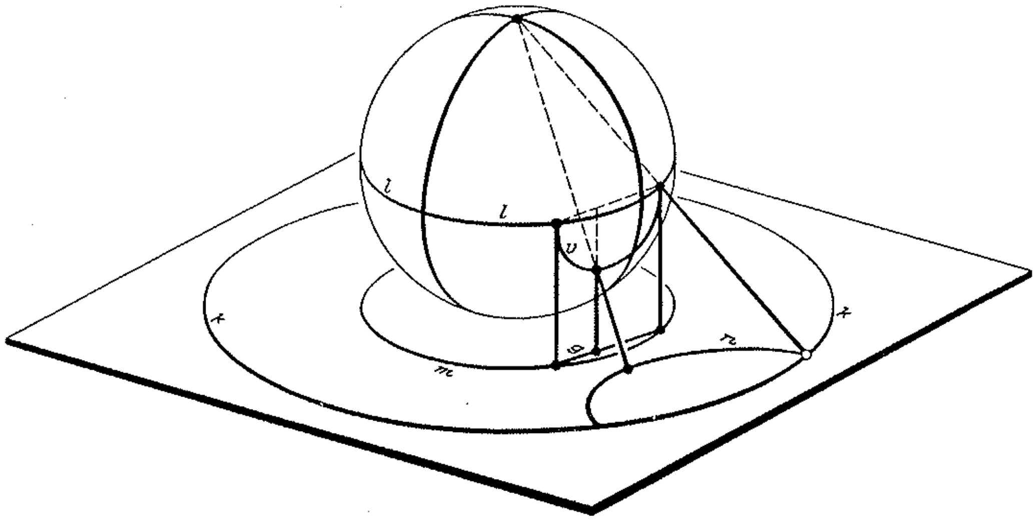


FIG. 246

Reader in
Classical Geometries
(Math 128A)

Table of Contents

1.

~~NOTES ON~~ WRITING UP PROOFS.

1. Make sure you understand the statement of the problem. What are the 'givens'? What are you to show?

2. The mental work: now you have to get from the givens to what there is to show. This often involves constructing another line segment or point which was not in the original problem.

3. Make a rough draft of your solution. Feel free to fill it with scribbles and equations and symbols. I DO NOT WANT TO SEE the rough draft. But it will help you get to a good write-up. And it may be essential to complete the crucial step, step 2!

4. The write-up itself. Start by restating what you are given and what it is you are to prove.

5. If there are lemmas or earlier exercises that your proof uses, and that are not so obvious, or are somewhat involved it is often helpful to state these seperately.

(SAS, the sum of the angles of a triangle are 180 DO NOT fall into this category.)

6. Break the write-up into bite-size (paragraph or smaller) pieces if the proof is long.

7. Include figures as necessary, or helpful. Feel free to refer to parts of your figure eg. "angle α of the figure". But do draw the figure clearly, without scratched up marks, and do label it clearly.

8. When you are done with the write-up , read what you wrote. Make sure it makes sense. Then read it out loud. For best results, have someone else read it!

! As in writing an English essay, you may need three to five rewrites before you arrive at a really decent proof!

9. Its always fun to end by "QED".

WARNING SIGNS, things to look out for:

a. **Circular arguments.** Make sure your argument is not circular. By this I mean, for example, if you are trying to prove somewhere that "ℓ and m are parallel" that you did say earlier "because ℓ and m are parallel"

b. **Unintelligible sentences.** If you read one of your sentences out loud and it sounds like nonsense, you've got a problem.

c. **Undefined objects.** Do not refer to objects (points, lines, angles,...) that are nowhere defined. If all of a sudden a point Z pops up in your argument this is not good. You must say something like "where Z is as indicated in the figure"' or "where $\vec{P}Z$ is the perpindicular to m passing through P" (assuming m and P were earlier defined.

d. **Repetitive statements.** Do not say the same thing two or three times in different ways. This does not fool me, or you. It just gives us both more to slog through. Try to just say it once, clearly.

2 Erastothene circumf of earth.

2

the world. It was an amazing achievement of human civilization that such an estimate had already been obtained in the 3rd century B.C. by Eratosthenes of Cyrene (284–192 B.C., director of the library of Alexandria). His estimate was based upon the following two pieces of geographic knowledge, namely

(i) At noon of the summer solstice, the sunlight was observed to shine directly down to the bottom of a deep well at Syene (modern Aswan).

(ii) While in Alexandria, which is within 1° of meridian due north of Syene, the sunlight made an angle roughly equal to 1/50 of 360°.

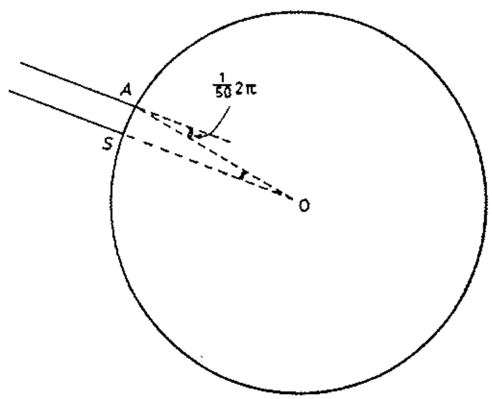


Fig. 36

In applying mathematics to analyze the above geographic facts, he drew a simple diagram (cf. Fig. 36) to represent the "common" meridian circle passing through Syene and Alexandria by a circle and the sunlights at noon of the summer solstice at the above two places by two parallel lines (the reason being that the sun is known to be far, far away as compared to the size of the earth).

Since the sunlight at S (Syene) is perpendicular to the surface it is pointing directly toward the center O of the circle, while the sunlight at A (Alexandria) makes an angle roughly equal to 1/50 o 360° with the direction of $\overline{AO}$. Thus he concluded that the centra angle $\angle AOS$ must be also roughly equal to 1/50 of 360° and hence the circumference of the meridian circle (i.e., the earth) is 50 times the distance between Alexandria and Syene. He estimated the latter simply by the fact that camel caravans, which usually travelled 100 stadia a day, took 50 days to reach Syene from Alexandria. Hence the circumference of the earth is roughly equal to

$$50 \times 50 \times 100 \text{ stadia} \tag{77}$$

which is quite close to the modern measurement of 40 000 km.

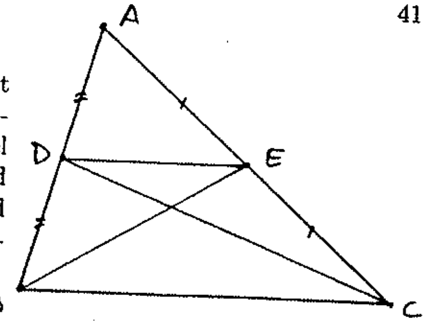
Remark. This is a masterpiece of mathematical analysis and a brilliant example of mathematical abstraction. The simple picture of Fig. 36 concisely organized the geographic events at Alexandria and Syene int ... ell-understood geometric configuration, thus enabled him to d ... ial correlation, namely, that the circumference of the ... the distance between Alexandria and Syene

2.1 ... *force*

To provide some ... kground on the development of integration theory, we discuss here a few pertinent examples which are, in fact, the prelude to modern integral calculus.

Example 2 (Archimedes proof of the area formula of the sphere). Inspired by the monumental success of Eudoxus in applying the approximation methodology to rebuild the foundation of quantitative geometry, geometers of the antiquity were energetically extending the method of approximation to study the areas and volumes of curved objects. They called it the *method of exhaustion*. Archimedes learned

3.11 (Campanus). Use the theory of content to show that the line DE joining the midpoint of two sides of a triangle is parallel to the third side. (*Hint*: Draw BE and DC. Show that the triangles BDC and BEC have the same content and then apply (I.39).)



4 Construction of the Regular Pentagon

One of the most beautiful results in all of Euclid's *Elements* is the construction of a regular pentagon inscribed in a circle (IV.11). The proof of this construction makes use of all the geometry he has developed so far, so that one could say to understand fully this single result is tantamount to understanding all of Euclid's geometry. It also raises questions of exposition which are central to our modern examination of Euclid's methods. For example, why does Euclid use the theory of area in proving a result about a polygon?

In this section we will present Euclid's construction of the regular pentagon, and begin discussing the issues raised by its proof. In a later chapter we will give other proofs using similar triangles or the complex numbers. Euclid's original geometric proof must be regarded as a tour de force of classical geometry. It depends on the theory of area, which we will discuss in more detail later. So this section can be regarded as a taste of things to come: a first meeting with one of the deeper topics which is central to Euclid's geometry.

The key point of the construction of the pentagon is the following problem.

Problem 4.1 To construct an isosceles triangle whose base angles are equal to twice the vertex angle.

Construction ((II.11), (IV.10)). Let A, B be two points chosen at random.

1. Draw line AB.

Next, construct a perpendicular to AB at A, as follows

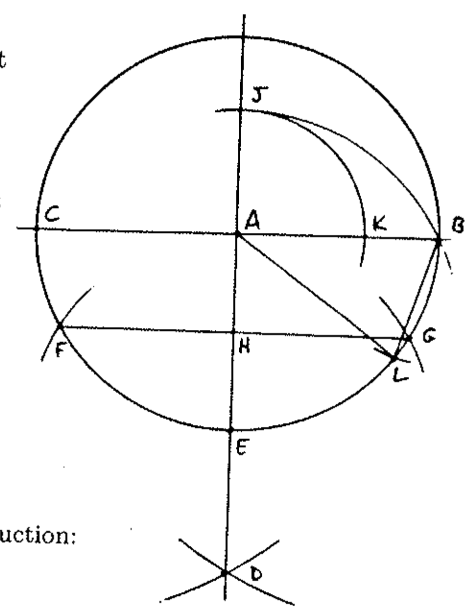
2. Circle AB, get C
3. Circle BC
4. Circle CB, get D
5. Line AD, get E.

