_2 (Figure 2.8). Hint: Invert with z_1 as center of inversion.

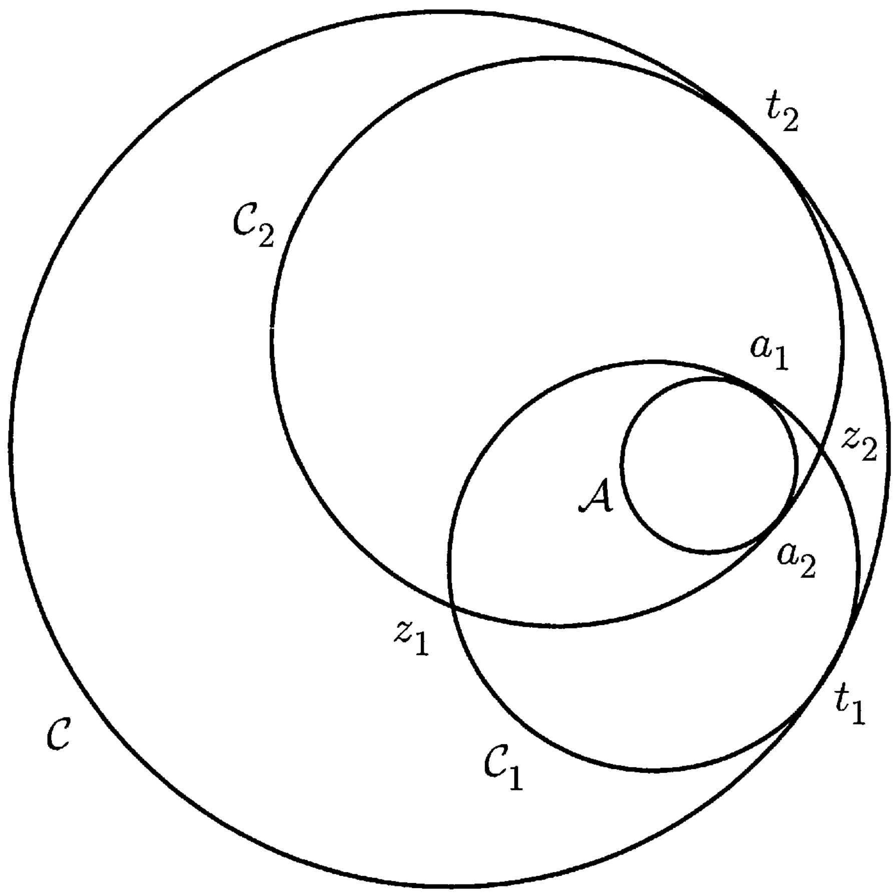

Figure 2.8

We have already seen that a linear fractional transformation maps lines and circles to lines and circles; here we consider a subcollection of these transformations and the set of pre-images of a fixed point on a segment. This problem is sometimes called "the distortion of a segment".

Let z_1 and z_2 be two points in the open unit disk, and w_1 and w_2 two points in the plane. Consider the set of all linear fractional transformations mapping z_i to w_i, $i = 1, 2$, and the unit disk to a disk or half plane (i.e. from all linear fractional transformations mapping z_i to w_i we exclude those mapping the unit disk to the exterior of some disk). Let $w = (1 - \lambda)w_1 + \lambda w_2$ be the line through w_1 and w_2. Now for a fixed value of $\lambda, 0 < \lambda < 1$, we seek the set $\mathfrak{A}$ of all points a which are mapped to the point $(1 - \lambda)w_1 + \lambda w_2$ by some linear fractional transformation in our subset.

Theorem 2.11. *In the notation of Exercise* 3 *above and Figure* 2.8, *set* $\mu = \frac{\lambda}{\lambda - 1}$. *Then* $\mathfrak{A}$ *consists of the circle* $\mathcal{A}$ *of Exercise* 3 *and its interior.*[3]

Proof. Since the transformations are mapping the unit disk onto a disk or half plane, the pre-image of ∞ must be a point ζ with $|\zeta| \geq 1$, and hence the pre-image of the segment $w_1 w_2$, which is a circular arc or segment, must lie in the spindle-shaped region between C_1 and C_2. Now the points a_1 and a_2 of Exercise 3 belong to $\mathfrak{A}$ because the linear fractional transformation mapping z_1, z_2, a_i to $w_1, w_2, (1-\lambda)w_1 + \lambda w_2$ is given by

$$(z_1, z_2, a_i, z) = (w_1, w_2, (1 - \lambda)w_1 + \lambda w_2, w)$$

and takes t_i to ∞ since

$$(w_1, w_2, (1 - \lambda)w_1 + \lambda w_2, \infty) = \frac{\lambda}{\lambda - 1}.$$

Now for general $a \in \mathfrak{A}$ the linear fractional transformation (z_1, z_2, a, z) $= (w_1, w_2, (1 - \lambda)w_1 + \lambda w_2, w)$ must be such that the pre-image ζ of ∞ satisfies $|\zeta| \geq 1$ and

$$(z_1, z_2, a, \zeta) = \frac{\lambda}{\lambda - 1}.$$

Solving for a, we have

$$a = \frac{-((1 - \lambda)z_1 + \lambda z_2)\zeta + z_1 z_2}{-\zeta + \lambda z_1 + (1 - \lambda)z_2}$$

and

$$[-((1 - \lambda)z_1 + \lambda z_2)][\lambda z_1 + (1 - \lambda)z_2] + z_1 z_2 = \lambda(\lambda - 1)(z_1 - z_2)^2 \neq 0.$$

Thus we have a linear fractional transformation for a in terms of ζ. It maps ∞ to $(1 - \lambda)z_1 + \lambda z_2$, so that the image of $\{\zeta \big| |\zeta| \geq 1\}$ must be a closed disk lying in the spindle-shaped region between C_1 and C_2. We have already noted that a_1 and a_2 are in the closed disk, so this disk must be the circle $\mathcal{A}$ and its interior. □

---

[3]An interesting, more general result is known in complex variable theory. Consider the set of all holomorphic functions mapping z_i to w_i, $i = 1, 2$, and the unit disk in a one-to-one manner onto a convex set. Then again the set of points mapped to $(1 - \lambda)w_1 + \lambda w_2$ by these functions is the circle $\mathcal{A}$ and its interior (Aumann [**3**]).

Chapter 3

Advanced Calculus and Conformal Maps

3.1. Review of advanced calculus

This section is intended primarily to provide a review of some results and notation from standard courses in advanced calculus and real analysis, and to set the stage for our later work. Consequently we will present results without proof, and the reader may consult any standard text for more details (e.g. [**2**], [**4**], [**14**]).

We denote by $\mathbb{R}^n$ the set of n-tuples $(x^1, \ldots, x^n)$ of real numbers and for simplicity write $\mathbf{x}$ for $(x^1, \ldots, x^n)$. Denote the distance, $\left(\sum_{i=1}^n (x^i)^2\right)^{1/2}$, of $\mathbf{x}$ from the origin by $|\mathbf{x}|$. Then $|\mathbf{x} - \mathbf{y}|$ is the distance between $\mathbf{x}$ and $\mathbf{y}$.

The *open ball* in $\mathbb{R}^n$ with center $\mathbf{x}_0 = (x_0^1, \ldots, x_0^n)$ and radius r, denoted $B_r(\mathbf{x}_0)$, is defined by

$$B_r(\mathbf{x}_0) = \{\mathbf{x} \big| |\mathbf{x} - \mathbf{x}_0| < r\}.$$

A subset $U \subset \mathbb{R}^n$ is said to be *open* if for every point $\mathbf{x} \in U$, there exists a ball $B_r(\mathbf{x}) \subset U$.

A set $S \subset \mathbb{R}^n$ is said to be *arcwise connected* if every pair of points in S may be joined by a continuous curve lying in S. An open arcwise connected set is called a *region*. A set S is *bounded* if there exists a ball $B_r(\mathbf{x})$ such that $S \subset B_r(\mathbf{x})$. A point $\mathbf{x}_0$ is an *accumulation point* or a *limit point* of a set S if every open ball $B_r(\mathbf{x}_0)$ contains a point $\mathbf{x} \neq \mathbf{x}_0$ belonging to S. The union of S and its accumulation points is called the *closure* of S. A set S is *dense* in a set T if the closure of S contains T; e.g. the rational numbers in $[0,1]$ form a dense subset of $[0,1]$.

Consider now a function f defined on an open set $U \subset \mathbb{R}^n$ with values in $\mathbb{R}^m$. Such a function may be described by m real-valued functions $f^1, \ldots, f^m$ defined on U. Let $\mathbf{x} = (x^1, \ldots, x^n) \in U$ and $f(\mathbf{x}) = (y^1, \ldots, y^m) \in \mathbb{R}^m$; then f is given by $y^1 = f^1(x^1, \ldots, x^n)$, $\ldots, y^m = f^m(x^1, \ldots, x^n)$. The functions $f^1, \ldots, f^m$ are called the *component functions* of f.

Example 3.1. *A parametric curve $x = x(t)$, $y = y(t)$ in the plane is a function $f : \mathbb{R} \longrightarrow \mathbb{R}^2$ whose component functions are $x = x(t)$ and $y = y(t)$.*

Example 3.2. *Inversion in the unit circle is a map $f : \mathbb{R}^2 - \{(0,0)\} \longrightarrow \mathbb{R}^2$ given by*

$$f(x,y) = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2} \right).$$

Example 3.3. *Rotation in the plane about the origin through an angle θ is a map $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by*

$$f(x,y) = (x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta).$$

Let $f : S \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be a function defined on a set S in $\mathbb{R}^n$, $\mathbf{x}_0$ an accumulation point of S, and $\mathbf{y}_0 \in \mathbb{R}^m$. Then

$$\lim_{\mathbf{x} \to \mathbf{x}_0} f(\mathbf{x}) = \mathbf{y}_0$$

means that for every ball $B_\epsilon(\mathbf{y}_0) \subset \mathbb{R}^m$ there exists a $B_\delta(\mathbf{x}_0) \subset \mathbb{R}^n$ such that $f(\mathbf{x}) \in B_\epsilon(\mathbf{y}_0)$ whenever $\mathbf{x} \neq \mathbf{x}_0$ is in $B_\delta(\mathbf{x}_0) \cap S$. It is easy to show that $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = \mathbf{y}_0$ if and only if each component function f^i of f has limit y_0^i. We say f is *continuous* at $\mathbf{x}_0$ if $f(\mathbf{x}_0)$ is defined and $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$; and we say f is *continuous* on a set S if it is continuous at each point of S.

Now let f be a real-valued function defined on an open set $U \subset \mathbb{R}^n$. Let $\mathbf{x}_0 = (x_0^1, \ldots, x_0^n) \in U$. By the *partial derivative* of f with respect to x^i at $\mathbf{x}_0$ we mean

$$\lim_{h\to 0} \frac{f(x_0^1,\ldots,x_0^{i-1},x_0^i+h,x_0^{i+1},\ldots,x_0^n)-f(x_0^1,\ldots,x_0^n)}{h}$$

provided this limit exists, and we denote the partial derivative by

$$\frac{\partial f}{\partial x^i}(\mathbf{x}_0).$$

Recall that in the calculus of one real variable, the existence of the derivative at a point implies the continuity of the function at the point. This is no longer true for functions of several variables, e.g. if $f(x,y) = x+y$ if $x=0$ or $y=0$ and 1 otherwise, then $\frac{\partial f}{\partial x}(0,0) = \frac{\partial f}{\partial y}(0,0) = 1$ but f is not continuous at $(0,0)$.

We shall use the following terminology. A function f is said to be of *class* C^0 at a point or on a set if it is continuous there. A function f is of *class* C^k if the partial derivatives of order k exist and are continuous, and a function f is of *class* C^∞ if it has continuous partial derivatives of all orders. Finally, f is of class C^ω or *real analytic* at a point if it is represented by its Taylor series on some neighborhood of the point. For example, $f(x) = e^{-1/x^2}, x \neq 0$, and $f(0) = 0$ is C^∞ at $x = 0$ but not analytic there. Informally we will sometimes say a function f is *smooth* if f has sufficiently many continuous derivatives for the purpose at hand.

Recall also the following theorem on the interchange of order of differentiation.

Theorem 3.1. *If f is of class C^2 on a neighborhood of $\mathbf{x}_0$, then*

$$\frac{\partial^2 f}{\partial x^j \partial x^i}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial x^i \partial x^j}(\mathbf{x}_0).$$

Let f be a real-valued function defined on an open set $U \subset \mathbb{R}^n$ and let $\mathbf{x}_0 \in U$. We say f is *differentiable* at $\mathbf{x}_0$ if there exists a linear map $df_{\mathbf{x}_0} : \mathbb{R}^n \longrightarrow \mathbb{R}$ such that

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0) - df_{\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)|}{|\mathbf{x}-\mathbf{x}_0|} = 0.$$

The map $df_{\mathbf{x}_0}$ is called the *differential* of f at $\mathbf{x}_0$. If f is differentiable at every point of U, we say f is *differentiable* on U. Notationally, for a vector $\mathbf{v} \in \mathbb{R}^n$ we write $df_{\mathbf{x}_0}(\mathbf{v})$ or just $df_{\mathbf{x}_0}\mathbf{v}$. The following theorems are proved in standard advanced calculus courses.

Theorem 3.2. *If f has a differential at $\mathbf{x}_0$, then the partial derivatives of f exist at $\mathbf{x}_0$ and*

$$df_{\mathbf{x}_0}(\mathbf{v}) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\mathbf{x}_0)v^i.$$

Theorem 3.3. *If f is of class C^1 at $\mathbf{x}_0$, then f has a differential at $\mathbf{x}_0$. If f has a differential at $\mathbf{x}_0$, then f is continuous at $\mathbf{x}_0$.*

As a corollary, note that if f is of class C^k, then f is also of all lower classes.

Theorem 3.4. *If $df_{\mathbf{x}} = 0$ holds throughout a region, then f is constant on the region.*

More generally if $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ and $\mathbf{x}_0 \in U$, we define the *differential* of f at $\mathbf{x}_0$ to be the map $df_{\mathbf{x}_0} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, if it exists, such that $df_{\mathbf{x}_0}$ is linear and

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0) - df_{\mathbf{x}_0}(\mathbf{x}-\mathbf{x}_0)|}{|\mathbf{x}-\mathbf{x}_0|} = 0.$$

Again the above theorems hold, and, in particular, if $f^1, \dots, f^m$ are the component functions of f, then

$$df_{\mathbf{x}_0}(\mathbf{v}) = \left(\sum_{i=1}^{n} \frac{\partial f^1}{\partial x^i}(\mathbf{x}_0)v^i, \dots, \sum_{i=1}^{n} \frac{\partial f^m}{\partial x^i}(\mathbf{x}_0)v^i\right).$$

In particular, the matrix of the linear map $df_{\mathbf{x}_0}$ with respect to the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$ is

$$\begin{pmatrix} \frac{\partial f^1}{\partial x^1}(\mathbf{x}_0) & \cdots & \frac{\partial f^1}{\partial x^n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \frac{\partial f^m}{\partial x^1}(\mathbf{x}_0) & \cdots & \frac{\partial f^m}{\partial x^n}(\mathbf{x}_0) \end{pmatrix}.$$

In the case that $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$, the determinant of the above matrix is called the *Jacobian* of f at $\mathbf{x}_0$ and denoted $J_f(\mathbf{x}_0)$. Clearly $df_{\mathbf{x}_0}$ is non-singular if and only if $J_f(\mathbf{x}_0) \neq 0$. The most important property of the Jacobian is the inverse function theorem:

Theorem 3.5. *Let f be a C^1, $\mathbb{R}^n$-valued function defined on an open set $U \subset \mathbb{R}^n$, and let $\mathbf{x}_0 \in U$. Suppose $J_f(\mathbf{x}_0) \neq 0$. Then there exist an open set V containing $\mathbf{x}_0$ and an open set W containing $f(\mathbf{x}_0)$ such* *that $f : V \longrightarrow W$ is one-to-one and onto. Moreover, $f^{-1} : W \longrightarrow V$ is of class C^1.*

Thus we see that the non-vanishing of the Jacobian (equivalently, the non-singularity of the differential) implies that the mapping is locally one-to-one. This is not a global result, however. For example, $f(x, y) = (e^x \cos y, e^x \sin y)$ is locally one-to-one since $J_f(x, y) = e^{2x} \neq 0$, but not globally one-to-one since f is periodic in y.

Now let $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^p$ and $g : V \subset \mathbb{R}^r \longrightarrow \mathbb{R}^n$ be mappings with $g(V) \subset U$. Then the composition $f \circ g$ is defined by $(f \circ g)(\mathbf{x}) = f(g(\mathbf{x}))$. If g is differentiable at $\mathbf{x}_0$ and f is differentiable at $\mathbf{y}_0 = g(\mathbf{x}_0)$, then $f \circ g$ is differentiable at $\mathbf{x}_0$ and we have the chain rule

$$d(f \circ g)_{\mathbf{x}_0} = df_{g(\mathbf{x}_0)} \circ dg_{\mathbf{x}_0}.$$

In terms of matrices we have (setting $\mathbf{y} = g(\mathbf{x})$ and $\mathbf{y}_0 = g(\mathbf{x}_0)$)

$$\begin{pmatrix} \frac{\partial (f\circ g)^1}{\partial x^1}(\mathbf{x}_0) & \cdots & \frac{\partial (f\circ g)^1}{\partial x^r}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \frac{\partial (f\circ g)^p}{\partial x^1}(\mathbf{x}_0) & \cdots & \frac{\partial (f\circ g)^p}{\partial x^r}(\mathbf{x}_0) \end{pmatrix}$$
$$= \begin{pmatrix} \frac{\partial f^1}{\partial y^1}(\mathbf{y}_0) & \cdots & \frac{\partial f^1}{\partial y^n}(\mathbf{y}_0) \\ \vdots & & \vdots \\ \frac{\partial f^p}{\partial y^1}(\mathbf{y}_0) & \cdots & \frac{\partial f^p}{\partial y^n}(\mathbf{y}_0) \end{pmatrix} \begin{pmatrix} \frac{\partial g^1}{\partial x^1}(\mathbf{x}_0) & \cdots & \frac{\partial g^1}{\partial x^r}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \frac{\partial g^n}{\partial x^1}(\mathbf{x}_0) & \cdots & \frac{\partial g^n}{\partial x^r}(\mathbf{x}_0) \end{pmatrix}.$$

If $r = n = p$, then $J_{(f\circ g)}(\mathbf{x}_0) = J_f(\mathbf{y}_0) J_g(\mathbf{x}_0)$; in particular, if $df_{\mathbf{y}_0}$ and $dg_{\mathbf{x}_0}$ are non-singular, so is $d(f \circ g)_{\mathbf{x}_0}$.

Recall that if $\mathbf{x}(t) = (x^1(t), \ldots, x^n(t))$ is a differentiable curve in $\mathbb{R}^n$, then $\mathbf{x}'(t) = (x^{1\prime}(t), \ldots, x^{n\prime}(t))$ is the *velocity vector* to the curve at the point $\mathbf{x}(t)$. If $\mathbf{x}'(t) \neq 0$, then the curve is *regular* at $\mathbf{x}(t)$ and $\mathbf{x}'(t)$ is called the *tangent vector* to the curve (or a *tangent vector*, as appropriate). Suppose $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is differentiable and that $df_{\mathbf{x}}$ is non-singular for every $\mathbf{x} \in U$. If $\mathbf{x}(t) \in U$ for all t in some interval I, then $\mathbf{y}(t) = f(\mathbf{x}(t))$ is a curve in $\mathbb{R}^n$ defined on I and its tangent is given by

$$\mathbf{y}'(t) = \left(\sum_{i=1}^n \frac{\partial f^1}{\partial x^i} x^{i\prime}(t), \ldots, \sum_{i=1}^n \frac{\partial f^n}{\partial x^i} x^{i\prime}(t) \right) = df_{\mathbf{x}(t)}(\mathbf{x}'(t)).$$

Thus the linear map $df_{\mathbf{x}(t)}$ maps the tangent vector of a curve at $\mathbf{x}(t)$ to the tangent vector of the image curve under the mapping f at the corresponding point.

Also, if s denotes arc length along the curve from some point, then

$$\frac{ds}{dt} = |\mathbf{x}'(t)| = \sqrt{\sum_{i=1}^{n}(x^{i\prime}(t))^2}.$$

In particular, for a regular curve, the length of the curve from $\mathbf{x}(t_1)$ to $\mathbf{x}(t_2)$ is $\int_{t_1}^{t_2} |\mathbf{x}'(t)|dt$.

Again suppose that $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$. We define the *second order differential* $d^2 f_{\mathbf{x}_0} : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$ in terms of the component functions of f by

$$d^2 f_{\mathbf{x}_0}(\mathbf{v},\mathbf{w}) = \left(\sum_{i,j=1}^{n} \frac{\partial^2 f^1}{\partial x^j \partial x^i}(\mathbf{x}_0)v^i w^j, \dots, \sum_{i,j=1}^{n} \frac{\partial^2 f^m}{\partial x^j \partial x^i}(\mathbf{x}_0)v^i w^j\right).$$

$d^2 f$ is $d(df)$ in the following sense. For a fixed vector $\mathbf{v} \in \mathbb{R}^n$ consider the map $df(\mathbf{v}) : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ defined by $df(\mathbf{v})(\mathbf{x}) = df_{\mathbf{x}}(\mathbf{v})$; then

$$d^2 f_{\mathbf{x}_0}(\mathbf{v},\mathbf{w}) = d(df(\mathbf{v}))_{\mathbf{x}_0}(\mathbf{w}).$$

Similarly, we define differentials of order $3, 4, \dots$; e.g.

$$d^3 f_{\mathbf{x}_0} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

is defined by

$$d^3 f_{\mathbf{x}_0}(\mathbf{u},\mathbf{v},\mathbf{w}) = \left(\sum_{i,j,k=1}^{n} \frac{\partial^3 f^1}{\partial x^k \partial x^j \partial x^i}(\mathbf{x}_0)u^i v^j w^k, \dots,\right.$$
$$\left.\sum_{i,j,k=1}^{n} \frac{\partial^3 f^m}{\partial x^k \partial x^j \partial x^i}(\mathbf{x}_0)u^i v^j w^k\right).$$

Note in particular that if f is of class C^k, then

$$d^2 f_{\mathbf{x}_0}(\mathbf{v}_1,\mathbf{v}_2), d^3 f_{\mathbf{x}_0}(\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3), \dots, d^k f_{\mathbf{x}_0}(\mathbf{v}_1,\dots,\mathbf{v}_k)$$

are symmetric in the $\mathbf{v}_i$'s.

Finally we state Taylor's theorem for functions of several variables.

Theorem 3.6. *Suppose $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is of class C^k. Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be points in U such that the line segment $l(\mathbf{x}, \mathbf{y})$ joining* $\mathbf{x}$ *and* $\mathbf{y}$ *lies in U. Then there exists a point $\mathbf{z} \in l(\mathbf{x}, \mathbf{y})$ such that*

$$f(\mathbf{y}) = f(\mathbf{x}) + \sum_{i=1}^{k-1} \frac{1}{i!} d^i f_{\mathbf{x}}(\mathbf{y} - \mathbf{x}) + \frac{1}{k!} d^k f_{\mathbf{z}}(\mathbf{y} - \mathbf{x}).$$

In the special case $k = 1$, this is just the mean value theorem. For example, if $f : U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}$, then there is a point $(\xi, \eta) \in l((x, y), (x_0, y_0))$ such that

$$f(x, y) - f(x_0, y_0) = \frac{\partial f}{\partial x}(\xi, \eta)(x - x_0) + \frac{\partial f}{\partial y}(\xi, \eta)(y - y_0).$$

Exercises

1. Let U be an open set in $\mathbb{R}^n$ and f an $\mathbb{R}^m$-valued function defined on U. Show that f is continuous if and only if each component function $f^i : U \longrightarrow \mathbb{R}$ is continuous.

2. Let

$$f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}$$

for $(x, y) \neq (0, 0)$, and let $f(0, 0) = 0$. Show that

$$\frac{\partial^2 f}{\partial y \partial x}(0, 0) \neq \frac{\partial^2 f}{\partial x \partial y}(0, 0).$$

3. Let $g(x) = x^2 \sin \frac{1}{x}$ for $x \neq 0$ and $g(0) = 0$; let $f(x, y) = g(x) + g(y)$. Show that f has a differential at $(0, 0)$ but is not of class C^1 there.

4. Suppose $f(x, y) = 0$ for (x, y) in the first and third quadrants and on the axes. In the second quadrant suppose that $f(x, y) = -x$ if $y \geq -x$ and $f(x, y) = y$ if $y \leq -x$. In the fourth quadrant suppose $f(x, y) = y$ if $y \geq -x$ and $f(x, y) = -x$ if $y \leq -x$. Show that both first partial derivatives exist at $(0, 0)$, but that f does not have a differential at $(0, 0)$.

5. We noted in Example 3.2 that inversion in the unit circle is given by

$$f(x,y) = \left(\frac{x}{x^2+y^2}, \frac{y}{x^2+y^2}\right), \qquad (x,y) \neq (0,0).$$

Find $df_{(x,y)}$ and $J_f(x,y)$.

6. Suppose $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ is differentiable and $df_{(x,y)}$ for all (x,y) is given by $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$, where a and b are constants such that $a^2+b^2=1$. Find f, and describe it geometrically.

7. Prove the following extension of Exercise 2 of Section 2.7. Given two points z_1, z_2 in the upper half plane, a unit vector $\mathbf{u}$ at z_1 and a unit vector $\mathbf{v}$ at z_2, show that there exists $T \in H$ such that $T(z_1) = z_2$ and $dT_{z_1}(\mathbf{u}) = \mathbf{v}$.

8. In Exercise 7, remove the hypothesis that $\mathbf{u}$ and $\mathbf{v}$ are unit vectors, but assume the conclusion that $dT_{z_1}(\mathbf{u}) = \mathbf{v}$. Show that $\mathbf{u}$ and $\mathbf{v}$ have the same length.

9. Similarly to Exercise 7, prove the following two-point homogeneity property. Given points z_1, w_1, z_2, w_2 in the upper half plane such that the distance between z_1 and w_1 is equal to the distance between z_2 and w_2, show that there exists $T \in H$ such that $T(z_1) = z_2$ and $T(w_1) = w_2$.

3.2. Inner products

For two points $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we define their *inner product* $\langle \mathbf{v}, \mathbf{w} \rangle$ (often called scalar product or dot product, $\mathbf{v}\cdot\mathbf{w}$) by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v^i w^i.$$

Clearly $\langle \mathbf{v}, \mathbf{v} \rangle = |\mathbf{v}|^2$, the square of the distance of $\mathbf{v}$ from the origin, and similarly $|\mathbf{v} - \mathbf{w}| = \langle \mathbf{v} - \mathbf{w}, \mathbf{v} - \mathbf{w} \rangle^{1/2}$ is the distance between $\mathbf{v}$ and $\mathbf{w}$. Note that the inner product is symmetric and bilinear, i.e. $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$, $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$, $a, b \in \mathbb{R}$, and similarly in the second variable.

Now apply the law of cosines to the following triangle (Figure 3.1).

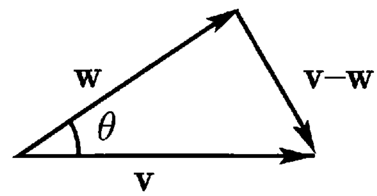

Figure 3.1

Then $|\mathbf{v}-\mathbf{w}|^2 = |\mathbf{v}|^2 + |\mathbf{w}|^2 - 2|\mathbf{v}||\mathbf{w}|\cos\theta$, but

$$|\mathbf{v}-\mathbf{w}|^2 = \langle \mathbf{v}-\mathbf{w}, \mathbf{v}-\mathbf{w}\rangle = \langle \mathbf{v},\mathbf{v}\rangle - 2\langle \mathbf{v},\mathbf{w}\rangle + \langle \mathbf{w},\mathbf{w}\rangle,$$

and hence the angle θ between two vectors $\mathbf{v}, \mathbf{w}$ is given by

$$\cos\theta = \frac{\langle \mathbf{v},\mathbf{w}\rangle}{|\mathbf{v}||\mathbf{w}|}.$$

In particular, if $\mathbf{v}$ and $\mathbf{w}$ are non-zero, then $\langle \mathbf{v},\mathbf{w}\rangle = 0$ if and only if $\mathbf{v}$ and $\mathbf{w}$ are orthogonal.

As before, let U be an open set in $\mathbb{R}^n$, and consider two differentiable maps $f, g : U \longrightarrow \mathbb{R}^m$. Then $\langle f(\mathbf{x}), g(\mathbf{x})\rangle = \sum_{j=1}^m f^j(\mathbf{x})g^j(\mathbf{x})$ defines a map $\langle f, g\rangle : U \longrightarrow \mathbb{R}$. Note that

$$\frac{\partial}{\partial x^i}\langle f(\mathbf{x}), g(\mathbf{x})\rangle = \sum_{j=1}^m \left(\frac{\partial f^j}{\partial x^i}(\mathbf{x})g^j(\mathbf{x}) + f^j(\mathbf{x})\frac{\partial g^j}{\partial x^i}(\mathbf{x})\right).$$

Consequently

$$\begin{aligned}
d\langle f, g\rangle_{\mathbf{x}}(\mathbf{v}) &= \sum_{i=1}^n \frac{\partial}{\partial x^i}\langle f(\mathbf{x}), g(\mathbf{x})\rangle v^i \\
&= \sum_{i=1}^n \sum_{j=1}^m \left(\frac{\partial f^j}{\partial x^i}(\mathbf{x})v^i g^j(\mathbf{x}) + f^j(\mathbf{x})\frac{\partial g^j}{\partial x^i}(\mathbf{x})v^i\right) \\
&= \langle df_{\mathbf{x}}(\mathbf{v}), g(\mathbf{x})\rangle + \langle f(\mathbf{x}), dg_{\mathbf{x}}(\mathbf{v})\rangle.
\end{aligned}$$

We shall also have occasion to differentiate inner products of differentials, e.g.,

$$\begin{aligned}
&d\langle df(\mathbf{u}), dg(\mathbf{v})\rangle_{\mathbf{x}}(\mathbf{w}) \\
&= \langle d^2 f_{\mathbf{x}}(\mathbf{u},\mathbf{w}), dg_{\mathbf{x}}(\mathbf{v})\rangle + \langle df_{\mathbf{x}}(\mathbf{u}), d^2 g_{\mathbf{x}}(\mathbf{v},\mathbf{w})\rangle.
\end{aligned}$$

Recall that in a vector space $\mathbb{V}$ of dimension n, n linearly independent vectors constitute a *basis* of $\mathbb{V}$. Moreover, given a basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, every vector $\mathbf{v} \in \mathbb{V}$ has a unique expression of the form

$\mathbf{v} = \sum_{i=1}^{n} v^i \mathbf{e}_i$, $v^i \in \mathbb{R}$. Also, given p, $p < n$, linearly independent vectors $\mathbf{v}_1, \ldots, \mathbf{v}_p$, we can find a basis of $\mathbb{V}$ such that $\mathbf{v}_1, \ldots, \mathbf{v}_p$ belong to the basis.

Let L be a linear functional on $\mathbb{V}$, i.e., L is a linear map of $\mathbb{V}$ into $\mathbb{R}$. Clearly if L vanishes on the elements of a basis, $L = 0$.

Regarding $\mathbb{R}^n$ as a vector space, let B be a symmetric bilinear form on $\mathbb{R}^n$, i.e. $B(\mathbf{v}, \mathbf{w}) \in \mathbb{R}$, $B(\mathbf{v}, \mathbf{w}) = B(\mathbf{w}, \mathbf{v})$, $B(a\mathbf{u} + b\mathbf{v}, \mathbf{w}) = aB(\mathbf{u}, \mathbf{w}) + bB(\mathbf{v}, \mathbf{w})$, $a, b \in \mathbb{R}$, and hence, of course, $B(\mathbf{u}, a\mathbf{v} + b\mathbf{w}) = aB(\mathbf{u}, \mathbf{v}) + bB(\mathbf{u}, \mathbf{w})$.

For later use we prove the following theorem.

Theorem 3.7. *Let B be a symmetric bilinear form on $\mathbb{R}^n$ which vanishes on all orthogonal pairs of vectors. Then there exists a real number α such that*

$$B(\mathbf{v}, \mathbf{w}) = \alpha\langle \mathbf{v}, \mathbf{w}\rangle.$$

Proof. For fixed $\mathbf{v} \in \mathbb{R}^n, \mathbf{v} \neq 0$, let

$$L(\mathbf{w}) = B(\mathbf{v}, \mathbf{w}) - \frac{B(\mathbf{v}, \mathbf{v})}{\langle \mathbf{v}, \mathbf{v}\rangle}\langle \mathbf{v}, \mathbf{w}\rangle.$$

Then $L(\mathbf{v}) = 0$ and $L(\mathbf{w}) = 0$ for every $\mathbf{w}$ orthogonal to $\mathbf{v}$; thus L vanishes on a basis and hence $L = 0$. Similarly, holding $\mathbf{w}$ fixed,

$$B(\mathbf{w}, \mathbf{v}) - \frac{B(\mathbf{w}, \mathbf{w})}{\langle \mathbf{w}, \mathbf{w}\rangle}\langle \mathbf{w}, \mathbf{v}\rangle = 0$$

for all $\mathbf{v}$. Therefore

$$\frac{B(\mathbf{v}, \mathbf{v})}{\langle \mathbf{v}, \mathbf{v}\rangle} = \frac{B(\mathbf{w}, \mathbf{w})}{\langle \mathbf{w}, \mathbf{w}\rangle}$$

for all non-zero $\mathbf{v}, \mathbf{w}$; calling this common value α, we have $B(\mathbf{v}, \mathbf{w}) = \alpha\langle \mathbf{v}, \mathbf{w}\rangle$. □

Exercises

1. Let B be a symmetric bilinear form and define a quadratic form Q by $Q(\mathbf{v}) = B(\mathbf{v}, \mathbf{v})$. Show that Q determines B.

2. Let U be an open set in $\mathbb{R}^n$ and f a C^1 function on U. Define the *gradient* of f at $\mathbf{x} \in U$ to be the unique vector $\mathbf{grad} f_{\mathbf{x}}$ such that

$$\langle \mathbf{grad} f_{\mathbf{x}}, \mathbf{v}\rangle = df_{\mathbf{x}}(\mathbf{v})$$

for all $\mathbf{v}$. Let $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be the standard basis on $\mathbb{R}^n$; show that this definition of gradient is equivalent to

$$\mathbf{grad} f_{\mathbf{x}} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(\mathbf{x})\mathbf{e}_i.$$

3. Consider a level surface of a smooth function f in $\mathbb{R}^3$, i.e. a surface given by the equation $f(x, y, z) = c$, c a constant. Show that $\mathbf{grad} f$ is normal to the surface.

4. Note the dependence on the inner product in the definition of gradient in the Exercise 2. Let B be a positive definite symmetric bilinear form. Define the *gradient* of f with respect to the inner product given by B by $B(\mathbf{grad}_B f, \mathbf{v}) = df(\mathbf{v})$. Suppose that on $\mathbb{R}^3$ with standard basis $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ an inner product is given by

$$\big(B(\mathbf{e}_i, \mathbf{e}_j)\big) = \begin{pmatrix} 1+y^2 & 0 & -y \\ 0 & 1 & 0 \\ -y & 0 & 1 \end{pmatrix}.$$

If $\mathbf{W} = \mathbf{grad}_B f = W^1\mathbf{e}_1 + W^2\mathbf{e}_2 + W^3\mathbf{e}_3$, find W^1, W^2, W^3 in terms of the partial derivatives of f and the variable y.

3.3. Conformal maps

We have already used the notion of a conformal map as an angle-preserving transformation. Let us now see how we may describe this notion analytically.

Theorem 3.8. *Let f be a one-to-one mapping of an open set U in $\mathbb{R}^n$ onto $f(U)$ such that $df_{\mathbf{x}}$ is non-singular for all $\mathbf{x} \in U$. Then f is conformal if and only if, for all vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$,*

$$\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w}\rangle = e^{2\sigma(\mathbf{x})}\langle \mathbf{v}, \mathbf{w}\rangle$$

for some real-valued function σ on U.

Proof. Suppose $\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w}\rangle = e^{2\sigma}\langle \mathbf{v}, \mathbf{w}\rangle$. If θ is the angle between $\mathbf{v}$ and $\mathbf{w}$ and ϕ the angle between $df_{\mathbf{x}}\mathbf{v}$ and $df_{\mathbf{x}}\mathbf{w}$, then, since $|df_{\mathbf{x}}\mathbf{v}| = e^{\sigma}|\mathbf{v}|$,

$$\cos\phi = \frac{\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w}\rangle}{|df_{\mathbf{x}}\mathbf{v}||df_{\mathbf{x}}\mathbf{w}|} = \frac{e^{2\sigma}\langle \mathbf{v}, \mathbf{w}\rangle}{e^{\sigma}|\mathbf{v}|e^{\sigma}|\mathbf{w}|} = \cos\theta,$$

so that f is angle preserving.