Next, we bisect AE as follows

6. Circle EA, get F, G
7. Line FG, get H.

Now comes the unusual part of the construction:

8. Circle HB, get J
9. Circle AJ, get K



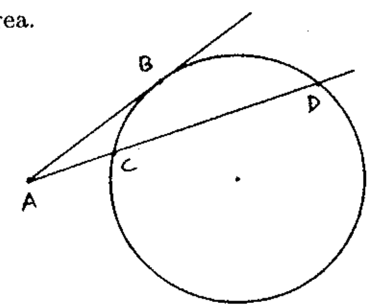
10. Circle center B, radius AK, get L
11. Line AL
12. Line BL.

Then ΔABL is the required triangle. The angles at B and at L will be equal to twice the angle at A.

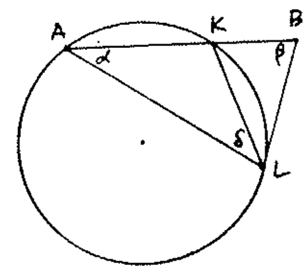
Proof. From a modern point of view, it would seem that some theory of quadratic equations is essential for the proof. Euclid did not have any algebra available to him, but he was able to deal with quantities essentially equivalent to quadratic expressions via the theory of area. We can think of a rectangle as representing the product of its sides, or a square as the square of its side. These areas, without even assigning a numerical value to them, can be manipulated by cutting up and adding or subtracting congruent pieces. In this way Euclid establishes a "geometrical algebra" for manipulating these quantities (always by geometrical methods), which acts as a substitute for our modern algebraic methods.

Let us then trace the steps by which Euclid proves (IV.10), which is the key point in the construction of the regular pentagon. In Book I, especially (I.35)–(I.47) he discusses the areas of triangles and parallelograms, leading up to the famous Pythagorean theorem (I.47) which is stated in terms of area: the square built on the hypotenuse of a right triangle has area equal to the area of the squares on the two sides. The theorem is proved by cutting these areas into triangles, and proving equality of areas using the cutting and pasting methods just developed.

Book II contains a number of results of geometrical algebra, as described above, all stated and proved geometrically in terms of areas. In particular, (II.5), (II.6) and (II.11) are used in the proof of (IV.10). Note that (II.11), which is sometimes called the division of a segment in extreme and mean ratio, states that the interval AB is divided by a point K (in our notation (4.1) above) such that the rectangle formed by BK and AB has area equal to the square on AK. In this way the property of extreme and mean ratio is expressed using area.



From Book III we need (III.36) and its converse (III.37). Proposition (III.36) says that if a point A lies outside a circle, and if AB is tangent to the circle at B, and if ACD cuts the circle at C and D, then the rectangle formed by AC and AD has area equal to the square on AB. This result is proved by several applications of (II.6) and (I.47).



Now Euclid can prove (IV.10) by a brilliant application of (III.37). Let A, K, B, L be as in the construction (4.1) above. Then by (II.11), the rectangle with sides BK and BA has area equal to the square on AK. Since BL was constructed equal to AK, this is also equal to the square on BL.

5 38 36

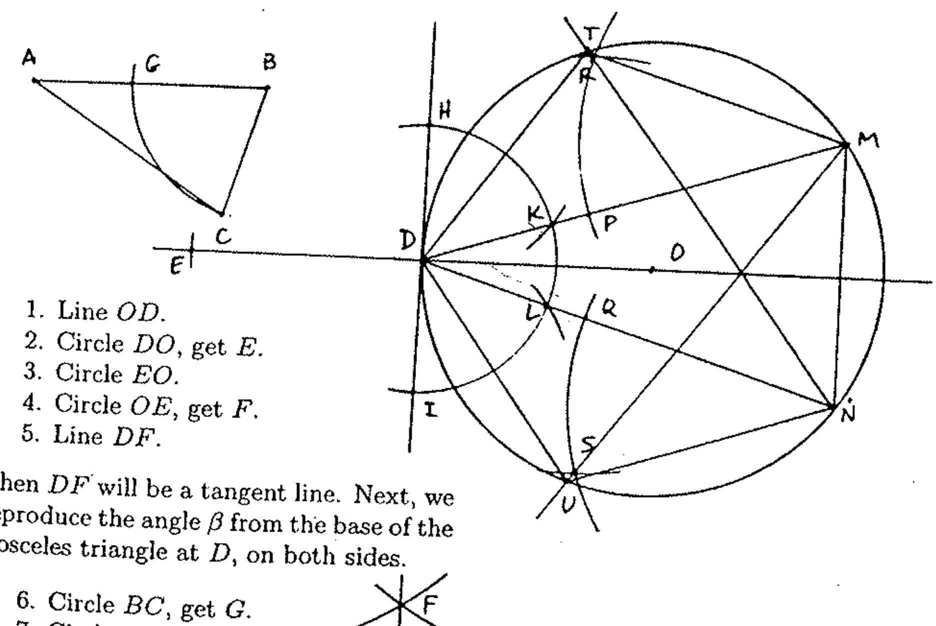
Now consider the circle passing through the three points A, K, L. Since the rectangle on BK and BA is equal to the square on BL, it follows that BL is tangent to this circle (III.37)!

Hence the angle $\angle BLK$ formed by the tangent BL and the line LK is equal to the angle α at A, which subtends the same arc (III.32). Let $\angle KLA = \delta$. Then $\angle BKL$ is an exterior angle to the triangle ΔAKL, so $\angle BKL = \alpha + \delta$ (I.32). But $\angle BLK = \alpha$, so $\alpha + \delta = \angle BLA$, and this angle is β because ΔABL is isosceles. Hence $\angle BKL = \beta$. Now it follows that ΔBKL is isosceles, so $KL = BL = AK$. Hence ΔAKL is also isosceles, so $\delta = \alpha$. Now $\beta = \angle BLA = 2\alpha$ as required. □

Once we have the isosceles triangle constructed in (4.1), the construction of the pentagon follows naturally. The idea is to inscribe in the circle a triangle equiangular with the given triangle, and then to bisect its two base angles.

Problem 4.2 Given an isosceles triangle whose base angles are equal to twice its vertex angle, and given a circle with its center, to construct a regular pentagon inscribed in the circle.

Construction ((IV.2) and (IV.11)). Let ΔABC be the given triangle and let O be the center of the given circle. The first part of the construction is to obtain a tangent line to the circle. Let D be any point on the circle.



1. Line OD.
2. Circle DO, get E.
3. Circle EO.
4. Circle OE, get F.
5. Line DF.

Then DF will be a tangent line. Next, we reproduce the angle β from the base of the isosceles triangle at D, on both sides.

6. Circle BC, get G.
7. Circle at D with radius equal to BC, get H, I.
8. Circle center H radius CG, get K.
9. Circle center I, radius CG, get L.
10. Line DK, get M.

6 37

11. Line DL, get N.
12. Line MN.

Then ΔDMN is a triangle inscribed in the circle, equiangular with ΔABC. Next we bisect the angles at M, N.

13. Circle MN, get P.
14. Circle NM, get Q.
15. Circle QN, get R.
16. Circle PM, get S.
17. Line NR, get T.
18. Line MS, get U.

Then D, M, N, T, U will be the vertices of the pentagon.

19. Line DT.
20. Line TM.
21. Line DU.
22. Line UN.

Then $DTMNU$ is the required pentagon.

Proof. We follow the geometric proof given by Euclid. First of all, the line DF is constructed perpendicular to a diameter of the circle, so it is a tangent line to the circle (III.16). Next, the triangles ΔDHK and ΔDLI are constructed so that their three sides are equal to the three sides of ΔBCG. Hence by (SSS) = (I.8), it follows that $\angle KDH$ and $\angle LDI$ are both equal to the angle β of the triangle ΔABC at B. From there it follows that the angles of ΔDMN at M and N are both equal to β, because they subtend the same arcs cut off by the tangent line and the angles β just constructed (III.32). Since the sum of the three angles of a triangle is constant $= 180°$ (I.32), it follows that the triangle ΔDMN is equiangular with the triangle ΔABC. In particular, if α is the angle at D, then $\beta = 2\alpha$.

The points T, U are constructed by taking the angle bisectors of ΔDMN at M and N. Since the angles at M and N are β, their halves are equal to α. Thus the arcs DT, TM subtend angles α at N; the arc MN subtends an angle α at D; and the arcs DU, UN subtend angles α at M. Hence these five arcs are all equal (III.26), and the line segments on them are also equal. So we have constructed an equilateral pentagon inscribed in the circle. The angle subtended by each side at the center of the circle will be $2\alpha = \beta$. It follows that the angles of the pentagon are also equal, so the pentagon is *regular* in the sense that its sides are all equal and its angles are all equal. □

This completes the presentation of Euclid's construction of the pentagon. As usual, his method is adapted to economy of proof, not economy of steps used. The whole construction, as we have presented it here, takes $12 + 22 = 34$ steps. By collapsing separate parts of the construction, in particular, by constructing the triangle of (4.1) on a radius of the given circle, one can make a construction with less than half as many steps (cf. (4.3)). Note also that Euclid's construction of the points T, U by bisecting the angles at M, N makes possible his elegant proof that the five sides of the pentagon are equal. However in retrospect we see that MN

is actually one side of the pentagon, so T and U could have been constructed in a single step by a circle with center D and radius MN.