Conversely, suppose that f is angle preserving; then, at each point $\mathbf{x} \in U$, $B(\mathbf{v}, \mathbf{w}) = \langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w} \rangle$ is a symmetric bilinear form which vanishes on orthogonal pairs $\mathbf{v}, \mathbf{w}$. Thus by Theorem 3.7 $\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w} \rangle$ is proportional to $\langle \mathbf{v}, \mathbf{w} \rangle$ at each point. Again since f is angle preserving, $\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w} \rangle$ and $\langle \mathbf{v}, \mathbf{w} \rangle$ have the same sign; thus, denoting the proportionality factor by $e^{2\sigma(\mathbf{x})}$, we have $\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w} \rangle = e^{2\sigma(\mathbf{x})} \langle \mathbf{v}, \mathbf{w} \rangle$ as desired. $\square$

In the above theorem the function e^{σ} is called the *characteristic function* of the conformal mapping f. Note that $e^{\sigma(\mathbf{x})} = |df_{\mathbf{x}}\mathbf{v}|/|\mathbf{v}|$ for all $\mathbf{v} \neq 0$, and therefore we always regard a conformal map as being non-singular in its domain of conformality.

EXERCISES

1. Suppose that $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ and $g : V \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ with $g(V) \subset U$ are both conformal with characteristic functions e^{σ} and e^{ρ} respectively. Show that $f \circ g$ is conformal with characteristic function $e^{\sigma \circ g + \rho}$.

2. If $f : U \subset \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is conformal with characteristic function e^{σ}, show that f^{-1} is conformal with characteristic function $e^{-\sigma \circ f^{-1}}$.

3. Show that the map $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ given by $f(x, y) = (x^2 - y^2, 2xy)$ is conformal except at the origin. Find the characteristic function.

Chapter 4

Conformal Maps in the Plane

4.1. Complex function theory

We have seen that the extended Möbius transformations (homographies and anti-homographies) are conformal mappings. However we shall see in this chapter that there are many more conformal maps in the plane. For this purpose we must first study a little complex function theory.

Let f be a complex-valued function of a complex variable $z = x + iy$. We denote by u and v the real and imaginary parts of f, i.e., $f(z) = u(x, y) + iv(x, y)$. In this way f is a map, $f : U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, with component functions u and v, where U will always be an open, arcwise connected set. We say that f as a function of z is *continuous* if f as a map is continuous. However we will see that the notion of the derivative of f is stronger than its differentiablility as a map given in the last chapter. The *derivative* of f at $z_0 \in U$, $f'(z_0)$, is defined by

$$f'(z_0) = \lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h}$$

provided the limit exists (and, of course, the limit is defined in terms of $h \in \mathbb{C}$ being in a neighborhood of $0 \in \mathbb{C}$). We say that f is *holomorphic* or *analytic* at z_0 if f has a derivative at every point of some neighborhood of z_0. In particular, if f has a derivative at every point of U, we say f is *holomorphic* on U. Note that since the limit is independent of how h tends to 0, the existence of $f'(z_0)$ implies the continuity of f at z_0.

Since we will not have many applications of differentiation to special cases, we simply remark without proof that the usual rules for differentiation hold for complex differentiation. The reader is referred to any standard text for details, e.g. [**8**], [**13**].

Theorem 4.1. *Let f be defined and continuous on an open set U, and let $z_0 = x_0 + iy_0 \in U$. Suppose $f'(z_0)$ exists. Then*

$$\frac{\partial u}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0), \quad \frac{\partial u}{\partial y}(x_0, y_0) = -\frac{\partial v}{\partial x}(x_0, y_0).$$

Conversely, if $u(x, y)$ and $v(x, y)$ are of class C^1 on a neighborhood of z_0 and their partial derivatives satisfy these equations at z_0, then $f(z) = u(x, y) + iv(x, y)$ has a derivative at z_0.

Proof. Supposing $f'(z_0)$ exists and taking h real, we get

$$\frac{f(z_0 + h) - f(z_0)}{h} = \frac{u(x_0 + h, y_0) - u(x_0, y_0)}{h} + i\frac{v(x_0 + h, y_0) - v(x_0, y_0)}{h}$$

and hence

$$f'(z_0) = \frac{\partial u}{\partial x}(x_0, y_0) + i\frac{\partial v}{\partial x}(x_0, y_0).$$

Similarly, taking h imaginary, say $h = ik$, k real,

$$\frac{f(z_0 + h) - f(z_0)}{h} = \left(\frac{1}{i}\right)\frac{u(x_0, y_0 + k) - u(x_0, y_0)}{k} + \frac{v(x_0, y_0 + k) - v(x_0, y_0)}{k},$$

and hence

$$f'(z_0) = -i\frac{\partial u}{\partial y}(x_0, y_0) + \frac{\partial v}{\partial y}(x_0, y_0).$$

Now comparing these two expressions for $f'(z_0)$ we see that

$$\frac{\partial u}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0), \quad \frac{\partial u}{\partial y}(x_0, y_0) = -\frac{\partial v}{\partial x}(x_0, y_0);$$

these equations are called the *Cauchy-Riemann equations.*

Conversely, suppose that u and v are of class C^1 on the disk $B_\epsilon(z_0)$ of radius ϵ about z_0. Now $h \longrightarrow 0$ corresponds to $z \longrightarrow z_0$, and it suffices to consider $z \in B_\epsilon(z_0)$. Write h as $a+ib$. Then $\frac{f(z_0+h)-f(z_0)}{h}$ becomes

$$\begin{aligned}
&\frac{u(x_0+a,y_0+b)-u(x_0,y_0)}{a+ib}+i\frac{v(x_0+a,y_0+b)-v(x_0,y_0)}{a+ib}\\
&\quad=\frac{a}{a+ib}\Big(\frac{u(x_0+a,y_0+b)-u(x_0,y_0+b)}{a}\\
&\qquad\qquad+i\frac{v(x_0+a,y_0+b)-v(x_0,y_0+b)}{a}\Big)\\
&\quad+\frac{b}{a+ib}\Big(\frac{u(x_0,y_0+b)-u(x_0,y_0)}{b}+i\frac{v(x_0,y_0+b)-v(x_0,y_0)}{b}\Big).
\end{aligned}$$

By the mean value theorem there are points (ξ_1, y_0+b) and (ξ_2, y_0+b) on the line segment $l((x_0+a,y_0+b),(x_0,y_0+b))$ and points (x_0,η_1) and (x_0,η_2) on $l((x_0,y_0+b),(x_0,y_0))$ such that $\frac{f(z_0+h)-f(z_0)}{h}$ becomes

$$\begin{aligned}
&\frac{a}{a+ib}\Big(\frac{\partial u}{\partial x}(\xi_1,y_0+b)+i\frac{\partial v}{\partial x}(\xi_2,y_0+b)\Big)\\
&\qquad+\frac{b}{a+ib}\Big(\frac{\partial u}{\partial y}(x_0,\eta_1)+i\frac{\partial v}{\partial y}(x_0,\eta_2)\Big).
\end{aligned}$$

Now, since u and v are of class C^1, we can write

$$\frac{\partial u}{\partial x}(\xi_1,y_0+b)=\frac{\partial u}{\partial x}(x_0,y_0)+\epsilon_1$$

where $\epsilon_1 \longrightarrow 0$ as $h \longrightarrow 0$, and similar expressions for the other partial derivatives, giving

$$\begin{aligned}
\frac{f(z_0+h)-f(z_0)}{h}&=\frac{a}{a+ib}\Big(\frac{\partial u}{\partial x}(x_0,y_0)+\epsilon_1+i\frac{\partial v}{\partial x}(x_0,y_0)+i\epsilon_2\Big)\\
&\quad+\frac{b}{a+ib}\Big(\frac{\partial u}{\partial y}(x_0,y_0)+\epsilon_3+i\frac{\partial v}{\partial y}(x_0,y_0)+i\epsilon_4\Big).
\end{aligned}$$

Thus, using the Cauchy-Riemann equations at (x_0,y_0), we have

$$\begin{aligned}
&\frac{f(z_0+h)-f(z_0)}{h}\\
&\quad=\frac{\partial u}{\partial x}(x_0,y_0)+i\frac{\partial v}{\partial x}(x_0,y_0)+\frac{a\epsilon_1+ia\epsilon_2+b\epsilon_3+ib\epsilon_4}{a+ib}.
\end{aligned}$$

Now $|a| \leq |a+ib|$ and $|b| \leq |a+ib|$, so that

$$\left|\frac{a\epsilon_1 + ia\epsilon_2 + b\epsilon_3 + ib\epsilon_4}{a+ib}\right| \leq |\epsilon_1 + i\epsilon_2| + |\epsilon_3 + i\epsilon_4|,$$

and hence, taking the limit as $h \longrightarrow 0$, we have

$$f'(z_0) = \frac{\partial u}{\partial x}(x_0, y_0) + i\frac{\partial v}{\partial x}(x_0, y_0).$$

$\square$

Using Theorem 4.1, we note that the function $f(z) = \bar{z} = x - iy$ is not holomorphic. However, if $g(z) = \phi(x,y) + i\psi(x,y)$ is holomorphic, then $f(z) \equiv u(x,y) + iv(x,y) = \overline{g(z)}$ satisfies $\frac{\partial u}{\partial x} = -\frac{\partial v}{\partial y}$ and $\frac{\partial u}{\partial y} = \frac{\partial v}{\partial x}$. Conversely, if $f(z) = u(x,y) + iv(x,y)$ is of class C^1 and satisfies these equations, then $f(z)$ is the conjugate of some holomorphic function. Indeed, setting $\phi(x,y) = u(x,y)$ and $\psi(x,y) = -v(x,y)$, we find that $\frac{\partial \phi}{\partial x} = \frac{\partial \psi}{\partial y}$ and $\frac{\partial \phi}{\partial y} = -\frac{\partial \psi}{\partial x}$. Thus $g = \phi + i\psi = \bar{f}$ is holomorphic, and hence $f = \bar{g}$. We say a function f is *anti-holomorphic* if f is the conjugate of a holomorphic function.

Exercises

1. Let $f(z) = z^5/|z|^4$ for $z \neq 0$ and $f(0) = 0$. Show that f satisfies the Cauchy-Riemann equations at $(0,0)$ but that $f'(0)$ does not exist; compare with Theorem 4.1.

2. If f is holomorphic on U, show that, as a map of $U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, its Jacobian is $|f'(z)|^2$.

3. If f is anti-holomorphic on U, show that, as a map of $U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, its Jacobian is non-positive.

4. Find the Jacobian of a linear fractional transformation T, and show that it is identically 1 if and only if $T(z) = e^{i\theta}z + b$.

4.2. Abundance of conformal maps

We now show the relation between conformality and holomorphy and anti-homorphy, and hence the abundance of conformal maps in the plane. Recall that we have taken conformality to mean the preservation of angle not including the sense or orientation of the angle. Many complex analysts require that conformality includes the preservation of sense, and hence they would not include the anti-holomorphic case in the next theorem.

Theorem 4.2. *Let $f : U \subset \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be of class C^1 with non-vanishing Jacobian. Then f as a map is conformal if and only if f, as a function of the complex variable z, is holomorphic or anti-holomorphic.*

Proof. By the inverse function theorem f is locally one-to-one. We must show that f satisfies

$$\langle df_z\mathbf{v}, df_z\mathbf{w}\rangle = e^{2\sigma(z)}\langle \mathbf{v}, \mathbf{w}\rangle$$

for some function σ and for all vectors $\mathbf{v}$ and $\mathbf{w}$ if and only if the component functions u and v of f satisfy

$$\frac{\partial u}{\partial x} = \pm\frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = \mp\frac{\partial v}{\partial x}. \tag{4.1}$$

Recall that $df_z\mathbf{v}$ is given by

$$\begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} \frac{\partial u}{\partial x}v^1 + \frac{\partial u}{\partial y}v^2 \\ \frac{\partial v}{\partial x}v^1 + \frac{\partial v}{\partial y}v^2 \end{pmatrix}$$

and $\langle df_z\mathbf{v}, df_z\mathbf{w}\rangle = e^{2\sigma}\langle \mathbf{v}, \mathbf{w}\rangle$ becomes

$$\Big(\frac{\partial u}{\partial x}v^1 + \frac{\partial u}{\partial y}v^2\Big)\Big(\frac{\partial u}{\partial x}w^1 + \frac{\partial u}{\partial y}w^2\Big) + \Big(\frac{\partial v}{\partial x}v^1 + \frac{\partial v}{\partial y}v^2\Big)\Big(\frac{\partial v}{\partial x}w^1 + \frac{\partial v}{\partial y}w^2\Big)$$

$$= e^{2\sigma}(v^1w^1 + v^2w^2). \tag{4.2}$$

If this holds for all $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^2$, then, choosing $\mathbf{v} = \mathbf{w} = \binom{1}{0}$ and $\mathbf{v} = \mathbf{w} = \binom{0}{1}$, we have

$$\Big(\frac{\partial u}{\partial x}\Big)^2 + \Big(\frac{\partial v}{\partial x}\Big)^2 = e^{2\sigma}, \quad \Big(\frac{\partial u}{\partial y}\Big)^2 + \Big(\frac{\partial v}{\partial y}\Big)^2 = e^{2\sigma}. \tag{4.3}$$

Similarly, setting $\mathbf{v} = \binom{1}{0}$ and $\mathbf{w} = \binom{0}{1}$ yields

$$\frac{\partial u}{\partial x}\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\frac{\partial v}{\partial y} = 0. \tag{4.4}$$

Equating the two expressions in (4.3), multiplying by $\left(\frac{\partial v}{\partial y}\right)^2$, and using (4.4), we have

$$\left(\frac{\partial u}{\partial y}\right)^2\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial v}{\partial y}\right)^4=\left(\frac{\partial u}{\partial x}\right)^2\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial v}{\partial x}\right)^2\left(\frac{\partial v}{\partial y}\right)^2$$
$$=\left(\frac{\partial u}{\partial x}\right)^2\left(\frac{\partial v}{\partial y}\right)^2+\left(\frac{\partial u}{\partial x}\right)^2\left(\frac{\partial u}{\partial y}\right)^2.$$

Therefore $\frac{\partial u}{\partial x}=\pm\frac{\partial v}{\partial y}$. If $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}$, (4.4) gives $\frac{\partial u}{\partial y}=-\frac{\partial v}{\partial x}$ unless $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}=0$, in which case (4.3) gives $\frac{\partial v}{\partial x}=\pm\frac{\partial u}{\partial y}$. Similarly, if $\frac{\partial u}{\partial x}=-\frac{\partial v}{\partial y}$, then $\frac{\partial u}{\partial y}=\frac{\partial v}{\partial x}$ unless $\frac{\partial u}{\partial x}=\frac{\partial v}{\partial y}=0$, and then $\frac{\partial v}{\partial x}=\pm\frac{\partial u}{\partial y}$. Consequently, conformality implies holomorphy or anti-holomorphy. Conversely, substitution of (4.1) into the left side of (4.2) yields the right side of (4.2) with $e^{2\sigma}=\left(\frac{\partial u}{\partial x}\right)^2+\left(\frac{\partial u}{\partial y}\right)^2$. □

Thus we see that the plane is rich in conformal maps; in particular, every holomorphic function on an open set U with non-vanishing derivative is a conformal map. To note this abundance more dramatically we state the celebrated Riemann Mapping Theorem.

We say that a set $U\subset\mathbb{R}^2$ is *simply-connected* if every simple closed curve lying in U is contractible within U (or equivalently has its interior in U).

Theorem 4.3. *Let U be an open, connected, simply-connected set $\subset\mathbb{R}^2$ other than $\mathbb{R}^2$ itself. Then there exists a one-to-one conformal map f of U onto the open unit disk (Figure 4.1).*

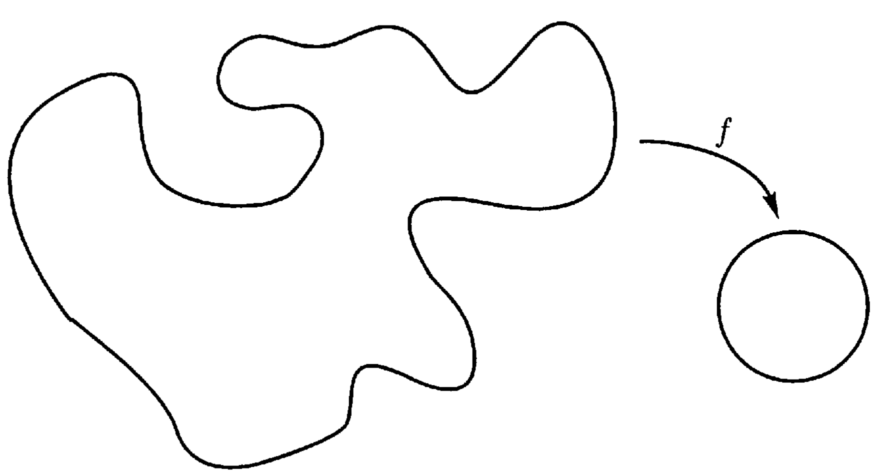

Figure 4.1

Of course f^{-1} is a conformal map of the open unit disk onto U. Thus, if U and V are two open, connected and simply-connected sets in the plane, neither being the plane itself, let f and g be conformal maps of U and V respectively onto the unit disk. Then $g^{-1} \circ f$ is a conformal map of U onto V.

The theorem says nothing about the extension of the mapping to the boundary of U. If the boundary of U is a piecewise differentiable curve, then such an extension is possible such that f is a continuous map, but it is not necessarily conformal on the boundary. For example, if U is a square and V is a rectangle which is not a square, then $g^{-1} \circ f$ as above is a conformal map of U onto V obtained by mapping through the unit disk. While the boundary is mapped to the boundary, it is known that there does not exist a conformal map of a square onto a non-square rectangle which maps the vertices to the vertices ([**21**], pp.14-15).

Exercises

1. Show that $w = \frac{4}{(z+1)^2}$ is a conformal map of the open unit disk onto the "exterior" of the parabola $v^2 = 4(1-u)$ (Figure 4.2). Show also that the images of the four concircular points $\frac{1}{2}$, $\frac{i}{2}$, $\frac{-1}{2}$, $\frac{-i}{2}$ are non-concircular.

2. A more dramatic example is given by $w = e^{-1/z}$ as a locally conformal map defined on $\mathbb{R}^2 - \{(0,0)\}$. Draw the images of circles centered at $z = 0$ with radii 1, $\frac{1}{2}$, $\frac{1}{4}$. (The reader with some knowledge of complex function theory will recognize the point $z = 0$ as an "essential singularity" of the function.)

3. Show that the Koebe function

$$f(z) = \frac{z}{(1-z)^2} = \frac{1}{4}\left[\left(\frac{1+z}{1-z}\right)^2 - 1\right]$$

defines a one-to-one mapping of the open unit disk onto the plane slit along the ray $(-\infty, -\frac{1}{4}]$.

4. Identitfying a vector $\mathbf{v}$ given in classical notation by $\mathbf{v} = v^1\mathbf{i} + v^2\mathbf{j}$ with the complex number $v = v^1 + iv^2$, show that the inner product is given by $\langle \mathbf{v}, \mathbf{w} \rangle = \Re v\bar{w}$.

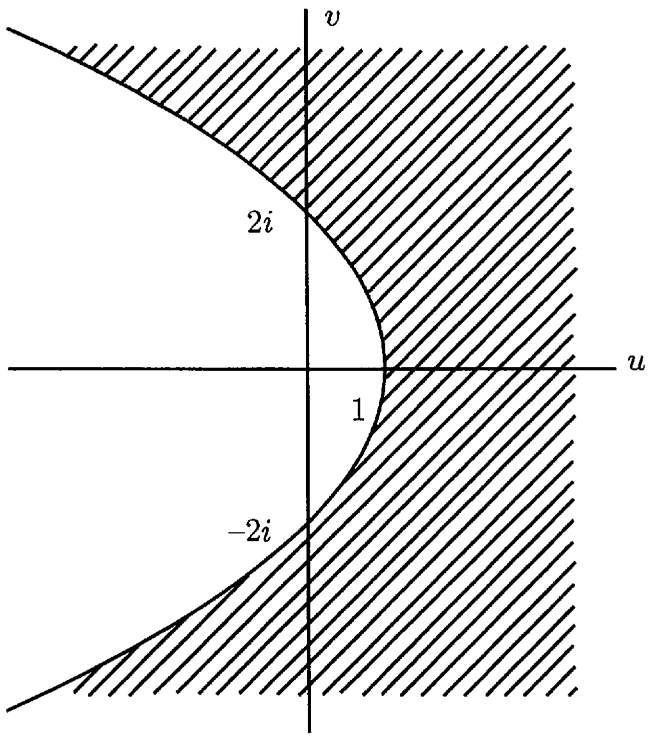

Figure 4.2

5. Let f be a holomorphic function with non-vanishing derivative on U. Using the identification of Exercise 4, show that $df_z\mathbf{v} = f'(z)\mathbf{v}$, the multiplication on the right being complex multiplication. Then show that

$$\langle df_z\mathbf{v}, df_z\mathbf{w}\rangle = \langle f'(z)\mathbf{v}, f'(z)\mathbf{w}\rangle = |f'(z)|^2\langle \mathbf{v}, \mathbf{w}\rangle,$$

again proving that f is a conformal map.

Chapter 5

Conformal Maps in Euclidean Space

5.1. Inversion in spheres

Having seen the abundance of conformal maps in the plane, we ask if there are as many conformal maps in $\mathbb{R}^n$ for $n \geq 3$. We shall see that dimension 2 is indeed exceptional in its richness of conformal mappings, and that the only conformal maps in Euclidean space are those that generalize the Möbius transformations—that is, mappings that are generated by similarities and inversions in spheres. This was first proved in dimension 3 by Liouville in 1850 [**22**]. The standard proof of this requires that the mapping be at least of class C^3 and requires some knowledge of differential geometry. In the next chapter we will briefly review some differential geometry and give the standard or classical proof. A more recent differential geometric proof was given by Huff [**20**]; it is also for C^3 mappings. The result is known for C^1 mappings [**19**], but this is difficult. In 1960 R. Nevanlinna [**26**] gave an elementary proof for C^4 mappings which we present here (see also [**9**], pp. 168-174, and [**10**], pp. 136-142).

We begin with a study of inversion in a sphere. Let $S_r(\mathbf{x}_0)$ be the sphere centered at $\mathbf{x}_0$ with radius r, i.e.

$$S_r(\mathbf{x}_0) = \{\mathbf{x} \in \mathbb{R}^n \big| |\mathbf{x} - \mathbf{x}_0| = r\}.$$

Inversion in $S_r(\mathbf{x}_0)$ is the mapping $g : \mathbb{R}^n - \{\mathbf{x}_0\} \longrightarrow \mathbb{R}^n$ defined by

$$g(\mathbf{x}) = \mathbf{x}_0 + r^2 \frac{\mathbf{x} - \mathbf{x}_0}{|\mathbf{x} - \mathbf{x}_0|^2}.$$

Note that $\mathbf{x}$ and $g(\mathbf{x})$ are on the same ray emmanating from $\mathbf{x}_0$, and that the product of the distances of $\mathbf{x}$ and $g(\mathbf{x})$ from $\mathbf{x}_0$ is equal to r^2.

Since g maps $\mathbb{R}^n - \{\mathbf{x}_0\}$ onto itself, $g \circ g$ is defined on $\mathbb{R}^n - \{\mathbf{x}_0\}$ and is the identity there. Thus by the chain rule the matrix of $dg_{g(\mathbf{x})} \circ dg_{\mathbf{x}}$ is the identity there, and, taking determinants, we see that $dg_{\mathbf{x}}$ is non-singular, i.e. the Jacobian of g at $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{x}_0\}$ is non-zero.

Without loss of generality take $\mathbf{x}_0$ to be the origin and $r = 1$; this may be thought of as a choice of coordinates or the result of composition of g with similarites. g is now given by $g(\mathbf{x}) = \mathbf{x}/|\mathbf{x}|^2$. The matrix of $dg_{\mathbf{x}}$ is then

$$\frac{1}{|\mathbf{x}|^4} \begin{pmatrix} |\mathbf{x}|^2 - 2(x^1)^2 & -2x^1x^2 & \cdots & \cdots & \cdots \\ -2x^2x^1 & |\mathbf{x}|^2 - 2(x^2)^2 & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & -2x^nx^{n-1} & |\mathbf{x}|^2 - 2(x^n)^2 \end{pmatrix}.$$

Applying this to a vector $\mathbf{v} \in \mathbb{R}^n$ as a column vector we obtain

$$\frac{1}{|\mathbf{x}|^4} \begin{pmatrix} v^1|\mathbf{x}|^2 - 2x^1\langle \mathbf{x}, \mathbf{v}\rangle \\ v^2|\mathbf{x}|^2 - 2x^2\langle \mathbf{x}, \mathbf{v}\rangle \\ \vdots \\ v^n|\mathbf{x}|^2 - 2x^n\langle \mathbf{x}, \mathbf{v}\rangle \end{pmatrix}.$$

Therefore

$$\langle dg_{\mathbf{x}}\mathbf{v}, dg_{\mathbf{x}}\mathbf{v}\rangle = \frac{1}{|\mathbf{x}|^8} \sum_{i=1}^{n} (v^i|\mathbf{x}|^2 - 2x^i\langle \mathbf{x}, \mathbf{v}\rangle)^2 = \frac{|\mathbf{v}|^2}{|\mathbf{x}|^4},$$

i.e., we have $\langle dg_{\mathbf{x}}\mathbf{v}, dg_{\mathbf{x}}\mathbf{v}\rangle = e^{2\sigma}\langle \mathbf{v}, \mathbf{v}\rangle$ for any $\mathbf{v} \in \mathbb{R}^n$ and any $\mathbf{x} \in \mathbb{R}^n - \{(0, \ldots, 0)\}$, where $e^{2\sigma} = 1/|\mathbf{x}|^4$. Replacing $\mathbf{v}$ by $\mathbf{v} + \mathbf{w}$, we see that

$$\langle dg_{\mathbf{x}}\mathbf{v}, dg_{\mathbf{x}}\mathbf{w}\rangle = e^{2\sigma}\langle \mathbf{v}, \mathbf{w}\rangle,$$

giving the following result.

Theorem 5.1. *Inversion in a sphere is a conformal map.*

In general the characteristic function e^σ of inversion in the sphere of radius r and centered at $\mathbf{x}_0$ is $r^2/|\mathbf{x}-\mathbf{x}_0|^2$.

We shall now show that inversion maps hyperplanes and spheres in $\mathbb{R}^n$ to hyperplanes and spheres. By a *hyperplane* we mean an $(n-1)$-dimensional linear subset of $\mathbb{R}^n$; it is given by an equation of the form $\sum_{i=1}^n A_i x^i = D$, $A_i, D \in \mathbb{R}$. The sphere $S_r(\mathbf{x}_0)$ of radius r about $\mathbf{x}_0$ is given by

$$\sum_{i=1}^n (x^i - x_0^i)^2 = r^2 \quad \text{or} \quad A\sum_{i=1}^n (x^i)^2 + \sum_{i=1}^n B_i x^i + C = 0, \ A \neq 0.$$

If $D = 0$, then $\sum_{i=1}^n A_i x^i = 0$ is a hyperplane through the origin, so, considering inversion in the unit sphere about the origin, every ray from the origin in the hyperplane remains in the hyperplane. If $\mathbf{y} = g(\mathbf{x}) = \frac{\mathbf{x}}{|\mathbf{x}|^2}$, then $|\mathbf{x}|^2 = \frac{1}{|\mathbf{y}|^2}$, so if $D \neq 0$, then g maps $\sum_{i=1}^n A_i x^i = D$ into

$$D|\mathbf{y}|^2 - \sum_{i=1}^n A_i y^i = 0,$$

which is a sphere passing through the origin. Similarly, a sphere passing through the origin ($C = 0$) is mapped to

$$\sum_{i=1}^n B_i y^i = -A,$$

which is a plane. Finally, the sphere $A\sum_{i=1}^n (x^i)^2 + \sum_{i=1}^n B_i x^i + C = 0$ is mapped to the sphere

$$C|\mathbf{y}|^2 + \sum_{i=1}^n B_i y^i + A = 0.$$

Thus, as in the case of plane inversion (Theorem 1.1), we have the following theorem.

Theorem 5.2. a) *The inverse of a hyperplane through the center of inversion is the hyperplane itself.*

b) *The inverse of a hyperplane not passing through the center of inversion is a sphere passing through the center of inversion.*

c) *The inverse of a sphere through the center of inversion is a hyperplane not passing through the center of inversion.*

d) *The inverse of a sphere not passing through the center of inversion is a sphere not passing through the center of inversion.*

Corollary 5.1. *Inversion in a sphere maps lines and circles to lines and circles.*

Proof. A line in $\mathbb{R}^n$ is the intersection of $n-1$ hyperplanes, and a circle is the intersection of $n-1$ hyperplanes and spheres with at least one sphere. In any case the inverse of these $n-1$ hyperplanes and spheres is a collection of $n-1$ intersecting hyperplanes and spheres. □

Theorem 5.3. *If two inversions g_1 and g_2 have the same center, their composition is a homothety.*

Proof. Suppose $g_1(\mathbf{x}) = r^2\frac{\mathbf{x}}{|\mathbf{x}|^2}$ and $g_2(\mathbf{x}) = R^2\frac{\mathbf{x}}{|\mathbf{x}|^2}$; then

$$(g_2 \circ g_1)(\mathbf{x}) = R^2\frac{\frac{r^2\mathbf{x}}{|\mathbf{x}|^2}}{\left|\frac{r^2\mathbf{x}}{|\mathbf{x}|^2}\right|^2} = \frac{R^2}{r^2}\mathbf{x}.$$

□

Theorem 5.4. *If two inversions g_1 and g_2 have centers $\mathbf{x}_0$ and $\mathbf{y}_0$ respectively and if $g_2 \circ g_1$ is a similarity, then $\mathbf{x}_0 = \mathbf{y}_0$.*

Proof. Suppose $\mathbf{x}_0 \neq \mathbf{y}_0$, and consider a hyperplane which does not pass through $\mathbf{x}_0$ and is such that its image under g_1 is a sphere passing through $\mathbf{x}_0$ and not through $\mathbf{y}_0$. Then the image of the hyperplane under $g_2 \circ g_1$ is a sphere, and hence, since similarities map hyperplanes to hyperplanes, $g_2 \circ g_1$ is not a similarity (Figure 5.1). □

Exercises

1. For dimension $n = 2$ show that the mapping

$$g(\mathbf{x}) = \mathbf{x}_0 + r^2(\mathbf{x} - \mathbf{x}_0)/|\mathbf{x} - \mathbf{x}_0|^2$$

is the same as inversion in a circle.

2. Find the image of the plane $x^3 = 0$ under inversion in the sphere $(x^1)^2 + (x^2)^2 + (x^3 - 1)^2 = 1$. Give another proof that stereographic projection is a conformal map.

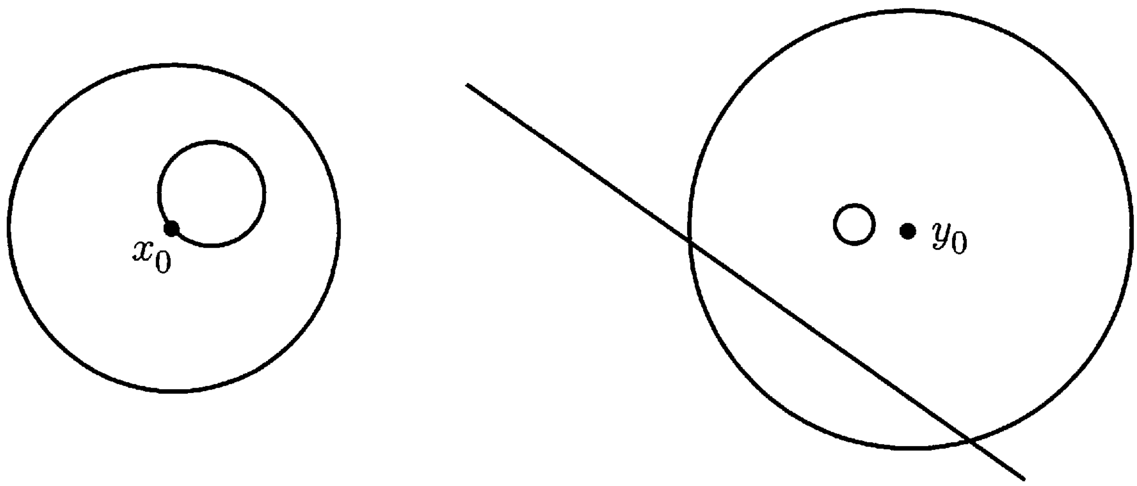

Figure 5.1

3. Define and give coordinate formulas for stereographic projection of $\mathbb{R}^n$ onto a sphere $S^n \subset \mathbb{R}^{n+1}$.

4. Consider a torus (anchor ring) in $\mathbb{R}^3$ and an interior point O of the solid torus. Invert the exterior of the torus in a sphere with center O. In Exercise 3 we saw that S^3 is the union of $\mathbb{R}^3$ with an ideal point ∞. Show that S^3 is the union of two solid tori with their boundaries identified.

5. Generalize Exercise 4 of Section 1.3.

6. We end with this exercise, which will be used in the next section. Let $|\mathbf{x} - \mathbf{x}_0| = r$, $r \in [r_1, r_2]$, be a family of concentric spheres and let $\mathbf{x}(t)$ be a C^1 curve intersecting the spheres such that $\mathbf{x}'(t)$ is orthogonal to the spheres. Show that $\mathbf{x}(t)$ is a segment of a ray from $\mathbf{x}_0$.

5.2. Conformal maps in Euclidean space

We now proceed to prove our main result that the only conformal maps in $\mathbb{R}^n$ are those generated by similarities and inversions. Let $f : U \longrightarrow \mathbb{R}^n$ be a conformal map defined on an open set $U \subset \mathbb{R}^n$. Then, as we have seen in Section 3.3,

$$\langle df_{\mathbf{x}}\mathbf{v}, df_{\mathbf{x}}\mathbf{w}\rangle = e^{2\sigma(\mathbf{x})}\langle \mathbf{v}, \mathbf{w}\rangle$$

for all vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and some real-valued function σ on U. In particular, the characteristic function e^σ is related to f by $e^{\sigma(\mathbf{x})} = |df_\mathbf{x}\mathbf{v}|/|\mathbf{v}|$ for all $\mathbf{v} \neq 0$. We begin with a difficult lemma giving a strong implication of the conformality of f on the function σ. Throughout the remainder of this chapter we will assume that f is of class C^4 and that the dimension $n \geq 3$.