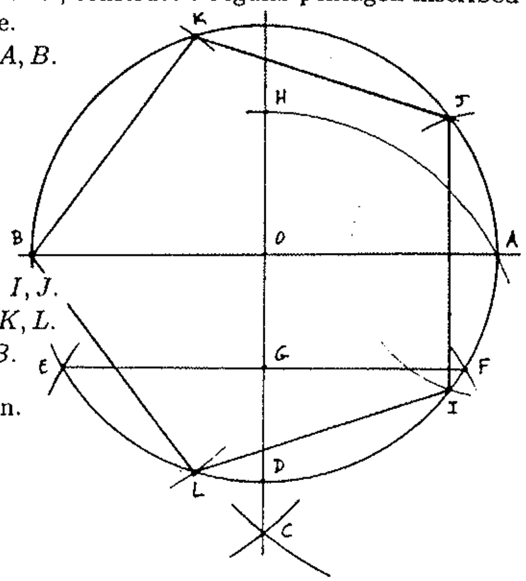
If there is such a thing as beauty in a mathematical proof, I believe that this proof of Euclid's for the construction of the regular pentagon sets the standard for a beautiful proof. In the words of Edna St. Vincent Millay, "Euclid alone has looked on beauty bare".

Now let us use the ideas of Euclid's method to construct a pentagon in as few steps as possible.

Problem 4.3 Given a circle with center O, construct a regular pentagon inscribed in the circle in as few steps as possible.

1. Draw any line through O. Get A, B.
2. Circle A, any radius.
3. Circle B, same radius, get C.
4. OC get D.
5. Circle DO. Get E, F.
6. EF, get G.
7. Circle GA, get H.
8. Circle center A, radius OH, get I, J.
9. Circle center B, radius IJ, get K, L.

10–14. Draw BK, KJ, JI, IL, LB.



Then $BKJIL$ is the required pentagon.

Exercises

4.1 Read Euclid, Book IV.

4.2 Explain why the construction of (4.3) gives a regular pentagon.

4.3 Given a circle, but not given its center, construct an inscribed equilateral triangle in as few steps as possible. (Par = 10).

4.4 Construct a square in as few steps as possible. (Par = 9).

4.5 Given a line segment AB, construct a regular pentagon having AB as a side.

4.6 In the construction of (4.3) in the text, show that AH is equal to the side of the pentagon.

Then, choose C_1 (resp. C_1') on $\overline{AC}$ (resp. $\overline{A'C'}$) such that $\overline{B_1C_1}//\overline{BC}$ (resp. $\overline{B_1'C_1'}//\overline{B'C'}$). It is easy to see that $\Delta AB_1C_1 \cong \Delta A'B_1'C_1'$ and the above proof applies to ΔABC and ΔAB_1C_1 (resp. $\Delta A'B'C'$ and $\Delta A'B_1'C_1'$). Hence

$$\begin{aligned} \overline{AC} &= m\overline{AC_1}, \ \overline{BC} = m\overline{B_1C_1} \\ \overline{A'C'} &= n\overline{A'C_1'}, \ \overline{B'C'} = n\overline{B_1'C_1'} \\ \overline{AC_1} &= \overline{A'C_1'}, \ \overline{B_1C_1} = \overline{B_1'C_1'}. \end{aligned} \tag{8}$$

Thus

$$\overline{AC} : \overline{A'C'} = \overline{BC} : \overline{B'C'} = \frac{m}{n}.$$

□

Historical remarks

(i) At the time of Pythagorean (6th–5th century B.C.), the "universal validity" of commensurability was taken as a self-evident axiom. Therefore, the above proofs of area formula and the similar triangle theorem are regarded as perfectly general and complete.

(ii) The Pythagoras Theorem and the above similar triangle theorem are the two fundamental theorems of quantitative geometry, moreover, the proof of Pythagoras Theorem is necessarily based upon the above area formula. Therefore, the area formula and the similar triangle theorem are truly of fundamental importance.

(iii) Unfortunately, the basic "axiom" of universal validity of commensurability is, in fact, *not* true. The *existence of non-commensurable* pairs of intervals was first discovered by Hippasus, a disciple of Pythagoras. Therefore, the above proofs are *only* the proofs of the *commensurable case* and hence theoretically *incomplete*!

Let us first discuss the discovery of Hippasus.

1.2. *Discovery and Proof of existence of non-commensurable pairs of intervals*

First of all, notice that the commensurability problem is a *purely theoretical* problem whose significance and importance lie in the foundation of geometry. (For practical purposes, any two intervals can be regarded as commensurable simply by omitting a very, very small, i.e., practically nil, remainder.) Therefore, the existence of non-commensurable pairs of intervals can only be demonstrated by theoretical proofs. Let us begin with a theoretical criterion of commensurability.

A criterion of commensurability. Suppose that a and b are a pair of commensurable intervals, namely, there exists a common "yardstick" c such that both a and b are integral multiples of c, say $a = m \cdot c$ and $b = n \cdot c$. Let $l = (m, n)$ be the greatest common divisor of m and n. Then $c' = l \cdot c$ is clearly the longest common yardstick of a and b. Corresponding to the Euclid algorithm of computing l from m and n, one has the following "*geometric algorithm*" of finding the longest common yardstick c' from a and b, namely:

Use the shorter one, say b, as the yardstick to measure the longer one, say a. If a is an integral multiple of b, then b itself is the longest common yardstick. Otherwise, one has a remainder r_1 shorter than b. Next use r_1 as the yardstick to measure b. If b is an integral multiple of r_1, then r_1 is the longest common yardstick. Otherwise, one has another remainder r_2 shorter than r_1. Keep going until the last remainder, say r_k, can integrally measure the preceding remainder r_{k-1} (i.e., without remainder). Then r_k is the longest common yardstick of a and b that we are seeking.

On the other hand, if one can *prove* that the above algorithm will never end for a specific, given pair of intervals, then such a pair is proven to be *non-commensurable*. This was exactly how Hippasus proved the non-commensurability of some specific pairs of intervals, the first one discovered by him is the following, namely

Example 1. Let a and b be the diagonal and the side of a regular pentagon. Then a and b are non-commensurable.

Proof. The proof is to show that the algorithm of alternating measurement, applying to such a pair $\{a, b\}$ will never end!

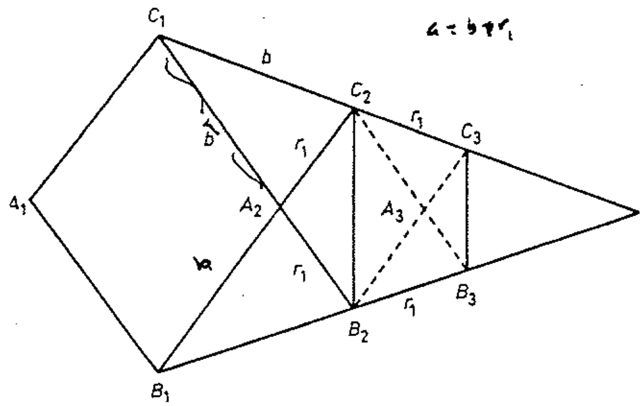


Fig. 3

As indicated in Fig. 3, $A_1B_1B_2C_2C_1$ is a regular pentagon whose side length and diagonal length are b and a respectively and its five inner angles are all equal to $\frac{3\pi}{5}$. $\Delta C_1B_2C_2$ is an isosceles triangle, thus

$$\angle C_1B_2C_2 = \angle B_2C_1C_2 = \frac{1}{2}\left(\pi - \frac{3\pi}{5}\right) = \frac{\pi}{5} \tag{9}$$

and the same reason shows that $\angle B_2C_2B_1 = \frac{\pi}{5}$. Therefore $\Delta A_2B_2C_2$ is also an isosceles triangle and $\angle B_2A_2C_2 = \pi - \frac{\pi}{5} - \frac{\pi}{5} = \frac{3\pi}{5}$. Moreover,

$$\angle C_1A_2C_2 = \pi - \frac{3\pi}{5} = \frac{2\pi}{5}, \ \angle C_1C_2A_2 = \frac{3\pi}{5} - \frac{\pi}{5} = \frac{2\pi}{5} \tag{10}$$

and hence $\Delta C_1A_2C_2$ is also an isosceles triangle.

Thus

$$a = \overline{C_1B_2} = \overline{C_1A_2} + \overline{A_2B_2} = b + r_1, \ r_1 = \overline{A_2B_2} = \overline{A_2C_2}. \tag{11}$$

Extend $\overline{B_1B_2}$ (resp. $\overline{C_1C_2}$) to B_3 (resp. C_3) such that $\overline{B_2B_3} = \overline{C_2C_3} = r_1$. Then, as indicated in Fig. 3, the pentagon $A_2B_2B_3C_3C_2$ is again a regular one! (The proof of this fact is a simple exercise.) Moreover, its diagonal length is b while its side length is r_1. Therefore, as one proceeds to measure b by r_1 as the yardstick, the geometric situation is exactly the same as before, namely, the remainder is the difference between the diagonal and the side of a regular pentagon. Thus

$$b = r_1 + r_2, r_1 = r_2 + r_3, \ldots, r_{k-1} = r_k + r_{k+1}, \ldots \tag{12}$$

where the pair $\{r_{k-1}, r_k\}$ are always the diagonal and the side of a regular pentagon! Of course, this algorithm can never end, although the size of the k-th regular pentagon gets smaller and smaller. This proves that $\{a, b\}$ are *non-commensurable*!

Example 2. After he discovered the above astonishing example of non-commensurable pair of intervals by the above simple ingenious proof, Hippasus naturally proceeded to analyze the commensurability problem between the diagonal and the side of a square, say a' and b'. As indicated by Fig. 4, it is not difficult to show that the algebraic relationships among the remainders of the algorithm of alternating measurements are as follows, namely

$$a' = b' + r_1, b' = 2r_1 + r_2, r_1 = 2r_2 + r_3, \ldots, r_{k-1} = 2r_k + r_{k+1}, \ldots \tag{13}$$

Therefore, the geometric situations from the second one onward are all the same and hence this algorithm can never end! Thus $\{a', b'\}$ is again a non-commensurable pair.

[We leave the geometric proof of (13) as an exercise.]

Historical remarks

(i) The above discovery of non-commensurable pairs by Hippasus is a monumental milestone in the entire human civilization of rational mind. However, to his fellow Pythagoreans and contemporary

10

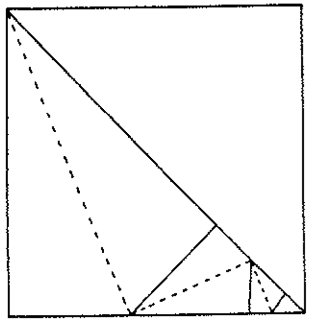


Fig. 4

geometers, this was a gigantic "*geoquake*" which rocked the whole foundation of quantitative geometry. The proofs of the area formula of rectangle and the similar triangle theorem that they prided were *no longer* complete proofs covering full generality, but rather, they were merely proofs for the *special commensurable case* only.

(ii) The historical record of this great event is unfortunately lost. However, according to some indirect sources, the following story may roughly serve as an account of what was happening to Hippasus and his great discovery: The initial reaction of his fellow Pythagoreans were shock and denial and, in order to avoid the unbearable embarrassment of public disgrace, they decided to cover it up and vowed to keep it as a secret. However, such a covering up of fundamental truth, eventually, became unbearable for the scholar Hippasus and he somehow leaked the truth of his great discovery to the outsiders (which, by the way, were often referred by the Pythagoreans simply as "the unworthies"). This made his fellow Pythagoreans furious and they condemned him to death! Naturally, he fled away. But unfortunately, the Pythagoreans were eventually able to track him down on a merchant ship in the Mediterranean and they pushed him overboard. Thus, a great hero of human civilization died for the truth. One might add here that the above story should probably be labelled as "*the first Pentagon Paper*".

(iii) Actually, his fellow Pythagoreans should be proud of suc a monumental discovery by their school. Moreover, although the first attempt in building a foundation of quantitative geometry wa not as perfect as they thought, it was still a major step forward an an impressive achievement by itself. Therefore, the proper reactio should be to celebrate the new discovery of their colleague, admittin the inadequacy of their proofs based upon the false axiom of univers validity of commensurability and then resolved to work for the proo of the remaining non-commensurable case. Of course, such proo were by no means easy to find and they naturally became the maj challenge to the entire community of Greek geometers of that tim The task of rebuilding a solid foundation of quantitative geomet was finally succeeded by Eudoxus (408–355 B.C.) and his successf story is naturally our next topic of discussion.

1.3. *Eudoxian principle, the origin of the methodology of approximation*

Let us begin with some analysis of the task that Eudoxus and h contemporary were facing.

Analysis

1. In the case that two intervals a and b are *commensurab* the ratio between their lengths has a clear simple meaning and it a rational number. However, in the case that two intervals a and are *non-commensurable* (such as the case of Examples 1 and 2), t meaning of the ratio between them is something yet to be defined a it is definitely *not* a rational number.

2. Although the meaning of the ratio between two non-comme surable pairs of intervals a and b is still undefined, the meaning inequality between such a yet to be defined ratio and a given ration number $\frac{m}{n}$ such as

$$a : b > \frac{m}{n} \text{ or } a : b < \frac{m}{n}$$

48

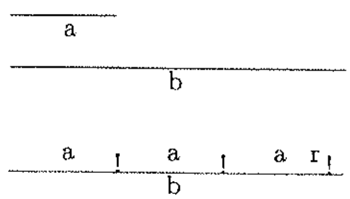
11 5/22

The Euclidean Algorithm, incommensurability and continued fractions

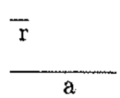
- Montgomery -

August 20, 2002

Given two line segments of lengths a and b we can use the shorter one as a measuring stick to measure the longer one. If $b = na$ with n an integer, then we stop. And $a/b = 1/n$, $b/a = n$. Otherwise there is a remainder r, with $0 < r < a$. We repeat the procedure, now using r as a measuring stick with which to measure a. If it fits exactly, then $a = n_1 r$ with n_1 a positive integer. And $a/b = n_1 r/(n(n_1 r + r) = 1/(n + (1/n_1))$. If not, we have a remainder r_2 and repeat the procedure.



We see that $b = 3a + r$, where the remainder r satisfies $0 < r < a$. Replace a by b and b by r. Repeat



12 ~~[illegible]~~

We will call this procedure the "Euclidean algorithm" as that is the name used for the same procedure applied to integers a and b. See below.

Definition 1 *The intervals a and b are commensurable if there is another interval c and positive integers p, q such that $a = pc$ and $b = qc$. In other words, p copies of c make up a and q copies of c make up b. The intervals a and b are incommensurable if there is no such interval.*

When we think of a and b as numbers rather than lengths, they are commensurable if their ratio a/b is rational and they are incommensurable if this ratio is irrational. Remember: the discovery of the existence of irrational ratios was a huge deal. We are told the first such was $\sqrt{2} = b/a$ where b is the hypotenuse of a right isosceles triangle whose equal sides are a .

Theorem 1 *The lengths a, b are commensurable if and only if the Euclidean algorithm eventually halts. In this case, the last remainder before the halt is the largest possible interval c for which $a = pc$ and $b = qc$ with p, q positive integers.*

Comparison with the Euclidean algorithm of Number theory.
The Euclidean algorithm is used in number theory to find the greatest common factor and to test for primality. Given two integers a, b we begin by testing which is greater. Say that $b > a$. Then we can write $b = n_1 a + r_1$ for integers, n_1, r_1. If b is a factor of a then $r_1 = 0$. Otherwise, $r_1 < a$. We repeat with the roles of a, b replaced by r_1, a. Thus $a = n_1 r_1 + r_2$. Example, $a = 32$, $b = 92$. $b = 2 \cdot 32 + 28$. So $r_1 = 28$. $a = 28 + 4$. So $r_2 = 4$. $28 = 7 \cdot 4$. The algorithm ends. $r = 4$ is the greatest common factor (or GCD) of 32 and 92.

The algorithm. We state the algorithm in full. Start with two "intervals", or lengths a and b. Suppose that $a < b$. In the following, $n_1, n_2, \ldots$ are positive integers. Find the largest integer n_0 such that n_0 intervals of length a fit within b. Thus:

$$b = n_0 a + r_1 \qquad ; 0 \leq r_1 < a.$$

If $r_1 = 0$ we are done: $b/a = n_0$. If not, we can write

$$a = n_1 r_1 + r_2 \qquad ; 0 \leq r_2 < r_1.$$

If $r_2 = 0$ we are done. If not, we can write:

$$r_1 = n_2 r_2 + r_3 \qquad ; 0 \leq r_3 < r_2.$$

If $r_3 = 0$ we are done. If not ... . The result at the $j + 2$nd step is:

$$r_j = n_{j+1} r_{j+1} + r_{j+2} \qquad ; 0 \leq r_{j+2} < r_{j+1}.$$

13 [illegible]

The Euclidean algorithm applied to the Golden Mean.

Let us see how this algorithm works on the Golden Mean. Recall that the defining property of the golden mean is that

$$\frac{a}{b} = \frac{b-a}{a}$$

with $0 < a < b$. Also recall that $a : b$ are the ratios of the sides of an isosceles triangles with angles $\alpha, 2\alpha, 2\alpha$ where $\alpha = \pi/5$. Thus:

$$b = a + r_1$$

Where $r_1 = b - a$. See figure! By cutting we see that (r_1, a) form the sides of another $\alpha, 2\alpha, 2\alpha$ isosceles triangle. Cut again:

$$a = r_1 + r_2$$

with $r_2 < r_1$ and

$$a/b = r_1/a = r_2/r_1.$$

The process continues ad infinitum!

$$r_2 = r_3 + r_4$$

$$\dots$$

$$r_j = r_{j+1} + r_{j+2}$$

with

$$a/b = r_1/a = r_2/r_1 = \dots = r_{j+1}/r_j = \dots.$$

14

Continued Fractions.

We use the output of the Euclidean algorithm above applied to a, b to express the real number a/b. We are assuming $0 < a < b$. The algorithm yields

$$b = n_0 a + r_1 \qquad ;0 \leq r_1 < a,$$

$$a = n_1 r_1 + r_2 \qquad ;0 \leq r_2 < r_1$$

$$r_1 = n_2 r_2 + r_3 \qquad ;0 \leq r_3 < r_2.$$

$$\ldots$$

$$r_j = n_{j+1} r_{j+1} + r_{j+2} \qquad ;0 \leq r_{j+2} < r_{j+1}.$$

$$\ldots$$

Set

$$s_1 = a/r_1$$

and

$$s_j = r_{j-1}/r_j; \qquad j = 2, 3, \ldots.$$

Then

$$s_j > 1$$

. Divide the equation for b by a, divide the equation for a by r_1, divide the equation for r_1 by r_2, and so on. We get:

$$b/a = n_0 + 1/s_1$$

$$s_1 = n_1 + 1/s_2$$

$$s_2 = n_2 + 1/s_3$$

$$\ldots$$

$$s_j = n_j + 1/s_{j+1}$$

with the n_j positive integers, and $0 < 1/s_j < 1$. Thus $a/b = 1/(n_0 + 1/s_1) = 1/(n_0 + (1/(n_1 + 1/s_2)) = 1/(n_0 + (1/(n_1 + 1/(n_2 + 1/s_2))) = \ldots$. Continuing, ad infinitum we get

$$a/b = \cfrac{1}{n_0 + \cfrac{1}{n_1 + \cfrac{1}{n_2 + \cfrac{1}{n_3 + \ldots}}}}$$

The Golden Mean.