Lemma 5.1. *Let f be a conformal map with characteristic function $e^{\sigma(\mathbf{x})}$. If σ is not a constant, then $d^2(e^{-\sigma})_\mathbf{x}(\mathbf{v}, \mathbf{w}) = \alpha\langle \mathbf{v}, \mathbf{w}\rangle$ for some constant α.*

Proof. Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be n mutually orthogonal unit vectors in $\mathbb{R}^n$. Then $\langle df_\mathbf{x}\mathbf{v}_i, df_\mathbf{x}\mathbf{v}_j\rangle = e^{2\sigma}\langle \mathbf{v}_i, \mathbf{v}_j\rangle$. Differentiating this, we have (see, e.g., Section 3.2)

$$\begin{aligned}\langle d^2 f_\mathbf{x}(\mathbf{v}_i, \mathbf{v}_k), df_\mathbf{x}(\mathbf{v}_j)\rangle &+ \langle df_\mathbf{x}(\mathbf{v}_i), d^2 f_\mathbf{x}(\mathbf{v}_j, \mathbf{v}_k)\rangle \\ &= 2e^{2\sigma}\langle \mathbf{v}_i, \mathbf{v}_j\rangle d\sigma_\mathbf{x}(\mathbf{v}_k).\end{aligned}$$

Cyclicly permuting the indices i, j, k, we also have

$$\begin{aligned}\langle d^2 f_\mathbf{x}(\mathbf{v}_k, \mathbf{v}_j), df_\mathbf{x}(\mathbf{v}_i)\rangle &+ \langle df_\mathbf{x}(\mathbf{v}_k), d^2 f_\mathbf{x}(\mathbf{v}_i, \mathbf{v}_j)\rangle \\ &= 2e^{2\sigma}\langle \mathbf{v}_k, \mathbf{v}_i\rangle d\sigma_\mathbf{x}(\mathbf{v}_j), \\ \langle d^2 f_\mathbf{x}(\mathbf{v}_j, \mathbf{v}_i), df_\mathbf{x}(\mathbf{v}_k)\rangle &+ \langle df_\mathbf{x}(\mathbf{v}_j), d^2 f_\mathbf{x}(\mathbf{v}_k, \mathbf{v}_i)\rangle \\ &= 2e^{2\sigma}\langle \mathbf{v}_j, \mathbf{v}_k\rangle d\sigma_\mathbf{x}(\mathbf{v}_i).\end{aligned}$$

Adding the last two equations and subtracting the first gives

$$\begin{aligned}&\langle d^2 f_\mathbf{x}(\mathbf{v}_j, \mathbf{v}_i), df_\mathbf{x}(\mathbf{v}_k)\rangle \\ &\quad = e^{2\sigma}(\langle \mathbf{v}_k, \mathbf{v}_i\rangle d\sigma_\mathbf{x}(\mathbf{v}_j) + \langle \mathbf{v}_j, \mathbf{v}_k\rangle d\sigma_\mathbf{x}(\mathbf{v}_i) - \langle \mathbf{v}_i, \mathbf{v}_j\rangle d\sigma_\mathbf{x}(\mathbf{v}_k)).\end{aligned}$$

Therefore, since $\{df_\mathbf{x}(\mathbf{v}_k)\}$ is an orthogonal basis of $\mathbb{R}^n$, it follows that $\{e^{-\sigma} df_\mathbf{x}(\mathbf{v}_k)\}$ is an orthonormal basis, and for $i \neq j$ we have

$$d^2 f_\mathbf{x}(\mathbf{v}_j, \mathbf{v}_i) = d\sigma_\mathbf{x}(\mathbf{v}_j) df_\mathbf{x}(\mathbf{v}_i) + d\sigma_\mathbf{x}(\mathbf{v}_i) df_\mathbf{x}(\mathbf{v}_j), \tag{5.1}$$

and for $i = j$

$$d^2 f_\mathbf{x}(\mathbf{v}_i, \mathbf{v}_i) = d\sigma_\mathbf{x}(\mathbf{v}_i) df_\mathbf{x}(\mathbf{v}_i) - \sum_{k \neq i} d\sigma_\mathbf{x}(\mathbf{v}_k) df_\mathbf{x}(\mathbf{v}_k). \tag{5.2}$$

Note that so far we have not required that σ be non-constant.

Suppose now that $i \neq j$. Then, multiplying equation (5.1) by $e^{-\sigma}$, we have

$$e^{-\sigma}d^2f_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_i)+d(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_j)df_{\mathbf{x}}(\mathbf{v}_i)+d(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_i)df_{\mathbf{x}}(\mathbf{v}_j)=0.$$

Differentiating again gives

$$\begin{gathered}d(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_k)d^2f_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_i)+e^{-\sigma}d^3f_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_i,\mathbf{v}_k)\\ +d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_k)df_{\mathbf{x}}(\mathbf{v}_i)+d(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_j)d^2f_{\mathbf{x}}(\mathbf{v}_i,\mathbf{v}_k)\\ +d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_i,\mathbf{v}_k)df_{\mathbf{x}}(\mathbf{v}_j)+d(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_i)d^2f_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_k)=0.\end{gathered}$$

Now the second, fourth, fifth and the sum of the first and sixth terms are symmetric in i and k; therefore, since the right hand side (trivially) is symmetric in i and k, the third term must also be symmetric in i and k. Thus, for distinct i,j,k,

$$d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_k)df_{\mathbf{x}}(\mathbf{v}_i)=d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{v}_j,\mathbf{v}_i)df_{\mathbf{x}}(\mathbf{v}_k).$$

Since $k \neq i$, it follows that $df_{\mathbf{x}}(\mathbf{v}_i)$ and $df_{\mathbf{x}}(\mathbf{v}_k)$ are linearly independent, and hence

$$d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{u},\mathbf{v})=0$$

for all orthogonal vectors $\mathbf{u}$ and $\mathbf{v}$. Note that we have used $n \geq 3$ in our argument. Now, by Theorem 3.7,

$$d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{u},\mathbf{v})=\alpha\langle\mathbf{u},\mathbf{v}\rangle.$$

Thus it remains only to show that α is a constant; but this is relatively easy. First, by differentiation,

$$d^3(e^{-\sigma})_{\mathbf{x}}(\mathbf{u},\mathbf{v},\mathbf{w})=d\alpha_{\mathbf{x}}(\mathbf{w})\langle\mathbf{u},\mathbf{v}\rangle;$$

notice that f must have derivatives of order 4. Interchanging $\mathbf{u}$ and $\mathbf{w}$, we have

$$\langle d\alpha_{\mathbf{x}}(\mathbf{w})\mathbf{u}-d\alpha_{\mathbf{x}}(\mathbf{u})\mathbf{w},\mathbf{v}\rangle=0$$

for any $\mathbf{v}$. Therefore $d\alpha_{\mathbf{x}}(\mathbf{w})\mathbf{u}-d\alpha_{\mathbf{x}}(\mathbf{u})\mathbf{w}=0$, but choosing $\mathbf{u}$ and $\mathbf{w}$ independent we see that $d\alpha_{\mathbf{x}}(\mathbf{w})=0$ for any $\mathbf{w}$, and hence that α is a constant. □

We now state and prove our main theorem.

Theorem 5.5. *Let f be a one-to-one C^4 conformal map of an open set $U \subset \mathbb{R}^n$ onto $f(U)$, and suppose that $n \geq 3$. Then f is a composition of similarities and inversions.*

Proof. By the preceeding lemma $d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{v},\mathbf{w}) = \alpha\langle\mathbf{v},\mathbf{w}\rangle$, where α is a constant. Thinking of $d^2(e^{-\sigma})_{\mathbf{x}}(\mathbf{v},\mathbf{w})$ as $d(d(e^{-\sigma})(\mathbf{v}))_{\mathbf{x}}(\mathbf{w})$, we see that

$$\sum_{j=1}^{n} \frac{\partial}{\partial x^j}\Big(\sum_{i=1}^{n} \frac{\partial e^{-\sigma}}{\partial x^i} v^i\Big) w^j = \alpha \sum_{j=1}^{n} v^j w^j$$

for all $\mathbf{v}$ and $\mathbf{w}$. Therefore

$$\frac{\partial}{\partial x^j}\Big(\sum_{i=1}^{n} \frac{\partial e^{-\sigma}}{\partial x^i} v^i\Big) = \alpha v^j$$

for every j, and hence

$$\sum_{i=1}^{n} \frac{\partial e^{-\sigma}}{\partial x^i} v^i = \sum_{j=1}^{n} \alpha v^j (x^j - x_0^j)$$

for some constant vector $\mathbf{x}_0$ and all $\mathbf{v}$; that is,

$$\frac{\partial e^{-\sigma}}{\partial x^i} = \alpha(x^i - x_0^i).$$

Integrating a second time, we see that

$$e^{-\sigma(\mathbf{x})} = A|\mathbf{x} - \mathbf{x}_0|^2 + B, \tag{5.3}$$

where $A = \frac{\alpha}{2}$ and B are constants.

Now, since f is a one-to-one mapping of U onto $f(U)$ with $df_{\mathbf{x}}$ non-singular for every $\mathbf{x} \in U$, it follows that $f^{-1} : f(U) \longrightarrow U$ also satisfies $df_{\mathbf{y}}^{-1}$ non-singular, $\mathbf{y} = f(\mathbf{x})$, and

$$\langle df_{\mathbf{y}}^{-1}\mathbf{v}, df_{\mathbf{y}}^{-1}\mathbf{w}\rangle = e^{-2\sigma}\langle\mathbf{v},\mathbf{w}\rangle$$

(see Exercise 2 in Section 3.3). Thus f^{-1} is a conformal map with characteristic function $e^{-\sigma\circ f^{-1}}$. Applying the above argument to f^{-1}, we have $e^{\sigma(\mathbf{x})} = e^{-(-\sigma\circ f^{-1})(\mathbf{y})} = C|\mathbf{y} - \mathbf{y}_0|^2 + D$, and hence

$$(A|\mathbf{x} - \mathbf{x}_0|^2 + B)(C|\mathbf{y} - \mathbf{y}_0|^2 + D) = 1.$$

Thus if $A \neq 0$, f maps the sphere $|\mathbf{x} - \mathbf{x}_0| = r$ to the sphere $|\mathbf{y} - \mathbf{y}_0| = R$, where $(Ar^2 + B)(CR^2 + D) = 1$.

Consider a segment of some ray emanating from $\mathbf{x}_0$ and lying in U, say $\mathbf{x}(t) = t\mathbf{a} + \mathbf{x}_0$, $t \in [t_1, t_2]$, where $\mathbf{a}$ is the unit vector in the direction of the ray. Since $\mathbf{x}(t)$ is orthogonal to the spheres $|\mathbf{x} - \mathbf{x}_0| = r$ and f is conformal, the image curve $\mathbf{y}(t)$ is orthogonal to the concentric spheres $|\mathbf{y} - \mathbf{y}_0| = R$ and hence $\mathbf{y}(t)$ is a segment of a ray from $\mathbf{y}_0$ (see Exercise 6, Section 5.1). From (5.3) the length of the image segment from t_1 to any $\tau \in [t_1, t_2]$ is

$$\int_{t_1}^{\tau} |\mathbf{y}'(t)|dt = \int_{t_1}^{\tau} e^{\sigma}|\mathbf{x}'(t)|dt = \int_{t_1}^{\tau} \frac{dt}{At^2+B}$$

which for $A \neq 0$ and $B \neq 0$ is a transcendental function of τ. On the other hand,

$$(At_1^2 + B)(C|\mathbf{y}(t_1) - \mathbf{y}_0|^2 + D) = 1$$

and

$$(A\tau^2 + B)(C|\mathbf{y}(\tau) - \mathbf{y}_0|^2 + D) = 1;$$

thus we see that the length of the image segment, $|\mathbf{y}(\tau) - \mathbf{y}(t_1)|$, is an algebraic function of τ. Consequently either $A = 0$ or $B = 0$.

Suppose $A = 0$; then $e^{\sigma} = 1/B$, which is a constant. Note from the proof of Lemma 5.1 that we still have equations (5.1) and (5.2), and hence that $d^2 f_{\mathbf{x}}(\mathbf{v}, \mathbf{w}) = 0$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Therefore

$$\frac{\partial}{\partial x^j}\Big(\sum_{i=1}^{n} \frac{\partial f^k}{\partial x^i} v^i\Big) = 0$$

for all $\mathbf{v}$, and hence each $\partial f^k/\partial x^i$ is a constant, say S_i^k. Integrating again gives $f^k(\mathbf{x}) = \sum_{i=1}^n S_i^k x^i + T^k$, or, letting S be the matrix whose components are the S_i^k's (k, the row index),

$$f(\mathbf{x}) = S\mathbf{x} + T,$$

where T is the column vector of T^k's. Also S is the matrix of $df_{\mathbf{x}}$, and hence $\langle S\mathbf{v}, S\mathbf{w}\rangle = \frac{1}{B^2}\langle \mathbf{v}, \mathbf{w}\rangle$ for all $\mathbf{v}, \mathbf{w}$. Now, for any two points $\mathbf{x}_1, \mathbf{x}_2$,

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| = |S(\mathbf{x}_1 - \mathbf{x}_2)| = \frac{1}{B}|\mathbf{x}_1 - \mathbf{x}_2|,$$

and f is a similarity.

If $B = 0$, let $\mathbf{y} = g(\mathbf{x}) = \mathbf{x}_0 + (\mathbf{x} - \mathbf{x}_0)/|\mathbf{x} - \mathbf{x}_0|^2$ be inversion with respect to the unit sphere about $\mathbf{x}_0$ and consider the conformal map $f^*(\mathbf{y}) = (f \circ g^{-1})(\mathbf{y})$. Since the characteristic function of g is $1/|\mathbf{x} - \mathbf{x}_0|^2$, that of g^{-1} is $1/|\mathbf{y} - \mathbf{y}_0|^2 = |\mathbf{x} - \mathbf{x}_0|^2$. Therefore the characteristic function of f^* is $e^{\sigma(\mathbf{x})}|\mathbf{x} - \mathbf{x}_0|^2 = 1/A$, and hence f^* is a similarity. Finally, $f = f^* \circ g$, so that f is a composition of an inversion and a similarity. □

Note that we have also shown that any composition of similarities and inversions is either a similarity f, an inversion g, or a composition $f \circ g$ of one of each, where g is an inversion in a sphere of radius 1. Moreover, the form $f \circ g$ is unique. For if $f \circ g = f_1 \circ g_1$, then $g_1 \circ g^{-1} = f_1^{-1} \circ f$ is a similarity, and hence by Theorem 5.4 g and g_1 have the same center. Since both g and g_1 are inversions in unit spheres, $g = g_1$. Therefore $f_1^{-1} \circ f$ is the identity, giving $f = f_1$.

5.3. Sphere preserving transformations

In Chapter 2 we read a paper of Carathéodory proving that transformations of the plane that map circles to circles are extended Möbius transformations, i.e. linear fractional transformations in z or $\bar{z}$, and hence are compositions of inversions and similarities. In this spirit we prove the following theorem of Möbius, which assumes the continuity of the mapping.

Theorem 5.6. *Let $f : U \longrightarrow f(U)$ be a continuous 1-1 mapping defined on an open set in $\mathbb{R}^n$, and suppose that f maps (pieces of) planes and spheres in U to (pieces of) planes and spheres in $f(U)$ (not necessarily respectively). Then f is a composition of similarites and inversions.*

Proof. For any $\mathbf{x} \in U$ choose $\mathbf{x}_0 \neq \mathbf{x}$ in U and a ball $B_r(\mathbf{x}_0)$ such that the closed ball $\bar{B}_r(\mathbf{x}_0) = B_r(\mathbf{x}_0) \cup S_r(\mathbf{x}_0) \subset U$ but $\mathbf{x} \notin \bar{B}_r(\mathbf{x}_0)$. Let $\mathbf{y}_0 = f(\mathbf{x}_0)$, and let $S_{r'}(\mathbf{y}_0)$ be any sphere about $\mathbf{y}_0$. Let g and h be inversions in $S_r(\mathbf{x}_0)$ and $S_{r'}(\mathbf{y}_0)$ respectively. Then $h \circ f \circ g$ is defined on the exterior of $\bar{B}_r(\mathbf{x}_0)$, and the image of a hyperplane lying in this exterior is a hyperplane. Thus, considering intersections, $h \circ f \circ g$ maps lines to lines. Also parallel lines l_1 and l_2 are mapped to parallel lines, even if the plane of the lines meets $\bar{B}_r(\mathbf{x}_0)$, for there exist non-parallel planes π_1 and π_2 containing l_1 and l_2 respectively but not meeting $\bar{B}_r(\mathbf{x}_0)$, and $l = \pi_1 \cap \pi_2$ is parallel to both l_1 and l_2.

Now $\mathbf{x} \notin \bar{B}_r(\mathbf{x}_0)$; therefore $\mathbf{x}$ is in the domain of $h \circ f \circ g$. Let $T_{\mathbf{x}} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be translation by $\mathbf{x}$, i.e. the vector $\mathbf{v} \in \mathbb{R}^n$ is mapped to the vector $\mathbf{v} + \mathbf{x}$. Then, setting $\mathbf{y} = (h \circ f \circ g)(\mathbf{x})$, we see that

$$\varphi = T_{-\mathbf{y}} \circ h \circ f \circ g \circ T_{\mathbf{x}}$$

maps the origin to the origin, lines to lines, and preserves parallelism. Therefore φ maps parallelograms to parallelograms, and so, as a vector space transformation, $\varphi(\mathbf{v}+\mathbf{w}) = \varphi(\mathbf{v}) + \varphi(\mathbf{w})$. Consequently for a positive integer m we have $\varphi(m\mathbf{v}) = m\varphi(\mathbf{v})$, and in turn, setting $\mathbf{u} = m\mathbf{v}$, $\varphi(\frac{1}{m}\mathbf{u}) = \frac{1}{m}\varphi(\mathbf{u})$. Therefore $\varphi(q\mathbf{v}) = q\varphi(\mathbf{v})$ for q a positive rational. Now $\varphi(\mathbf{0}) = \mathbf{0}$ implies $\varphi(-\mathbf{v}) = -\varphi(\mathbf{v})$, so $\varphi(q\mathbf{v}) = q\varphi(\mathbf{v})$ for q any rational; but φ is continuous and hence $\varphi(c\mathbf{v}) = c\varphi(\mathbf{v})$ for any $c \in \mathbb{R}$. Thus φ is linear, and since it is one-to-one on a neighborhood of the origin, it is non-singular. Therefore φ may be given with respect to some basis by a non-singular matrix A. By polarization we can write A uniquely as FG, where F is an orthogonal matrix and G is a positive definite symmetric matrix. Viewing F and G as linear transformations, F leaves spheres about the origin invariant. Now A maps spheres to spheres; so, diagonalizing G, we see that the eigenvalues of G are all equal and hence that φ is a similarity, namely the isometry given by F composed with the homothety given by G. Thus

$$f = h \circ T_{\mathbf{y}} \circ \varphi \circ T_{-\mathbf{x}} \circ g$$

is a composition of similarities and inversions. □

Chapter 6

The Classical Proof of Liouville's Theorem

6.1. Surface theory

In this chapter we give the standard or "classical" proof of Liouville's theorem that in dimension 3, the only conformal maps are those generated by similarities and inversions. This is not Liouville's proof [**22**] and requires some knowledge of the differential geometry of surfaces in Euclidean space. We develop it very briefly in this section; for more detail see any of [**9**], [**23**], [**27**]. Chapter 7 also requires this knowledge.

We will begin, however, with a review of curve theory in the plane. Curve theory in the plane will be developed further in Section 7.1. In classical notation, let $\mathbf{x}(t) = x^1(t)\mathbf{i} + x^2(t)\mathbf{j}$ be a curve in the plane given parametrically in terms of the position vector of each point, and suppose that $x^1(t)$ and $x^2(t)$ are differentiable functions of t. In particular, we will assume that $x^1(t)$ and $x^2(t)$ have continuous derivatives of sufficiently high order for our arguments; class C^3 will be sufficient. We will often simply refer to the curve as being smooth.