All the n_j's were 1 when we applied the Euclidean algorithm to the Golden mean. It follows that

$$\frac{\sqrt{5}-1}{2} = \cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\cfrac{1}{1+\ldots}}}}$$

EXTRA MATERIAL. Analysis and number theory of continued fractions.

Halt the Euclidean algorithm at the jth step and drop the remainder. In this way we get the rational number

$$\alpha_j = \cfrac{1}{n_0+\cfrac{1}{n_1+\cfrac{1}{n_2+\cfrac{1}{n_3+\ldots+\cfrac{1}{n_j}}}}}$$

which ought to be a good approximation to our starting real number, the ratio $\alpha := a/b$.

Theorem 2 *The $\alpha_j \to \alpha = a/b$ as $j \to \infty$. Moreover, for α irrational:*

$$\alpha_0 < \alpha_2 < \ldots < \alpha < \ldots \alpha_3 < \alpha_1.$$

Definition 2 *Let α be a number. We say that the rational number $x = p/q$, p, q relatively prime integers is a best approximant to α if, for all other rationals $\tilde{x} = \tilde{p}/\tilde{q}$ whose denominator $\tilde{q}$ is less than or equal to that of x's, i.e for which $\tilde{q} \leq q$ we have*

$$|\alpha - x| \leq |\alpha - \tilde{x}|$$

Theorem 3 *The truncated continued fraction expansions α_j of α are the best rational approximants for α.*

References. Khinchin. Continued Fractions. Dover press.

of its transformations that do not use reflections) form a subgroup, and this subgroup is of a type we can identify by the methods we used before. The division of the two-dimensional and three-dimensional crystallographic groups of motions into classes can also be extended to groups containing reflections. For, just like screws with parallel axes and equal angles, two reflections in parallel planes or parallel lines can only differ by a translation. A summary of all the classes and groups obtained in this way is given in the following table.

	Plane		Space	
	Crystallographic Groups	Crystallographic Classes	Crystallographic Groups	Crystallographic Classes
Proper Rigid Motions	5	5	65	11
Added by Inclusion of Reflections . .	12	5	165	21
Total	17	10	230	32

Only by supplementing the proper motions with reflections do we get all the various crystal structures found in nature. In the construction of systems of pointers either in the plane or in space it is necessary in each case to add one extra pointer; for, in the plane a single pointer is invariant under reflection in the straight line containing the pointer; and similarly in space, the figure consisting of two unequal pointers is unchanged by a reflection in the plane of the pointers. In space, then, we have to use a point carrying three pointers of unequal length that do not all lie in one plane.

Instead of using geometric methods, we may also find the discontinuous groups of symmetry transformations by using algebraic methods. In the plane case, this leads to remarkable relationships among complex numbers; in space, the method is based on hypercomplex number systems.

It would be an interesting problem to generalize the present discussion to spaces of higher dimensionality. Some results relating to the discontinuous groups of symmetry transformations of higher-dimensional spheres have been found, the analogues of the regular polyhedra being known for spaces of any number of dimensions. We shall have more to say about these higher-dimensional figures in the next chapter. Moreover, Bieberbach has proved that there is only a finite number of n-dimensional crystallographic groups for every n, and that each of these groups contains n linearly independent translations.

§ 14. The Regular Polyhedra

The construction of the crystallographic classes led us to the regular tetrahedron and the regular octahedron. We shall now define regular polyhedra in general and find out what regular polyhedra are possible besides the tetrahedron and the octahedron.

We shall require that all the vertices of a regular polyhedron be equivalent and that the same be true for all the edges and all the faces. Furthermore, we require that all the faces be regular polygons.

First of all, a polyhedron satisfying these conditions cannot have any re-entrant vertices or edges. For, it is clear that the vertices cannot all be re-entrant, and therefore the presence of any re-entrant vertices would imply that not all the vertices are equivalent; the same argument holds for edges. It follows that the sum of the face angles at a vertex is always less than 2π. For otherwise all the faces meeting at the vertex would have to lie in one plane, or some of the edges ending at the vertex would have to be re-entrant. Furthermore, since at least three faces must meet at every vertex and the regularity conditions imply the equality of all the face angles, the magnitude of all of these angles must be less than $2\pi/3$. But the angle of the regular hexagon is exactly $2\pi/3$, and the angle of a regular n-sided polygon increases with n. Therefore the only polygons that can occur as faces of a regular polyhedron are the regular polygons having three, four, and five sides. Now the angles of the regular 4-sided polygon, i.e. the square, are right angles, so that no more than three squares can meet at a vertex without the sum of the angles at the vertex being equal to at least 2π; by the same token, more than three pentagons can certainly not meet at a vertex of a regular polyhedron. Now the shape of a regular polyhedron is completely determined by the number of faces meeting at a vertex and the number of sides of each polygon forming a face. There can be, accordingly, at most one regular polyhedron bounded by squares and one bounded by regular pentagons. On the other hand, three, four, or five equilateral triangles can meet at a vertex D, since it takes six of them to make the sum of the angles at the vertex equal to 2π. Therefore equilateral triangles can form the faces of three different regular polyhedra, bringing the total number of possible regular polyhedra up to five. Now all five of these possible forms actually exist.

16

41 57

They were well known as early as Plato, and he gave them a very important place in his Theory of Ideas, which is why they are often known as the "Platonic Solids." The most important data on the regular polyhedra are tabulated below. Figs. 95-99 show parallel projections of the regular polyhedra.

Name of the Polyhedron	Polygons Forming the Faces	Number of Vertices	Number of Edges	Number of Faces	Number of Faces Meeting at a Vertex
Tetrahedron (Fig. 95) . . .	Triangles	4	6	4	3
Octahedron (Fig. 96) . . .	"	6	12	8	4
Icosahedron (Fig. 97) . . .	"	12	30	20	5
Cube (Hexahedron) (Fig. 98)	Squares	8	12	6	6
Dodecahedron (Fig. 99) . .	Pentagons	20	30	12	3

All the regular polyhedra have a relation to the sphere much the same as that we have already described for the tetrahedron and octahedron in the last section. All of them can be inscribed in a sphere, and each of them generates a discontinuous group of motions of the sphere under which the vertices form a system of equivalent points. Now the planes that are tangent to the sphere at the vertices of such a polyhedron must bound another polyhedron which is also brought into self-coincidence by the motions of the group. It is to be expected that the new polyhedron is regular too; and thus the construction sets up a pairwise correspondence between the five polyhedra. If the construction is applied to the octahedron it does indeed lead to a regular polyhedron, namely, the cube; Fig. 100 illustrates the two polyhedra in the positions indicated. Thus the reduced group O could have been defined just as well by means of the cube as the octahedron. In the table, the relation between the two polyhedra is expressed by the fact that each has as many vertices as the other has faces, that both have the same number of edges, and finally, that the number of faces meeting at every vertex of either of them equals the number of vertices on each face of the other. Hence the octahedron can also be circumscribed about the cube (see Fig. 101).

The table shows that the dodecahedron and the icosahedron are related in the same way. Therefore both figures give rise to the same group, which is usually called the icosahedral group. Our

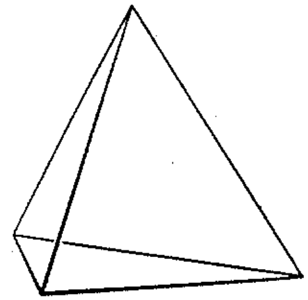


FIG. 95

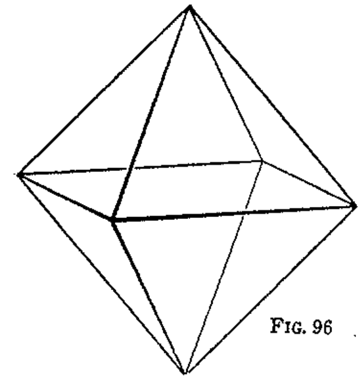


FIG. 96

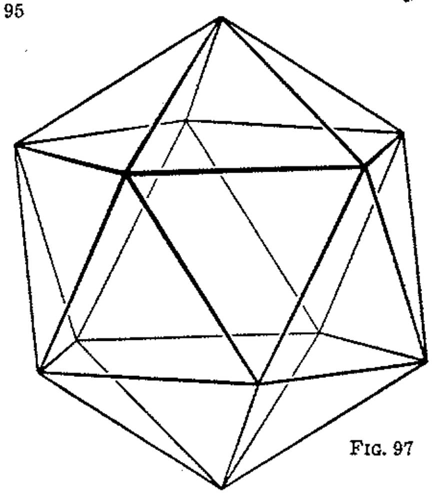


FIG. 97

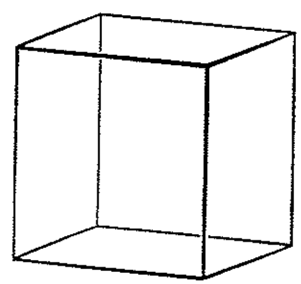


FIG. 98

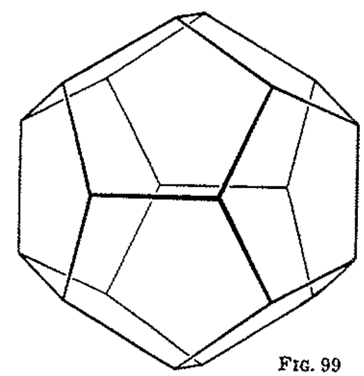


FIG. 99

18 ~~57~~

Complex numbers and geometry.

Punchline 1: The Euclidean plane IS the set of complex numbers.

Punchline 2: The complex arithmetical operations of addition and multiplication REALIZE the rigid motions (superpositions) of Euclid. Adding a fixed complex number translates the plane. Multiplying by a fixed complex number rotates the plane if the number has unit modulus, otherwise it rotates and dilates the plane.

The number i.

We suppose ourselves familiar with the real numbers and the number line. There is no real number whose square is -1. To start complex numbers we introduce or "invent" a new number called i which is to be thought of as orthogonal to the real number line, and whose square is -1: or $(i)^2 = -1$.

$$i = \sqrt{-1}.$$

As with real numbers, -1 has two square roots. The other one is $-i$, since $(-i)^2 = (-i)(-i) = (-1)^2(i)^2 = (1)(-1) = -1$.

We list the first few powers of i:

$$i, i^2 = -1, i^3 = -i, i^4 = 1, i^5 = i, \ldots$$

The pattern repeats with a period of 4. (This 4 is from 1/4 of a full circle. and the fact that $i = e^{i2\pi/4}$.)

COMPLEX ARITHMETIC.

A complex number is a number of the form $x + iy$ with x, y real. All the usual rules of arithmetic apply to complex numbers. Thus if $x + iy$ and $u + iv$ are two complex numbers then their product is

$$\begin{aligned}(x + iy)(u + iv) &= xu + x(iv) + iy(u) + i^2yv \\ &= xu + i(xv + yu) - yv \\ &= xu - yv + i(xv + yu)\end{aligned}$$

NOTATION and WARNING. We often use z to denote a complex number, or complex variable. So

$$z = x + iy.$$

This z is not to be confused with the z of a three-dimensional cartesian coordinate system. We write

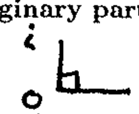
$$Re(z) = x$$

for the real part of z, and

$$Im(z) = y$$

for the imaginary part of z. Thus $z = Re(z) + iIm(z)$.

Equality of complex numbers. Suppose $w = u + iv$ is another complex number. To say that $z = w$ means that $x = u$ and $y = v$, or that $Re(z) = Re(w)$ and $Im(z) = Im(w)$. **An equation involving a complex numbers is two real equations. It says that both the real and imaginary parts of the right and left hand sides of the equation are equal.**



The set of COMPLEX NUMBERS IS EUCLID'S PLANE

Think of (x, y) as the usual Cartesian coordinates in the plane. Then (x, y) corresponds to $z = x + iy$. So 1 is the vector $(1, 0)$ and i is the vector $(0, 1)$. The word 'argand plane" is used for thinking of the cartesian xy plane as the space of all complex numbers.

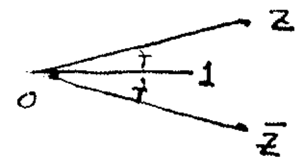
WARNING. There is a conceptual distinction between complex numbers and Euclid's plane. The complex number system comes with a distinguished point, namely 0, a distinguished line, the real line, and a distinguished length scale – the distance fromt he number 0 to the number 1. Euclid's plane has none of these: it has no special point – all points "look and act the same. It has no special directions. Any direction is as good as any other. And it has no distinguished ruler. Any choice of measurement is as good as any other.

It is conceptually more accurate to say that the complex number system is identical to the Cartesian plane: Euclid's plane with graph paper and an origin imposed.

Conjugation. The reflection of (x, y) about the x-axis is $(x, -y)$. We reserve a special name for this in complex variables.

Conjugation.

$$\bar{z} = x - iy$$



Formulae.

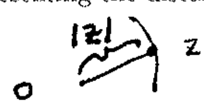
$$Re(z) = \frac{1}{2}(z + \bar{z})$$

$$Im(z) = \frac{1}{2i}(z - \bar{z})$$

$$z\bar{z} = x^2 + y^2.$$

We write $|z| = \sqrt{z\bar{z}}$. It is a positive real number (unless $z = 0$), representing the distance of the point z in the argand plane from the origin 0. We also write

$$r = |z|.$$



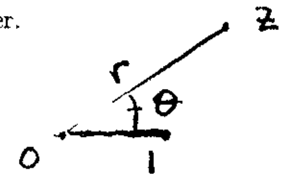
Inverting complex numbers. The same algebraic trick that enabled you to write $1/(1 - \sqrt{5})$ in the form $a + b\sqrt{5}$ with a, b rational, lets you divide by any complex number.

Here is the trick

$$\begin{aligned} \frac{1}{z} &= \frac{1}{z}\frac{\bar{z}}{\bar{z}} \\ &= \frac{\bar{z}}{z\bar{z}} \\ &= \frac{x}{r^2} - i\frac{y}{r^2} \end{aligned}$$

EXAMPLE: $1/(1 + i) = \frac{1}{2} - \frac{i}{2}$.

★ What is the 'geometry' behind inversion? I hope to delve into this question later.



Polar Coordinates. Euler's formula.

We have the very important formula:

$$e^{i\theta} = cos(\theta) + isin(\theta).$$

For θ a real number, representing an angle. From this formula we see that the polar representation of z is

$$\boxed{z = re^{i\theta}.}$$

which is to say

$$x = rcos(\theta), y = rsin(\theta).$$

There are two ways to see the validity of this formula for $e^{i\theta}$, recall the power series for e^x. It is

$$e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{3!}x^2 + \ldots \frac{1}{n!}x^n + \ldots$$

Now, replace x by $i\theta$, and equate real and imaginary parts. Note from the powers of i that the real terms correspond to all the even powers, and the odd terms correspond to all the odd powers. Thus

$$e^{i\theta} = (1 - \frac{1}{2!}\theta^2 + \frac{1}{4!}\theta^4 + \ldots) + i(\theta\frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 - \ldots).$$

20

489

The real part is the Taylor expansion of $cos(\theta)$, and the imaginary part is the Taylor expansion of $sin(\theta)$.

Let us check consistency. For real exponents we have $e^a e^b = e^{a+b}$. So we should have $e^{i\theta_1}e^{\theta_2} = e^{i(\theta_1+\theta_2}$. But this says that

$$(cos(\theta_1) + isin(\theta_1)(cos(\theta_2) + isin(\theta_2) = (cos(\theta_1 + \theta_2) + isin(\theta_1 + \theta_2).$$

Multiplying out the left hand side we find:

$$(cos(\theta_1) + isin(\theta_1)(cos(\theta_2) + isin(\theta_2) = [cos(\theta_1)cos(\theta_2) - sin(\theta_1)sin(\theta_2)] + i[cos(\theta_1)sin(\theta_2) + cos(\theta_2)sin(\theta_1)].$$

But the angle addition formulas from trigonometry assert that $cos(\theta_1+\theta_2) = cos(\theta_1)cos(\theta_2) - sin(\theta_1)sin(\theta_2)$ and that $sin(\theta_1 + \theta_2) = cos(\theta_1)sin(\theta_2) + cos(\theta_2)sin(\theta_1)$.

Some examples:

$$1 = e^{i2\pi} = e^{i0}.$$

$$-1 = e^{i\pi}.$$

$i = e^{i\pi/2}$.

$$-i = e^{i\pi/2} = e^{i3\pi/2}$$

This is consistent with the powers of i at the beginning. Thus: $i^2 = -1$ corresponds to $2\pi/2 = \pi$, i.e. $(e^{i\pi/2})^2 = e^{2(i\pi/2)} = e^{i\pi}$. And $i^4 = 1$ to $4(\pi/2) = 2\pi$.

To see the power of polar coordinates, consider: **Exercise.** Compute $(1 + i)^{100}$.

Solution. $1 + i = \sqrt{2}e^{i\pi/4}$. So $(1 + i)^{100} = (\sqrt{2})^{100}(e^{i\pi/4})^{100} = 2^50e^{25\pi} = -2^{50}$.

Compare this with what happens when you try to do the computation directly.

Unit complex numbers. The set of unit complex numbers is, by definition, the set of all z in the argand plane with $|z| = 1$. In cartesian coordinates this says that $x^2 + y^2 = 1$. So the set of unit complex numbers is the same as the circle. In polar coordinates, a number with $|z| = 1$ is saying that $r = 1$, θ is arbitrary. So any complex number z with $|z| = 1$ can be written $z = e^{i\theta}$.

Transformations of the plane via complex arithmetic.

First, the number line.

A translation of the number line is implemented by adding: $x \mapsto x + 1$ shifts all points to the right by one unit and thereby preserves distances. Scalings, or stretchings are represented by multiplication. $x \mapsto 2x$ doubles the size of any interval. Composing the two kinds of operations we get $x \mapsto ax + b$, which dilates by the factor a and translates by b. Let us see what it does to distances:

$$\begin{aligned}|F(s) - F(t)| &= |(as + b) - (at + b)|\\ &= |as - at|\\ &= |a||s - t|.\end{aligned}$$

Thus this scales distances by a factor $|a|$. The only time we get an isometry of the line is when $|a| = 1$, which is to say $a = \pm 1$. The case $a = -1$, with $b = 0$ corresponds to a reflection through the origin.

Transformations of the complex plane. Now we move to the argand plane, which is the Cartesian plane. First, note that the distance between two points $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$ is $|z_1 - z_2|$. A map of the complex plane to itself will can be written $z \mapsto F(z)$ where F is a complex function of the complex variable z.

Translations of the complex plane. These are implemented by $z \mapsto z+b$, with b a complex constant. This b is the translation vector. It tells us in what direction we are moving, and by how much.

Examples.

The case $b = 1$ corresponds to translating in the x-direction 1 unit.

The case $b = i$ corresponds to translating in the y-direction 1 unit.

The case $b = 1 + 3i$ corresponds to translating in the x-direction 1 unit, and then the y direction 3 units. Since $1 + 3i = 2e^{i\pi/3}$, this is the same as translating in the direction of 60 degrees towards 'north' by 2 units.

21 180

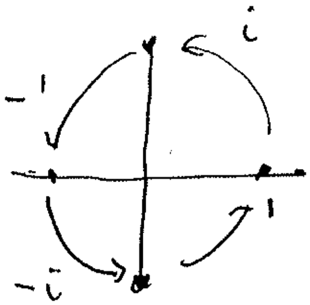
Rotations. A rotation by θ_0 radians about the origin $z = 0$ is represented by a map of the form $z \mapsto az$ where $a = e^{i\theta_0}$.