Recall that the derivative of a vector valued function is defined by the following limit, if it exists:

$$\mathbf{x}'(t) = \lim_{h\to 0} \frac{\mathbf{x}(t+h) - \mathbf{x}(t)}{h}.$$

A curve $\mathbf{x}(t)$ is said to be *regular* if $\mathbf{x}'(t) \neq 0$ for all t in the domain of $\mathbf{x}$. For a regular curve, $\mathbf{x}'(t)$ is said to be *tangent* to the curve, and, in general, $\mathbf{x}'(t)$ is called the *velocity vector* of $\mathbf{x}(t)$, regarded, from the dynamical point of view, as the trajectory of a moving particle.

Letting s denote arc length along the curve from some point, recall that

$$\frac{ds}{dt} = |\mathbf{x}'(t)| = \sqrt{\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2}.$$

Thus we may define the *unit tangent field* to a regular curve by $\mathbf{T} = \frac{\mathbf{x}'(t)}{|\mathbf{x}'(t)|}$. By the chain rule, $\mathbf{T} = \frac{d\mathbf{x}}{ds}$. Thus if arc length is chosen as the parameter, the derivative gives the unit tangent field along the curve.

Since $\langle \mathbf{T}, \mathbf{T}\rangle = 1$, differentiation yields $\langle \mathbf{T}, \frac{d\mathbf{T}}{ds}\rangle = 0$. Thus $\frac{d\mathbf{T}}{ds}$ is perpendicular to the curve. We define the *principal normal* $\mathbf{N}$ to a regular plane curve to be the unit vector at each point that advances the unit tangent by $\frac{\pi}{2}$ (see Figure 6.1). Intuitively, curvature should mean the rate of change of direction as we move along the curve; thus we define the *curvature* $\kappa(s)$ of $\mathbf{x}(t)$ by

$$\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N};$$

this equation is known as the *Frenet equation* of the plane curve. Many differential geometers define $\mathbf{N}$ to be the unit vector field in the direction of $\frac{d\mathbf{T}}{ds}$ and the curvature by $\kappa = \left|\frac{d\mathbf{T}}{ds}\right|$ for plane curves, and almost universally the principal normal and curvature of space curves are defined this way. For our purposes, especially in Chapter 7, it will be essential to have the notion of a signed curvature (Figure 6.1).

Also from $\langle \mathbf{N}, \mathbf{N}\rangle = 1$, we have that $\frac{d\mathbf{N}}{ds}$ is perpendicular to $\mathbf{N}$ and hence collinear with $\mathbf{T}$. Differentiating $\langle \mathbf{N}, \mathbf{T}\rangle = 0$ then gives the second Frenet equation, $\frac{d\mathbf{N}}{ds} = -\kappa\mathbf{T}$.

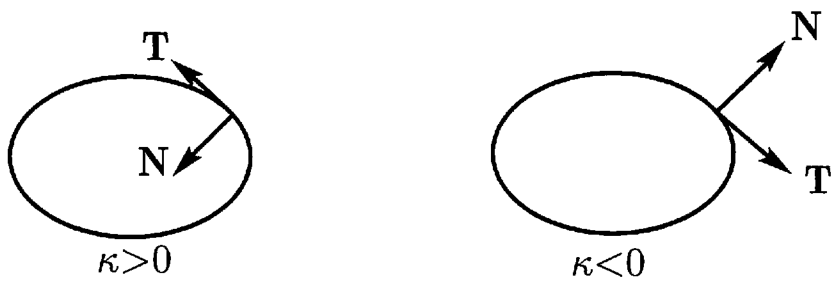

Figure 6.1

Since $\mathbf{T}$ is a unit vector, we can write $\mathbf{T} = \cos\phi\mathbf{i} + \sin\phi\mathbf{j}$. Then $\mathbf{N} = -\sin\phi\mathbf{i} + \cos\phi\mathbf{j}$ and

$$\frac{d\mathbf{T}}{\mathbf{ds}} = -(\sin\phi)\frac{d\phi}{ds}\mathbf{i} + (\cos\phi)\frac{d\phi}{ds}\mathbf{j}.$$

Thus from the Frenet equation we see that

$$\kappa(s) = \frac{d\phi}{ds}.$$

For example, for the circle $\mathbf{x} = a\cos\theta\mathbf{i} + a\sin\theta\mathbf{j}$ of radius a, arc length is given by $s = a\theta$ and $\phi = \theta + \frac{\pi}{2}$. Thus $\kappa(s) = \frac{d\phi}{ds} = \frac{1}{a}$.

Turning now to surface theory, a *piece of a regular surface* or a *coordinate patch* is a smooth one-to-one map $\mathbf{x}$ of a neighborhood $\mathcal{U} \subset \mathbb{R}^2$ into $\mathbb{R}^3$ given explicitly as a vector-valued function of two parameters: $\mathbf{x}(u^1, u^2) = (x^1(u^1, u^2), x^2(u^1, u^2), x^3(u^1, u^2))$ such that $\mathbf{x}$ is C^2 (C^3, C^4, etc.) and

$$\mathbf{x}_1 \times \mathbf{x}_2 \neq 0 \quad \text{(the regularity)},$$

where $\mathbf{x}_1 = \frac{\partial \mathbf{x}}{\partial u^1}$, $\mathbf{x}_2 = \frac{\partial \mathbf{x}}{\partial u^2}$.

A *surface* in $\mathbb{R}^3$ is a subset $M \subset \mathbb{R}^3$ such that for every point $p \in M$ there exists a coordinate patch $\mathbf{x} : \mathcal{U} \longrightarrow \mathbb{R}^3$ with $\mathbf{x}(\mathcal{U}) \subset M$, $p \in \mathbf{x}(\mathcal{U})$, and if $\mathbf{x} : \mathcal{U} \longrightarrow M$ and $\mathbf{y} : \mathcal{V} \longrightarrow M$ are coordinate patches with images $\mathcal{U}'$ and $\mathcal{V}'$ with $\mathcal{U}' \cap \mathcal{V}' \neq \emptyset$, then the map

$$\mathbf{y}^{-1} \circ \mathbf{x} : \mathbf{x}^{-1}(\mathcal{U}' \cap \mathcal{V}') \longrightarrow \mathbf{y}^{-1}(\mathcal{U}' \cap \mathcal{V}')$$

has non-vanishing Jacobian (see Figure 6.2).

The regularity condition $\mathbf{x}_1 \times \mathbf{x}_2 \neq 0$ enables us to define a local *unit normal field* $\mathbf{n}$ by

$$\mathbf{n} = \frac{\mathbf{x}_1 \times \mathbf{x}_2}{|\mathbf{x}_1 \times \mathbf{x}_2|}.$$

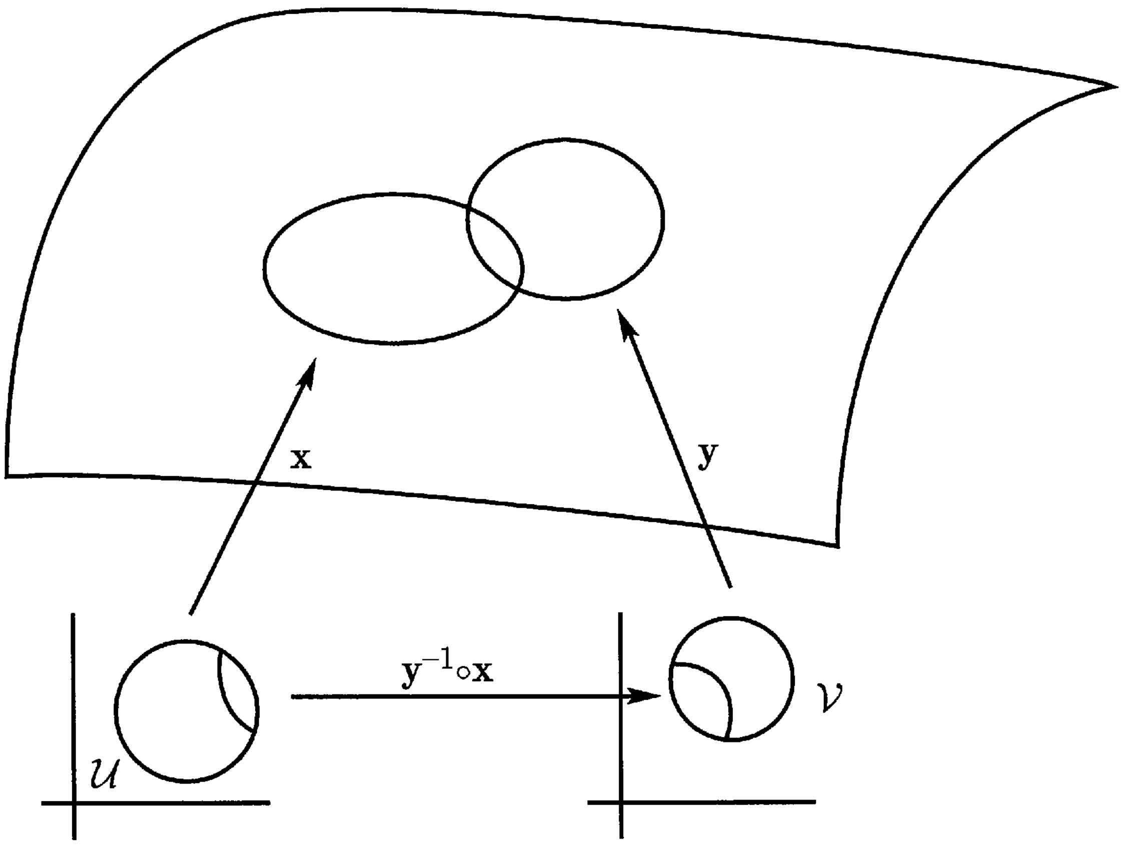

Figure 6.2

If a surface M is orientable, $\mathbf{n}$ may be taken globally (see e.g. [**9**]).

Consider a curve $\mathbf{x}(u^1(t), u^2(t))$ on the surface, where $u^1(t), u^2(t)$ define a regular curve in the parameter domain $\mathcal{U} \subset \mathbb{R}^2$. Then

$$\frac{d\mathbf{x}}{dt} = \frac{du^1}{dt}\mathbf{x}_1 + \frac{du^2}{dt}\mathbf{x}_2 \neq 0.$$

Now $\mathbf{x}_1$ and $\mathbf{x}_2$ are tangents to the parameter curves, and hence $\frac{d\mathbf{x}}{dt}$ lies in the plane determined by $\mathbf{x}_1$ and $\mathbf{x}_2$. In particular, the tangents to all regular curves through a point of a surface M and lying in the surface are coplanar, and we call this plane the *tangent plane* to the surface at the point. We denote the tangent plane to M at $p \in M$ by T_pM. Note that at each point of the surface the unit normal $\mathbf{n}$ is perpendicular to the tangent plane at the point.

The length of a regular curve $\mathbf{x}(u^1(t), u^2(t))$ on the surface between the points given by $t = t_0$ and $t = t_1$ is given by $\int_{t_0}^{t_1} |\mathbf{x}'|dt$.

Thus, if we denote arc length by s, then

$$\begin{aligned}\left(\frac{ds}{dt}\right)^2 &= \langle \mathbf{x}', \mathbf{x}' \rangle \\ &= \langle \mathbf{x}_1, \mathbf{x}_1 \rangle \left(\frac{du^1}{dt}\right)^2 + 2\langle \mathbf{x}_1, \mathbf{x}_2 \rangle \frac{du^1}{dt}\frac{du^2}{dt} + \langle \mathbf{x}_2, \mathbf{x}_2 \rangle \left(\frac{du^2}{dt}\right)^2\end{aligned}$$

or $ds^2 = E(du^1)^2 + 2F du^1 du^2 + G(du^2)^2$, where $E = \langle \mathbf{x}_1, \mathbf{x}_1 \rangle$, $F = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle$ and $G = \langle \mathbf{x}_2, \mathbf{x}_2 \rangle$. Now by the chain rule one shows that ds^2 is invariant under a change of coordinates, so ds^2 may be regarded as globally defined on the surface. The expression $ds^2 = E(du^1)^2 + 2F du^1 du^2 + G(du^2)^2$ is called the *first fundamental form* of the surface. Viewing ds^2 as being given by the matrix $\left(\begin{smallmatrix} E & F \\ F & G \end{smallmatrix}\right)$, it follows that ds^2 is a positive definite symmetric bilinear form.

Note also that $EG - F^2 = \langle \mathbf{x}_1, \mathbf{x}_1 \rangle \langle \mathbf{x}_2, \mathbf{x}_2 \rangle - \langle \mathbf{x}_1, \mathbf{x}_2 \rangle^2 = |\mathbf{x}_1 \times \mathbf{x}_2|^2$. We define the area of a piece of a surface by $\iint_{\mathcal{U}} \sqrt{EG - F^2}\, dA$, where $\mathcal{U}$ is the coordinate domain. Under a change of coordinates, $\sqrt{EG - F^2}$ is multiplied by the absolute value of the Jacobian of the coordinate transformation, in agreement with the usual change of variables in a multiple integral, and so we may define the *area* of a surface by this integral.

We now want to differentiate the unit normal field as we move around on the surface. For a regular curve $\mathbf{x}(u^1(t), u^2(t))$ on the surface, restrict $\mathbf{n}$ to the curve. Then, since $\mathbf{n}$ is of unit length, $\mathbf{n}'$ is orthogonal to $\mathbf{n}$ and hence tangent to the surface. Now define a linear map $A : T_pM \longrightarrow T_pM$, called the *Weingarten map*, as follows. For a unit tangent vector $\mathbf{v} \in T_pM$

$$A\mathbf{v} = -\frac{d\mathbf{n}}{ds},$$

where the derivative is taken along a curve on M through p and tangent to $\mathbf{v}$; one then shows that this derivative at p is independent of the choice of the curve tangent to $\mathbf{v}$. In particular, if $\mathbf{v} = \mathbf{x}_j / |\mathbf{x}_j|$, then

$$A\mathbf{v} = -\frac{d\mathbf{n}}{ds} = -\frac{1}{|\mathbf{x}_j|}\frac{\partial \mathbf{n}}{\partial u^j},$$

and hence $A\mathbf{x}_j = -\partial \mathbf{n}/\partial u^j$. Since $\langle \mathbf{n}, \mathbf{x}_i \rangle = 0$, we have

$$0 = \frac{\partial}{\partial u^j}\langle \mathbf{n}, \mathbf{x}_i \rangle = \langle -A\mathbf{x}_j, \mathbf{x}_i \rangle + \left\langle \mathbf{n}, \frac{\partial^2 \mathbf{x}}{\partial u^j \partial u^i} \right\rangle,$$

and hence, by commutativity of the second partials, we see that A is a symmetric operator, i.e. $\langle A\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, A\mathbf{w}\rangle$.

Since A is a symmetric operator, its eigenvalues κ_1 and κ_2 are real and are called the *principal curvatures* of the surface at the point p. The corresponding unit eigenvectors are called *principal directions.* We now give a geometric interpretation of these curvatures.

Consider a plane in $\mathbb{R}^3$ containing the unit normal to M at p; this plane intersects the surface in a plane curve, called a *normal section* of the surface (Figure 6.3). Consider the signed curvature of this plane curve.

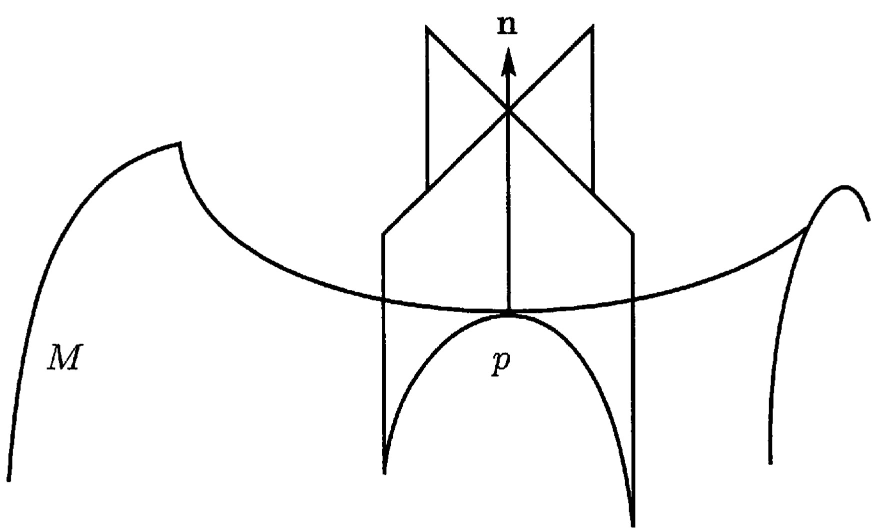

Figure 6.3

For each unit tangent vector $\mathbf{w}$ we have such a normal section. Now at the point p consider the curvatures of the normal sections for all directions $\mathbf{w}$ at the point. The maximum and minimum of these curvatures are the principal curvatures at the point. To see this, recall that the curvature κ of a plane curve is given by the Frenet equation $\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}$, where $\mathbf{T}$ is the unit tangent and $\mathbf{N}$ the principal normal, which is defined to be the unit vector at each point that advances the unit tangent by $\frac{\pi}{2}$. Thus for normal sections at p we may identify the principal unit normal $\mathbf{N}$ and the surface normal $\mathbf{n}$, the unit tangent at p being $\mathbf{w}$. Then

$$\kappa = \left\langle \frac{d\mathbf{T}}{ds}, \mathbf{n} \right\rangle = -\left\langle \mathbf{w}, \frac{d\mathbf{n}}{ds} \right\rangle = \langle A\mathbf{w}, \mathbf{w}\rangle.$$

If $\kappa_1 \neq \kappa_2$, let $\mathbf{w}_1$, $\mathbf{w}_2$ be unit eigenvectors corresponding to the eigenvalues κ_1, κ_2 respectively. Writing a unit tangent vector $\mathbf{w}$ as $\mathbf{w} = \cos\theta \mathbf{w}_1 + \sin\theta \mathbf{w}_2$, we have

$$\kappa = \langle A\mathbf{w}, \mathbf{w}\rangle = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta.$$

Differentiating gives $\kappa' = 2(\kappa_2 - \kappa_1)\sin\theta\cos\theta$. Thus the extrema of κ occur for the directions of the eigenvectors of A. Moreover, if $\kappa_2 = \kappa_1$, then all normal sections through p have the same curvature at p, and p is called an *umbilical point* of the surface.

Theorem 6.1. *If every point of a (piece of a) C^3 surface is an umbilical point, then the surface is a (piece of a) sphere or a plane.*

Proof. We first show that if every point of a surface is an umbilical point, then $\kappa = \kappa_1 = \kappa_2$ is constant on the surface, i.e. is independent of the point. Since κ is the same for all directions at a point, $\frac{\partial \mathbf{n}}{\partial u^i} = -\kappa \mathbf{x}_i$. Differentiating this, we have

$$\frac{\partial^2 \mathbf{n}}{\partial u^j \partial u^i} = -\frac{\partial \kappa}{\partial u^j}\mathbf{x}_i - \kappa \frac{\partial^2 \mathbf{x}}{\partial u^j \partial u^i}.$$

Interchanging the order of differentiation yields

$$\frac{\partial \kappa}{\partial u^1}\mathbf{x}_2 = \frac{\partial \kappa}{\partial u^2}\mathbf{x}_1,$$

but $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent, so that $\frac{\partial \kappa}{\partial u^i} = 0$ and hence κ is a constant.

If $\kappa = 0$, then $\frac{\partial \mathbf{n}}{\partial u^i} = 0$, giving that the field of unit normals is constant on the surface. Thus the surface is a plane perpendicular to $\mathbf{n}$.

If $\kappa \neq 0$, consider $\mathbf{x} + \frac{1}{\kappa}\mathbf{n}$. Differentiating gives

$$\frac{\partial}{\partial u^i}\left(\mathbf{x} + \frac{1}{\kappa}\mathbf{n}\right) = \mathbf{x}_i - \frac{1}{\kappa}\kappa \mathbf{x}_i = 0,$$

and hence that $\mathbf{x} + \frac{1}{\kappa}\mathbf{n}$ is a constant vector, say $\mathbf{c}$. Therefore

$$\langle \mathbf{x} - \mathbf{c}, \mathbf{x} - \mathbf{c}\rangle = \frac{1}{\kappa^2},$$

the equation of the sphere of center $\mathbf{c}$ and radius $\frac{1}{|\kappa|}$. □

A curve on a surface M is called a *line of curvature* if its unit tangent at each point is a principal direction. In particular, if $\mathbf{x}(s)$ is a line of curvature on the surface parametrized by arc length and $\mathbf{T}$ is its unit tangent, then $A\mathbf{T} = \kappa(s)\mathbf{T}$.

Now for a curve $\mathbf{x}(s)$ on the surface, let $\mathbf{v} = \mathbf{n} \times \mathbf{T}$. We define the *geodesic torsion* τ_g of the curve by

$$\tau_g = \left\langle \frac{d\mathbf{n}}{ds}, \mathbf{v} \right\rangle = -\langle A\mathbf{T}, \mathbf{v} \rangle.$$

Thus a curve $\mathbf{x}(s)$ on a surface is a line of curvature if and only if $\tau_g = 0$ along the curve.

The product of the prinicipal curvatures, $K = \kappa_1\kappa_2$, is called the *Gaussian curvature* of the surface. The Gaussian curvature has the following geometric interpretation. Let S^2 be the unit sphere in $\mathbb{R}^3$ centered at the origin. For a surface $M \subset \mathbb{R}^3$ we now define a map $\nu : M \longrightarrow S^2$, called the *Gauss map.* For a point $p \in M$, move the unit normal $\mathbf{n}(p)$ parallel to itself to the origin, and let $\nu(p)$ be the point on the sphere with position vector $\mathbf{n}(p)$. Now let $\mathcal{U}$ be a small neighborhood of p on M; its image $\nu(\mathcal{U})$ is a neighborhood of $\nu(p)$ on the sphere. Then, taking the limit as $\mathcal{U}$ shrinks to p, it can be shown that

$$K(p) = \pm \lim_{\mathcal{U} \to p} \frac{\text{Area}(\nu(\mathcal{U}))}{\text{Area}(\mathcal{U})}.$$

The most remarkable property of the Gaussian curvature is the famous "Theorema Ergregium" of Gauss that K is intrinsic to the surface; that is, despite its definition or the above geometric interpretation, which depends on how the surface sits in $\mathbb{R}^3$, K depends only on the functions E, F, G of the first fundamental form and their derivatives.

If the surface M is a sphere of radius a, then every point is an umbilical point, and hence the two principal curvatures are everywhere equal to $\frac{1}{a}$. Thus the Gaussian curvature of a sphere of radius a is $\frac{1}{a^2}$.

Exercises

1. Let $\mathbf{x}(t)$ be a plane curve and recall its velocity vector $\mathbf{x}'(t) = \frac{ds}{dt}\mathbf{T}$. Show that its acceleration $\mathbf{x}''(t)$ and its curvature are given by

$$\mathbf{x}''(t) = \frac{d^2s}{dt^2}\mathbf{T} + \kappa\left(\frac{ds}{dt}\right)^2\mathbf{N} \quad \text{and} \quad \kappa = \frac{\langle \mathbf{x}''(t), \mathbf{N} \rangle}{\left(\frac{ds}{dt}\right)^2}.$$

2. As we noted in passing, for a space curve $\mathbf{x}(s)$ we also have the Frenet equation $\frac{d\mathbf{T}}{ds} = \kappa\mathbf{N}$. If the curve lies on a surface and $\mathbf{v} = \mathbf{n}\times\mathbf{T}$, then

$$\kappa\mathbf{N} = \kappa_n\mathbf{n} + \kappa_g\mathbf{v};$$

κ_n and κ_g are called the *normal curvature* and *geodesic curvature* of the curve on the surface. If $\mathbf{x}(s)$ is a normal section corresponding to a principal direction at a point p, show that κ_n is the principal curvature and $\kappa_g = 0$ at p.

3. If $\mathbf{x}(s)$ is a curve on a surface, show that $\mathbf{x}(s)$ is a line of curvature if and only if $\frac{d\mathbf{n}}{ds} + \kappa_n\mathbf{T} = 0$, where κ_n is the normal curvature of $\mathbf{x}(s)$ (Theorem of Rodriques).

4. If two surfaces intersect along a curve that is a line of curvature for both surfaces, prove that the angle between the (tangent planes of the) surfaces is a constant along the curve. Conversely, if two surfaces intersect along a curve that is a line of curvature of one of them and the angle between the surfaces is a constant along the curve, prove that the curve is a line of curvature of the other surface (Theorem of Joachimsthal).

6.2. The classical proof

To give the classical proof of Liouville's theorem, we need the idea of a triply orthogonal family of surfaces and a theorem of Dupin. A *triply orthogonal system* consists of three families of surfaces in an open set in $\mathbb{R}^3$ with one surface from each family passing through each point and such that the tangent planes at each point are mutually perpendicular, as, for example, the level surfaces of three smooth functions of three variables for which their gradient vectors are mutually orthogonal. Two particular examples are the following:

Example 6.1. *In cartesian coordinates x^1, x^2, x^3, the planes $x^i =$ const., $i = 1, 2, 3$, form a triply orthogonal system.*

Example 6.2. *The family of spheres concentric about the origin, the family of right circular cones with axes coinciding with the x^3-axis, and the family of planes through the x^3-axis, form a triply orthogonal* *system; these are the level surfaces of the standard spherical coordinate functions.*

We now prove the following theorem of Dupin.

Theorem 6.2. *The surfaces of a triply orthogonal system intersect each other in lines of curvature.*

Proof. Let M_1, M_2, M_3 be three surfaces, one from each family, and let $\mathbf{x}_1(s) = M_2 \cap M_3$, $\mathbf{x}_2(s) = M_3 \cap M_1$, $\mathbf{x}_3(s) = M_1 \cap M_2$ be the curves of intersection parametrized by arc length with $\mathbf{x}_1', \mathbf{x}_2', \mathbf{x}_3'$ a right-handed triple at $p = M_1 \cap M_2 \cap M_3$. As in our discussion of geodesic torsion, set $\mathbf{v}_{ab} = \mathbf{n}_b \times \mathbf{T}_a$, $a, b = 1, 2, 3$. Consider $\mathbf{x}_1(s)$. Then as a curve on M_2, $\mathbf{v}_{12} = \mathbf{n}_2 \times \mathbf{T}_1 = -\mathbf{n}_3$; as a curve on M_3, $\mathbf{v}_{13} = \mathbf{n}_3 \times \mathbf{T}_1 = \mathbf{n}_2$. Now on M_2, the geodesic torsion is given by

$$\left\langle \frac{d\mathbf{n}_2}{ds}, \mathbf{v}_{12} \right\rangle = \left\langle \frac{d\mathbf{n}_2}{ds}, -\mathbf{n}_3 \right\rangle = \left\langle \mathbf{n}_2, \frac{d\mathbf{n}_3}{ds} \right\rangle = \left\langle \mathbf{v}_{13}, \frac{d\mathbf{n}_3}{ds} \right\rangle,$$

which is the geodesic torsion on M_3; we denote this common value of the geodesic torsion by τ_g^1. Also from this computation note that

$$\left\langle \mathbf{n}_2, \frac{d\mathbf{n}_3}{ds} \right\rangle = -\langle A_3 \mathbf{T}_1, \mathbf{n}_2 \rangle.$$

For the curve $\mathbf{x}_2(s)$,

$$\tau_g^2 = \left\langle \frac{d\mathbf{n}_3}{ds}, \mathbf{v}_{23} \right\rangle = -\left\langle \frac{d\mathbf{n}_3}{ds}, \mathbf{n}_1 \right\rangle = \langle A_3 \mathbf{T}_2, \mathbf{n}_1 \rangle.$$

Therefore at the point p of intersection of the three surfaces we have

$$\tau_g^1 + \tau_g^2 = -\langle A_3 \mathbf{T}_1, \mathbf{n}_2 \rangle + \langle A_3 \mathbf{T}_2, \mathbf{n}_1 \rangle.$$

Noting that $\mathbf{T}_a = \mathbf{n}_a$ and using the symmetry of the Weingaten map A_3, we have

$$\tau_g^1 + \tau_g^2 = -\langle A_3 \mathbf{n}_1, \mathbf{n}_2 \rangle + \langle A_3 \mathbf{n}_2, \mathbf{n}_1 \rangle = 0.$$

Similarly $\tau_g^2 + \tau_g^3 = 0$ and $\tau_g^3 + \tau_g^1 = 0$, and therefore $\tau_g^1 = \tau_g^2 = \tau_g^3 = 0$. □

We can now give the following theorem of Liouville and its proof; recall that conformal maps are regarded as being non-singular.

Theorem 6.3. *Let f be a one-to-one C^3 conformal map of an open set $U \subset \mathbb{R}^3$ onto $f(U)$. Then f is a composition of similarities and inversions.*

Proof. Let $M \subset U$ be a piece of a plane or sphere, $p \in M$ and $\mathbf{v} \in T_pM$. By rotating and translating the triply orthogonal systems of Examples 6.1 or 6.2, as necessary, we obtain a triply orthogonal family of surfaces with M in one of the families and with $\mathbf{v}$ tangent to a curve of intersection. Now by the conformality f maps this triply orthogonal system to another triply orthogonal system. By the theorem of Dupin the curves of intersection on $f(M)$ are lines of curvature. Therefore, since f is non-singular, there exists a line of curvature in any direction at any point of $f(M)$. Therefore every point of $f(M)$ is umbilic, and hence f maps (pieces of) planes and spheres to (pieces of) planes and spheres. The result now follows from Theorem 5.6 of Möbius. $\square$

Chapter 7

When Does Inversion Preserve Convexity?

7.1. Curve theory and convexity

In this chapter we pose and answer the following question: Given a smooth closed convex curve in the plane, what is the set of points in the plane as centers of inversion for which the image of the given curve will again be a convex curve? In this section we continue our treatment of curve theory from Section 6.1, and then in the next section we prove our result. Since the proof involves checking the convexity of the image curve, one of the main ideas of the present section is a test for convexity.

In Section 6.1 we introduced the unit tangent and principal normal to a plane curve and the signed curvature κ (see Figure 6.1).

Recall also the simple example, that, for the circle $\mathbf{x} = a\cos\theta\mathbf{i} + a\sin\theta\mathbf{j}$ of radius a, arc length is given by $s = a\theta$ and $\phi = \theta + \frac{\pi}{2}$. Thus $\kappa(s) = \frac{d\phi}{ds} = \frac{1}{a}$. The circle tangent to a curve at a point $\mathbf{x}(s)$, of radius $\frac{1}{|\kappa(s)|}$ and on the side of the tangent line determined by the direction of $\frac{d\mathbf{T}}{ds}$ is called the *osculating circle* of the curve at $\mathbf{x}(s)$. If κ is zero at $\mathbf{x}(s)$, the tangent line has "second order contact" at $\mathbf{x}(s)$, and we call this line the *osculating circle*; moreover, since we will be dealing with convex curves, the *interior* of the osculating circle will, in that case, be the open half plane containing the curve.

Example 7.1. *For example, $x^4 + y^4 = 1$ is a simple closed convex curve whose curvature at the points $(\pm 1, 0)$ and $(0, \pm 1)$ is zero and whose osculating circle is the tangent line. At all other points the osculating circle is indeed a circle; e.g. at $(1/2^{1/4}, 1/2^{1/4})$ the osculating circle is*

$$\left(x - \frac{2}{3(2^{1/4})}\right)^2 + \left(y - \frac{2}{3(2^{1/4})}\right)^2 = \frac{\sqrt{2}}{9}.$$

It is well know that an arc of a smooth curve $\mathbf{x}(s)$ with increasing positive curvature determined by an interval $[s_0, s_1]$ lies within the osculating circle at $\mathbf{x}(s_0)$. It is less well-known that the entire osculating circles at the points of the arc lie in the osculating circle at $\mathbf{x}(s_0)$. This is a result of Kneser (1912) (see [**17**]), which we give here as a lemma.

Lemma 7.1. *Any osculating circle of an arc of a smooth curve with monotonic curvature of constant sign contains every smaller osculating circle of the arc and is contained in every larger osculating circle of the arc.*

Proof. Let $\mathbf{x}(s)$ be a smooth curve and $\mathbf{x}(s_0)$ a point on the curve. Suppose that $\kappa(s)$ is positive and increasing on $[s_0, s_1]$. We shall show that for every $s \in (s_0, s_1]$ the osculating circle at $\mathbf{x}(s)$ lies in the osculating circle at $\mathbf{x}(s_0)$. Let $\mathbf{c}(s) = \mathbf{x}(s) + \frac{1}{\kappa}\mathbf{N}$ be the curve of centers of the osculating circles, and denote differentiation with respect to s by $'$. Then $\mathbf{c}' = -\frac{\kappa'}{\kappa^2}\mathbf{N}$, and hence, setting $R = \frac{1}{\kappa(s)}$ and $R_0 = \frac{1}{\kappa(s_0)}$,

$$|\mathbf{c}(s) - \mathbf{c}(s_0)| \leq \int_{s_0}^{s} \frac{\kappa'(u)}{\kappa(u)^2}du = R_0 - R.$$

Thus, if $\mathbf{y}$ is in or on the osculating circle at $\mathbf{x}(s)$, then

$$|\mathbf{y} - \mathbf{c}(s_0)| \leq |\mathbf{y} - \mathbf{c}(s)| + |\mathbf{c}(s) - \mathbf{c}(s_0)| \leq R + R_0 - R = R_0,$$

completing the proof. □

A set of points in the plane (or in $\mathbb{R}^n$) is said to be *convex* if for any two points in the set, the line segment joining them is contained in the set. A simple closed curve in the plane is said to be *convex* if its union with its interior is a convex set.

In some differential geometry books where the emphasis is on smooth curves, we find the following definition. A regular curve in the plane is *convex* if it lies on one side of each tangent line. This is not unreasonable in view of the following basic theorem of convexity theory. A *supporting line* for a set in the plane with interior points is a line through a boundary point such that all points of the set are in the same closed half plane determined by the line.

Theorem 7.1. *A simple closed curve is convex if and only if through each of its points there is at least one supporting line for the interior.*

For our study of the problem of when inversion preserves convexity we well need a test for convexity in differential geometric terms. Fortunately we have the following classical and natural theorem at our disposal (cf. [**23**], Section 3.3).

Theorem 7.2. *A simple closed regular smooth plane curve* $\mathbf{x}(s)$ *is convex if and only if* $\kappa(s)$ *does not change sign.*

Proof. Suppose κ does not change sign and $\mathbf{x}(s)$ is not convex. Then there is a point A such that $\mathbf{x}(s)$ does not lie on one side of the tangent line l at A. Since $\mathbf{x}(s)$ is closed, there are points B and C on the curve on opposite sides of l which are farthest from l. Now the tangent lines at A, B and C must be distinct and mutually parallel. Thus at two of A, B, C the tangent vectors point in the same direction. Therefore there exist $s_1 < s_2$ with $\mathbf{T}(s_1) = \mathbf{T}(s_2)$ and $\phi(s_2) = \phi(s_1) + 2\pi n$. Since $\kappa(s) = \frac{d\phi}{ds}$ does not change sign, ϕ is a monotonic function of s. If $n = 0$, then ϕ is constant on $[s_1, s_2]$. If ϕ is non-decreasing and $n \neq 0$, then $n = 1$, since $\mathbf{x}(s)$ is simple (this implication, as obvious as it seems, also requires proof and it is not entirely trivial; it is often refered to as the *Rotation Index Theorem*, see e.g. [**23**], Section 3.2). Therefore ϕ is constant on $[0, s_1]$ and $[s_2, L]$, where L is the length of the closed curve. Therefore one of the segments between $\mathbf{x}(s_1)$ and $\mathbf{x}(s_2)$ is a straight line, contradicting the distinctness of the tangent lines at A, B and C.

Conversely, suppose that $\mathbf{x}(s)$ is convex but that ϕ is not monotonic. Then there exist $s_1 < s_0 < s_2$ in $[0, L]$ with $\phi(s_1) = \phi(s_2) \neq$

$\phi(s_0)$. We shall contradict this by showing that ϕ is constant on $[s_1, s_2]$. Now $\mathbf{T}(s_1) = \mathbf{T}(s_2)$, but since $\mathbf{x}(s)$ is a simple closed curve, $\mathbf{T}$ points in every direction. Thus there exists s_3 with $\mathbf{T}(s_3) = -\mathbf{T}(s_1)$. If the tangent lines at s_1, s_2, s_3 are distinct, then one is between the other two, contradicting the convexity. Therefore two of these lines coincide and there are points A and B of $\mathbf{x}(s)$ lying on the same tangent line l. If $C \in AB$ and C is not on $\mathbf{x}(s)$, let l' be the perpendicular to l at C. Again by convexity, l' is not tangent to $\mathbf{x}(s)$, and hence it meets $\mathbf{x}(s)$ at two points on the same side of l, say D and E, with D closer to l. Then at least one of A, B and E is on each side of the tangent at D, again contradicting convexity. Thus C must lie on $\mathbf{x}(s)$, and hence ϕ is constant on $[s_1, s_2]$, completing the proof. □

7.2. Inversion and convexity

We now turn to the question raised at the beginning of this chapter, namely: Given a smooth closed convex curve in the plane, what is the set of points in the plane for which, when taken as centers of inversion, the image of the give curve will again be a convex curve? The question was first prompted by a question in fluid mechanics where the given curve was an ellipse, and this special case of the result was obtained by J. B. Wilker [**28**]. The general question was answered by Wilker and the author in [**5**].