Example. Rotation by π: We have $e^{i\pi} = -1$, so $z \mapsto -z$ is rotation by π.

Rotation by $\pi/2$: We have $i = e^{i\pi/2}$ so $z \mapsto iz$ represents rotation by 90 degrees, counterclockwise.

Check: $1 \mapsto i$, $i \mapsto -1$, $-1 \mapsto -i$ and $-i \mapsto 1$ under this map.

INSERT FIGURE



multiplication by i rotates by $90°$.

EXERCISE: Express rotation by $\pi/4$ radians counterclockwise in terms of complex multiplication.

SOLUTION. We have $e^{i\pi/4} = 1/\sqrt{2} + i/sqrt2$, so the map is $z \mapsto (1/\sqrt{2} + i/sqrt2)z$

General affine transformations of z. An affine transformation is, by definition, one of the form (linear) plus (translation). So it is of the form $z \mapsto az + b$ with a, b complex constants. Write $z' = az + b$. Let z, w be two points in the argand plane, and write z', w' for their image points. We want to relate the distance between these image points and that of the original points.

$$|z' - w'| = |(az + b) - (aw + b)| = |a(z - w)|.$$

Now we use the following fact from complex numbers:

FACT about modulus. $|z_1 z_2| = |z_1||z_2|$ for any two complex numbers z_1, z_2. PROOF OF FACT: $|z_1 z_2|^2 = z_1 z_2 \bar{z}_1 \bar{z}_2 = z_1 \bar{z}_1 z_2 \bar{z}_2 = |z_1|^2 |z_2|^2$.

Continuing...

$$|z' - w'| == |a||z - w|.$$

Since, the distance between z and w is $|z - w|$, and since the points z and w were arbitrary, we have that $z \mapsto az + b$ is an isometry if and only if $|a| = 1$.

Interpretation. Write $a = e^{i\theta_0}$ Then $z \mapsto az + b$ is rotation by θ_0 radians about the origin , followed by translation with translation vector b.

Scaling. $z \mapsto 2z$ scales everying up by the factor 2. $z \mapsto \frac{1}{2}z$ shrinks every figure by the factor 1/2. Then $z \mapsto az$, with $a = r_0 e^{i\theta_0}$ scales by the factor r_0 and rotates by θ_0.

FUNDAMENTAL THEOREM OF ALGEBRA.

This theorem was Gauss's thesis.

In its most basic form, the theorem asserts that every polynomial has a root in the complex numbers. Equivalently, it asserts that any complex polynomial of degree n has n roots when the roots are counted with their multiplicity.

This count of roots over the real numbers is FALSE. For example the polynomial $x^2 + 1$ has no real roots. It does have 2 complex roots, namely i and $-i$. The typical complex polynomial of degree n has exactly n roots. There is a notion of the *multiplicity of a root*: and counted with multiplicity every polynomial has exactly n roots.

22

88

Box 4: Complex number arithmetic

Complex numbers are manipulated according to the ordinary rules of arithmetic, always remembering that $i^2 = -1$. For example,

$$(3+4i)+(2-3i)=(3+2)+(4i-3i)=5+i$$

and

$$(3+4i)(2-3i)=6-9i+8i-12i^2=6-i+12=18-i$$

and

$$\frac{1}{2+i}=\frac{2-i}{2^2+1^2}=\frac{2}{5}-\frac{i}{5}.$$

The formal rules are:

$$\begin{aligned}
(a+ib)+(x+iy)&=(a+x)+i(b+y)\\
(a+ib)-(x+iy)&=(a-x)+i(b-y)\\
(a+ib)\times(x+iy)&=(ax-by)+i(ay+bx)\\
1\div(x+iy)&=\frac{x}{x^2+y^2}-i\,\frac{y}{x^2+y^2}\\
(a+ib)\div(x+iy)&=\frac{ax+by}{x^2+y^2}+i\,\frac{bx-ay}{x^2+y^2}.
\end{aligned}$$

we have done. Ordinary numbers have immediate connection to the world around us; they are used to count and measure every sort of thing. Adding, subtracting, multiplying and dividing all have simple interpretations in terms of the objects being counted and measured. When we pass to complex numbers, though, the arithmetic takes on a life of its own. Since -1 has no square root, we decided to create a new number game which supplies the missing piece. By adding in just this one new element $\sqrt{-1}$, we created a whole new world in which everything arithmetical, miraculously, works out just fine.

The world of complex numbers leads to elegant and profound connections between geometry, algebra and analysis. In fact, much of their importance stems from the fact that they give us a simple method of representing points in the plane. The germ of this idea is to be found in John Wallis' 1685 *Treatise on Algebra*, but was not fully and clearly developed until the work of the Norwegian surveyor Caspar Wessel (1745–1818)[1], and did not gain general acceptance until the work of Gauss (1777–1830)[2]. Let's go back to our description of points in front of us by positive numbers and behind as negative. The left frame of Figure 2.1 shows Dr. Stickler standing on an infinite flat plain facing

[1] A translation of Wessel's paper, with much interesting background about the history of complex numbers, is to be found in bicentennial commemorative volume *Caspar Wessel* by B. Branner and J. Lützen, published by the Royal Danish Academy of Sciences and Letters, C. A. Reitzel, 1999.

[2] See Gauss' *Collected Works*, Vol. 2, pp. 174–178. Although Gauss claimed that he already understood the complex plane in his first proof of the Fundamental Theorem of Algebra (see Note 2.3) in 1799, he did not actually publish anything on the subject until 1831.

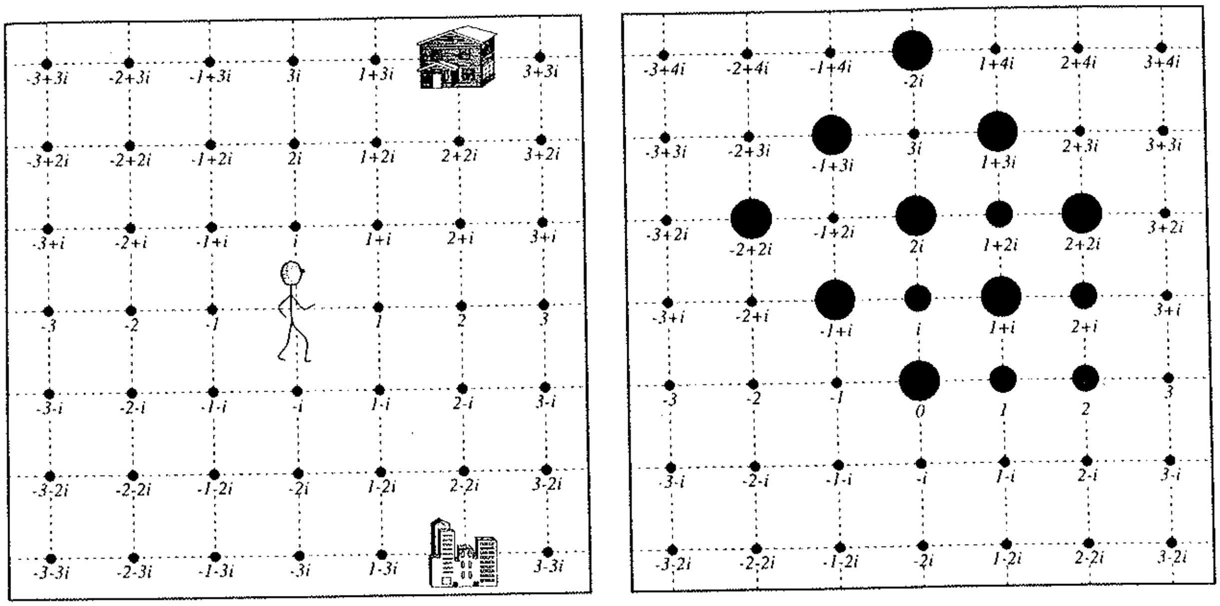


Figure 2.1. Left: Dr. Stickler stands on an infinite complex plain [*sic*], on which locations are marked by dots with complex numbers as addresses. Thus, his house is at $2 + 3i$, while his office is at $2 - 3i$. (You were expecting somewhat more austere accommodations?) Right: The nine red points are the products of the 9 blue points by $1 + i$. See how the grid is rotated and expanded by this operation.

along the horizontal x-axis, so that points directly ahead of him are marked out by positive numbers and points behind him with negative ones. Here's the idea: he can also describe points on the line stretching out at right angles to his left (the positive y-axis) by imaginary numbers $i, 2i, 3i, \ldots$; and points on the line to his right (the negative y-axis) by $-i, -2i, -3i, \ldots$. Even better, he can describe points ahead and to his left as complex numbers like $2 + 3i$, points ahead and to the right as $2 - 3i$, and so on. What this comes down to is this: the complex number $x + iy$ can be thought of as representing the point on the plane with coordinates (x, y). Dr. Stickler is standing at the origin $(0, 0)$. When we make this association, the coordinate plane is known as the **complex plane**.[1] The x-axis is now called the **real axis** because it represents those complex numbers $x + 0i$ which are actually ordinary real numbers, and the y-axis is called the **imaginary axis** because it represents all the numbers $0 + yi$ which are purely imaginary. In the complex plane, addition of complex numbers is the same as vector addition $(a, b) + (x, y) = (a + x, b + y)$. As you can see, this is just the same as complex addition $(a + ib) + (x + iy) = (a + x) + i(b + y)$.

[1] The complex plane is sometimes called the **Argand diagram**, after the rather obscure French mathematician Argand whose work on the geometry of complex numbers, originally published anonymously in 1806, also played a significant role.