Theorem 7.3. *Let C be a smooth simple closed convex curve. A point interior to C is a center of inversion preserving convexity if and only if it lies in the intersection of the interiors and boundaries of the osculating circles at the maxima of the curvature (C traversed counterclockwise). A point exterior to C is such a center if and only if it lies in the intersection of the exteriors and boundaries of the osculating circles at the minima of the curvature. If C is not a circle, inversion with center on C does not preserve convexity.*

Proof. The main part of the proof will be to show that a point serving as a center of inversion preserving the convexity of the curve lies in the intersection of the interiors (resp. exteriors) and boundaries of all the osculating circles. The fact that it is enough to take the osculating circles at the maxima and minima of the curvature is then a consequence of the lemma of Kneser.

Since the composition of two inversions with the same center is a homothety (Theorem 5.3), which preserves convexity, the radius of the circle of inversion is immaterial in our problem and we take it to be 1. If the vector $\mathbf{x}_0$ denotes a center of inversion and $\mathbf{x}^*$ the inverse of a point $\mathbf{x}$, then inversion in the unit circle about $\mathbf{x}_0$ is given by

$$\mathbf{x}^* = \mathbf{x}_0 + \frac{\mathbf{x} - \mathbf{x}_0}{|\mathbf{x} - \mathbf{x}_0|^2}.$$

As $\mathbf{x}$ traverses $\mathcal{C}$ we have the following computations:

$$\mathbf{x}^{*\prime} = \frac{1}{|\mathbf{x} - \mathbf{x}_0|^2}\mathbf{T} - \frac{2\langle \mathbf{T}, \mathbf{x} - \mathbf{x}_0\rangle}{|\mathbf{x} - \mathbf{x}_0|^4}(\mathbf{x} - \mathbf{x}_0),$$

$$|\mathbf{x}^{*\prime}| = \frac{ds^*}{ds} = \frac{1}{|\mathbf{x} - \mathbf{x}_0|^2}, \qquad \mathbf{T}^* = \mathbf{T} - \frac{2\langle \mathbf{T}, \mathbf{x} - \mathbf{x}_0\rangle}{|\mathbf{x} - \mathbf{x}_0|^2}(\mathbf{x} - \mathbf{x}_0),$$

$$\left\langle \frac{d\mathbf{T}^*}{ds^*}, \mathbf{x}^* - \mathbf{x}_0 \right\rangle = -\kappa\langle \mathbf{N}, \mathbf{x} - \mathbf{x}_0\rangle - 2 + 2\frac{\langle \mathbf{T}, \mathbf{x} - \mathbf{x}_0\rangle^2}{|\mathbf{x} - \mathbf{x}_0|^2}.$$

Also

$$\frac{\langle \mathbf{N}^*, \mathbf{x}^* - \mathbf{x}_0\rangle}{|\mathbf{x}^* - \mathbf{x}_0|} = \frac{\langle \mathbf{N}, \mathbf{x} - \mathbf{x}_0\rangle}{|\mathbf{x} - \mathbf{x}_0|}$$

and

$$\left(\frac{\langle \mathbf{T}, \mathbf{x} - \mathbf{x}_0\rangle}{|\mathbf{x} - \mathbf{x}_0|}\right)^2 + \left(\frac{\langle \mathbf{N}, \mathbf{x} - \mathbf{x}_0\rangle}{|\mathbf{x} - \mathbf{x}_0|}\right)^2 = 1.$$

Thus from $d\mathbf{T}^*/ds^* = \kappa^*\mathbf{N}^*$ we have

$$\kappa^* = -\kappa|\mathbf{x} - \mathbf{x}_0|^2 - 2\langle \mathbf{N}, \mathbf{x} - \mathbf{x}_0\rangle.$$

We now regard $\mathcal{C}$ as being traversed counterclockwise, so that $\kappa \geq 0$. If $\mathbf{x}_0$ is interior to $\mathcal{C}$, the inverted curve will also be traversed counterclockwise, and hence, if it is also convex, then $\kappa^* \geq 0$ (see Figure 7.1). Therefore $-\kappa|\mathbf{x} - \mathbf{x}_0|^2 - 2\langle \mathbf{N}, \mathbf{x} - \mathbf{x}_0\rangle \geq 0$. Writing in standard cartesian coordinates $\mathbf{N} = a\mathbf{i} + b\mathbf{j}$, $\mathbf{x} = (x, y)$ and setting $R = \frac{1}{\kappa}$ for $\kappa \neq 0$, this inequality becomes

$$(x_0 - (x + Ra))^2 + (y_0 - (y + Rb))^2 \leq R^2.$$

Therefore $\mathbf{x}_0 = (x_0, y_0)$ lies in or on the osculating circle at $\mathbf{x} \in \mathcal{C}$. For $\kappa = 0$ the inequality becomes

$$a(x_0 - x) + b(y_0 - y) \geq 0,$$

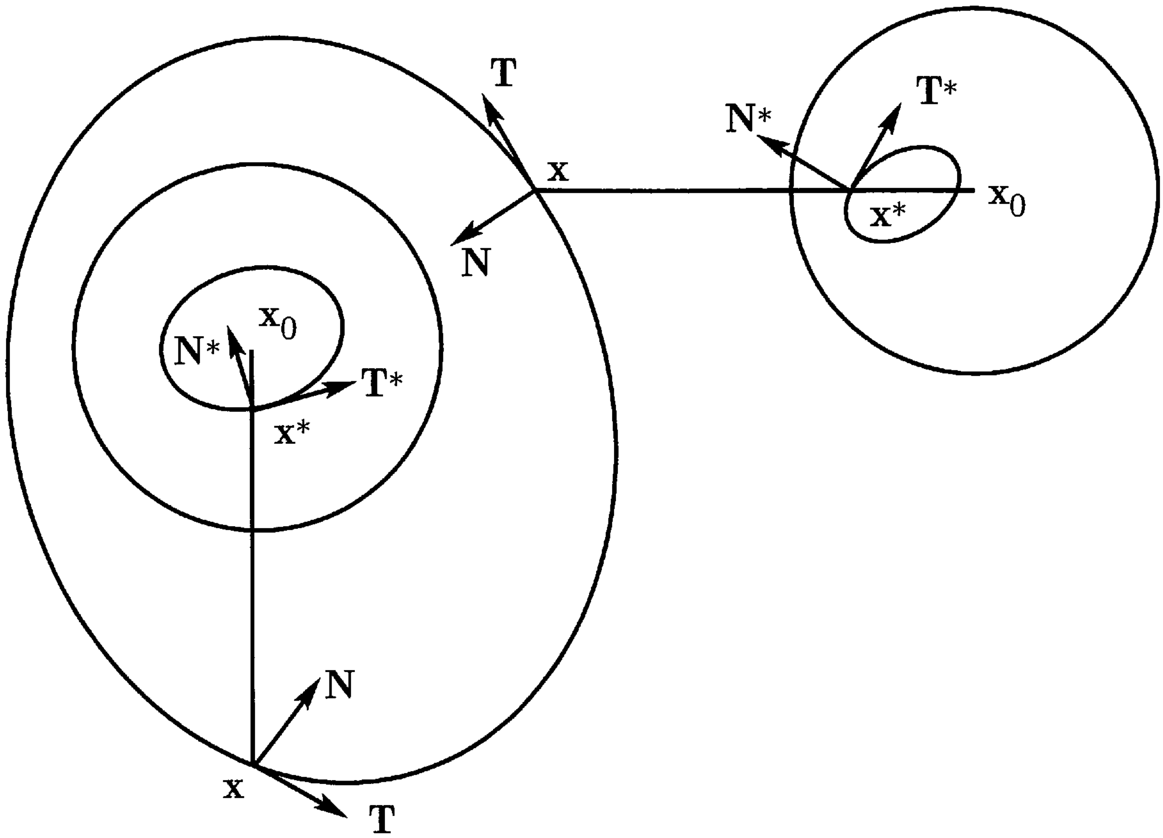

Figure 7.1

and $\mathbf{x}_0$ is in the desired half plane.

If $\mathbf{x}_0$ is exterior to $\mathcal{C}$ (see Figure 7.1), the above inequalities are reversed. Thus a point serving as a center of inversion preserving the convexity of the curve lies in the intersection of the interiors (resp. exteriors) and boundaries of all the osculating circles. Now by Lemma 7.1 we see that it is enough to take the osculating circles at the maxima and minima of the curvature. Conversely, one can reverse our argument.

Finally, if $\mathcal{C}$ is a circle, the set of centers of inversion preserving convexity is the entire plane. If $\mathcal{C}$ is not a circle, inversion about any point $\mathbf{x}_0 \in \mathcal{C}$ destroys the convexity. For consider the osculating circle at $\mathbf{x}_0$ and invert both it and the curve with center $\mathbf{x}_0$; the circle inverts to a line and $\mathcal{C}$ inverts to a curve which is asymptotic to the line in both directions and is therefore not convex. $\square$

We remark that for some convex curves there may be no inversions that will preserve convexity, e.g. the curve of Example 7.1.

Our proof was given in the context of real geometry, but it is interesting to relate some geometric ideas to complex analysis, as we do in the following exercises. In the exercises below we suppose that $f(z)$ is holomorphic and one-to-one on the unit disk, and consider the image of a circle $r = \text{const.} < 1$. Typically one also assumes that $f(0) = 0$. For further study along these lines, see e.g. [**15**]. Recall that in Exercise 4 of Section 4.2 we identified a vector $\mathbf{v} = v^1\mathbf{i} + v^2\mathbf{j}$ with the complex number $v = v^1 + iv^2$ and saw that the inner product of two vectors was given by $\langle \mathbf{v}, \mathbf{w} \rangle = \Re v\bar{w}$.

EXERCISES

1. For the curve $w = f(re^{i\theta})$, $r = \text{const.} < 1$, with θ as increasing parameter, show that the velocity vector, the principal normal and the acceleration are given by

$$\frac{dw}{d\theta} = izf'(z), \quad \mathbf{N} = -\frac{zf'(z)}{|zf'(z)|}, \quad \frac{d^2w}{d\theta^2} = -z(f'(z) + zf''(z)).$$

2. Using Exercise 1 of Section 6.1, show that the curvature of the above curve is

$$\kappa = \frac{1}{|zf'(z)|}\Re\Big(1 + \frac{zf''(z)}{f'(z)}\Big),$$

and hence that $w = f(re^{i\theta})$ is convex if and only if

$$\Re\Big(1 + \frac{zf''(z)}{f'(z)}\Big) \geq 0$$

(E. Study, 1913).

3. In Exercise 3 of Section 4.2 we saw that the Koebe function $f(z) = \frac{z}{(1-z)^2}$ maps the unit disk onto a slit plane. Show that for $2 - \sqrt{3} < r < 1$, the curvature of the curve $w = f(re^{i\theta})$ at the point $f(-r)$ is negative.

4. A smooth simple closed curve C is said to be *strongly star-shaped* with respect to a point P if no tangent line passes through P. Show that the curve $w = f(re^{i\theta})$, $r = \text{const.} < 1$, with $f(0) = 0$ is strongly star-shaped with respect to the origin if and only if

$$\Re\Big(\frac{zf'(z)}{f(z)}\Big) > 0.$$

7.3. The problem for convex bodies

In this section we state without proof the corresponding result in Euclidean 3-space. That is, given a smooth convex surface in space, what is the set of points as centers of inversion for which the image of the surface will again be convex? The main ingredient in the above proof for convex curves was Theorem 7.2, giving a test for convexity. For surfaces in 3-space we do have such a test. The following theorem is due to Hadamard [**18**] for positive Gaussian curvature and to Chern and Lashof [**7**] for non-negative Gaussian curvature. See Section 6.1 for a discussion of the Gaussian curvature.

Theorem 7.4. *A smooth closed surface M in $\mathbb{R}^3$ is convex if and only if $K \geq 0$.*

The initial difficulty with this problem is that, given a surface, one does not in general have an osculating sphere. One could define a sphere at each point of the surface, tangent to the surface, on the appropriate side of the tangent plane and whose Gaussian curvature equals the Gaussian curvature of the surface at the point. Such a sphere has second order contact with the surface only at an umbilical point. Recall, however, that we do have two principal curvatures κ_1 and κ_2, which for $K > 0$ have the same sign (again see Section 6.1). Thus we define the *principal spheres* at $p \in M$ to be the spheres of radius $1/|\kappa_i|$, $i = 1, 2$, $\kappa_i \neq 0$, tangent to M at p and on the side of the tangent plane containing the surface (our surfaces being convex). If one (or both) of the κ_i vanish, define the corresponding *principal sphere* to be the tangent plane and its "interior" to be the half space containing the surface.

We can now state the result given by J. B. Wilker and the author in [**5**].

Theorem 7.5. *Let M be a smooth closed convex surface in $\mathbb{R}^3$. A point interior to M is a center of inversion preserving convexity if and only if it lies in the intersection of the interiors and boundaries of all the (small) principal spheres. A point exterior to M is such a center if and only if it lies in the intersection of the exteriors and boundaries of all the (large) principal spheres. If M is not a sphere, inversion with center on M does not preserve convexity.*

Bibliography

[1] Ahlfors, L., *Complex Analysis*, 3rd ed., McGraw-Hill, New York, 1979.

[2] Apostol, T., *Mathematical Analysis*, 2nd ed., Addison-Wesley, Reading, MA, 1974.

[3] Aumann, G., *Distortion of a segment under conformal mapping and related problems*, C. Carathéodory International Symposium (Athens, 1973), Greek Math. Soc., Athens, 1974, pp. 46-53.

[4] Bartle, R. G., *The Elements of Real Analysis*, 2nd ed., Wiley, New York, 1976.

[5] Blair, D. E. and Wilker, J. B., *When does inversion preserve convexity?* Kōdai Math. J. **6** (1982), 186-192.

[6] Carathódory, C., *The most general transformations of plane regions which transform circles into circles*, Bull. Amer. Math. Soc. **43** (1937), 573-579.

[7] Chern, S. S. and Lashof, R. K., *On the total curvature of immersed manifolds.* II, Michigan Math. J. **5** (1958), 5-12.

[8] Churchill, R. and Brown, J., *Complex Variables and Applications*, 6th ed., McGraw-Hill, New York, 1996.

[9] do Carmo, M., *Riemannian Geometry*, Birkhäuser, Boston et al., 1992.

[10] Dubrovin, B. A., Fomenko, A. T. and Novikov, S. P., *Modern Geometry, Methods and Applications*, Part 1, 2nd. ed, Springer-Verlag, Berlin et al., 1992.

[11] Eves, H. *Modern Elementary Geometry*, Jones and Bartlett, Boston, London, 1992.

[12] Ewald, G., *Geometry: An Introduction*, Wadsworth, Belmont, CA, 1971.

[13] Fisher, S. D., *Complex Variables*, 2nd ed., Wadsworth & Brooks/Cole, Pacific Grove, CA, 1990.

[14] Fleming, W., *Functions of Several Variables*, 2nd ed., Springer-Verlag, Berlin et al., 1977.

[15] Goodman, A. W., *Univalent Functions*, Mariner, Tampa, FL, 1983.

[16] Greenberg, M. J., *Euclidean and Non-Euclidean Geometries*, 3rd ed., Freeman, New York 1993.

[17] Guggenheimer, H., *Differential Geometry*, McGraw Hill, New York, 1963.

[18] Hadamard, J., *Sur certaines propriétés des trajectoires en dynamique*, J. Math. Pures Appl. (5) **3** (1897), 331-387.

[19] Hartman, P., *On isometries and on a theorem of Liouville*, Math. Z. **69** (1958), 202-210.

[20] Huff, M., *Conformal maps on Hilbert space*, Bull. Amer. Math. Soc. **82** (1976), 147-149.

[21] Lehto, O. and Virtanen, K. I., *Quasiconformal Mappings in the Plane*, Springer-Verlag, Berlin et al., 1973.

[22] Liouville, J., *Extension au cas des trois dimensions de la question du trace géographique*, Note VI in the Appendix to G. Monge, *Application de l'Analyse à la Géométrie*, 5th ed. (J. Liouville, editor), Bachelier, Paris, 1850, pp. 609-616.

[23] Millman, R. S. and Parker, G. D., *Elements of Differential Geometry*, Prentice-Hall, Englewood Cliffs, NJ, 1977.

[24] Miquel, A., *Théorèmes de géométrie*, J. Math. Pures Appl. **3** (1838), 485-487.

[25] Miquel, A., *Mémoire de géométrie*, J. Math. Pures Appl. **9** (1844), 20-27.

[26] Nevanlinna, R., *On differentiable mappings*, in Analytic Functions, Princeton Math. Ser., vol. 24, Princeton Univ. Press, 1960, pp. 3-9.

[27] Spivak, M. *Differential Geometry*, Vol. 3, Publish or Perish, Boston, 1975.

[28] Wilker, J. B., *When is the inverse of an ellipse convex?* Utilitas Math. **17** (1980), 45-50.

Index

Inversion Theory and Conformal Mapping

David E. Blair

It is rarely taught in an undergraduate or even graduate curriculum that the only conformal maps in Euclidean space of dimension greater than two are those generated by similarities and inversions in spheres. This is in stark contrast to the wealth of conformal maps in the plane.

The principal aim of this text is to give a treatment of this paucity of conformal maps in higher dimensions. The exposition includes both an analytic proof in general dimension and a differential-geometric proof in dimension three. For completeness, enough complex analysis is developed to prove the abundance of conformal maps in the plane. In addition, the book develops inversion theory as a subject, along with the auxiliary theme of circle-preserving maps. A particular feature is the inclusion of a paper by Carathéodory with the remarkable result that any circle-preserving transformation is necessarily a Möbius transformation—not even the continuity of the transformation is assumed.

The text is at the advanced undergraduate level and is suitable for a capstone course, topics course, senior seminar or independent study. Students and readers with university courses in differential geometry or complex analysis bring with them background to build on, but such courses are not essential prerequisites.

ISBN 0-8218-2636-0

9 780821 826362

STML/9