Vector addition and subtraction is useful, but the beauty and power of complex numbers stems from the fact that the other algebraic operations of multiplication and division have an important geometrical interpretation as well. We could go on and just show you the formulas and figures, but we want you to read this book more interactively!

24

Here's a little experiment. If you haven't worked with complex numbers before, please try it to flex your mental muscles. In the complex plane, the nine complex numbers 0, 1, 2, i, $i+1$, $i+2$, $2i$, $2i+1$, $2i+2$ form a little 3×3 square with its bottom left corner at the origin. Now multiply all these nine numbers by $(1+i)$ and plot your answers. This is easy, for instance $(1+i) \times (2i+1) = -1+3i$. The effect is startling: as you can see in the right frame of Figure 2.1, the resulting red dots again form a 3×3 square but compared to the original square of blue dots it is larger and turned on its side. In fact, the square has expanded by a factor of $\sqrt{2}$ and has rotated anti-clockwise through 45°.

This simple experiment shows the general story: when complex numbers are thought of as points in the plane, multiplying by the complex numbers $x+iy$ has two effects: it stretches (or shrinks) everything by some factor r and rotates the plane around the origin through some angle θ. To understand this in more detail, we first need to look at the relationship between complex numbers and polar coordinates.

The **polar coordinates** of a point P in the plane are (r, θ), where r is the distance of P from the origin O, and θ is the angle the line joining P to O makes with the x-axis, measured anticlockwise. (Angles measured in the clockwise direction are counted as negative.) As you can see in Figure 2.2, if P has coordinates (x, y), then from Pythagoras' theorem it follows that $r = \sqrt{x^2+y^2}$ and from trigonometry we get $\tan\theta = y/x$. Thus θ is given by the inverse of the tangent function applied to y/x, the **arctangent** of y/x.

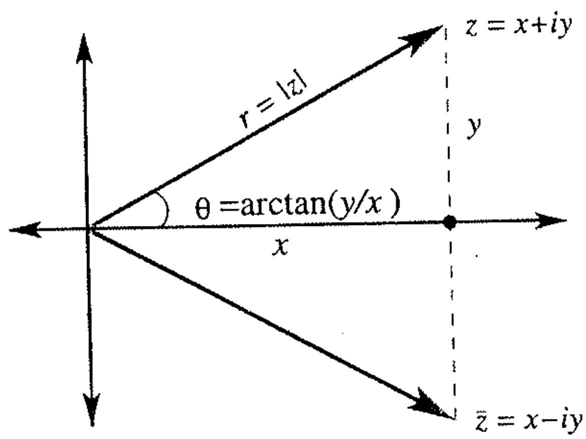


Figure 2.2. A complex number z can be described either via its real and imaginary parts x and y or via its modulus r and argument θ. The complex conjugate $\bar{z} = x - iy$ is obtained by reflecting in the real axis.

Now we can apply this to a complex number $z = x+iy$, thinking of z as a point $P = (x, y)$ in the plane. The distance r is called the **modulus** or **absolute value** of z, written $|z|$ (pronounced 'mod z') and the angle θ is called the **argument** of z, written $\arg z$. We should probably mention that, although angles are measured in degrees by ordinary people, mathematicians and scientists primarily use **radians**. The radian measure is based on the circumference being 2π times the radius of the circle, where π is the famous and mysterious number 3.14159... . A full circle of 360° measures 2π radians, an angle of 180° (or a half circle) is π radians, and 90° is $\pi/2$ radians, so one radian is about 57°.

Suppose now that we are given the modulus $r = |z|$ and argument $\theta = \arg z$ of a complex number z. How can we recover the real and imaginary parts x and y so as to be able to write $z = x + iy$? Using trigonometry in the right-angled triangle in the previous figure, we see that $x = |z|\cos\theta$ and $y = |z|\sin\theta$, so that $z = |z|(\cos\theta + i\sin\theta)$. Thus we can go back and forth between writing a complex number using its real and imaginary parts, or using its modulus and argument. The form $z = |z|(\cos\theta + i\sin\theta)$ is called the **polar representation** of the z.

Using the polar representation, the geometry of complex multiplication looks easy: suppose z and w are two complex numbers. Then zw is that complex number whose modulus is the product of the moduli of z and w and whose argument is the sum of their arguments. In symbols:

$$|zw| = |z| \times |w|,$$
$$\arg(zw) = \arg z + \arg w.$$

These formulas can be proved using high school trigonometry as we have done in Note 2.2. They are deceptively simple: it is a remarkable fact that to multiply two complex numbers you just *multiply* the moduli and *add* the arguments. This idea is fundamental to manipulating complex numbers and, as we shall see in the next section, it imparts a geometrical significance to all the arithmetical operations we shall want to perform.

Using the polar form, we get a nicer expression for the complex number $1/z$. Suppose we write $z = |z|(\cos\theta + i\sin\theta)$ and we set $w = (1/|z|)(\cos(-\theta) + i\sin(-\theta))$. Then from our formula, zw has modulus $|z||w| = |z| \times 1/|z| = 1$ and argument $\arg z + \arg w = \theta + (-\theta) = 0$. We immediately conclude that $w = 1/z$, because our reasoning shows that w must be the number such that $z \times w = 1$.

Complex numbers with modulus one play a rather special role. If $z = x + iy$ has modulus 1, in symbols if $|z| = 1$, then by definition $x^2 + y^2 = 1^2 = 1$. This means that z lies on the circle of radius 1 with centre the origin, usually known as the **unit circle**. Using the polar representation, any such point z can be written in the form $z = \cos\theta + i\sin\theta$ with $\theta = \arg z$. The angle θ of course is the angle θ shown in Figure 2.2. For example, $i = \cos 90° + i\sin 90°$. Again if

26
84

$z = \frac{1+i}{\sqrt{2}}$ then $\left|\frac{1+i}{\sqrt{2}}\right| = \sqrt{1/2 + 1/2} = 1$ and $\frac{1+i}{\sqrt{2}} = \cos 45° + i \sin 45°$, which is the point on the unit circle making the angle 45° (or $\pi/4$ radians) with the x axis. From the multiplication formulas you can see that the effect of multiplying two points z and w on the unit circle is to produce another point zw on the unit circle whose argument is exactly the sum of the arguments of z and w!

When we wanted to divide a complex number by $x + iy$, its 'twin' or complex conjugate $x - iy$ came to the rescue. The complex conjugate also has a nice geometrical interpretation: you can see from Figure 2.2 that it is just the reflection of $x + iy$ in the real axis. From Dr. Stickler's perspective in Figure 2.1, complex conjugation interchanges points on his left with those on his right and vice versa. If we write $z = x + iy$, its conjugate is written $\bar{z}$. The operation of passing from z to $\bar{z}$ is called **complex conjugation.**[1] The complex conjugate relates to the modulus like this:

$$z\bar{z} = (x + iy)(x - iy) = x^2 + y^2 = |z|^2,$$

which explains the trick we used in the last section when we wanted to divide.

[1] Not to be confused with the other use of the word conjugation for changing maps which we discussed in the last chapter!

Complex conjugation has an equally nice interpretation in terms of the polar representation. Notice that z and its conjugate have the same modulus: $|z| = \sqrt{x^2 + y^2} = |\bar{z}|$. If $z = |z|(\cos\theta + i \sin\theta)$, then as we have seen, θ is defined by the equation $\tan\theta = y/x$. Replacing y by $-y$ we find that $\arg\bar{z} = -y/x = -\theta$. Thus $\bar{z} = |z|(\cos(-\theta) + i \sin(-\theta))$, confirming that $\bar{z}$ is the reflection of z in the horizontal x-axis. We have summarized these various handy formulas in Box 5.

You might ask if we can take the principle which led to complex numbers further. For example, can we find a complex number $z = x + iy$ such that $z^2 = (x + iy)^2 = i$? In other words, can we find a complex number which is the square root of the number i, or do we have to invent yet another sort of 'super-complex' number? It turns out we can always find square roots using the polar representation $z = |z|(\cos(\theta) + i \sin(\theta))$.

Note 2.2: **Multiplication in polar coordinates**

Suppose $z = r(\cos\theta + i \sin\theta)$ and $w = s(\cos\phi + i \sin\phi)$. Then using the trigonometric identities for the sine and cosine of the sum of 2 angles, we get:

$$\begin{aligned} zw &= rs(\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi) \\ &= rs\,[(\cos\theta\cos\phi - \sin\theta\sin\phi) \\ &\quad + i(\cos\theta\sin\phi + \sin\theta\cos\phi)] \\ &= rs(\cos(\theta+\phi) + i\sin(\theta+\phi)). \end{aligned}$$

Box 5: **Complex numbers and polar coordinates**
Here are the main formulas used in the polar representation of complex numbers:

- The complex conjugate of $z = x + iy$ is $\bar{z} = x - iy$.
- The modulus or absolute value of $z = x + iy$ is the number $|z|$ found from the formula:
$$z\bar{z} = (x + iy)(x - iy) = x^2 + y^2 = |z|^2.$$
- The polar representation of the complex number $z = x + iy$ is
$$z = |z|(\cos\theta + i\sin\theta)$$
where $\theta = \arg z = \arctan y/x$.
- In polar form, the complex conjugate of $z = |z|(\cos\theta + i\sin\theta)$ is $\bar{z} = |z|(\cos(-\theta) + i\sin(-\theta))$.
- In polar form, complex numbers are multiplied by the rule:
$$|zw| = |z| \times |w|,$$
$$\arg(zw) = \arg(z) + \arg(w).$$
- If $z = |z|(\cos\theta + i\sin\theta)$ then $1/z = (1/|z|)(\cos(-\theta) + i\sin(-\theta))$.

From the formulas in Box 5:

$$\left(\pm\sqrt{|z|}\left(\cos(\theta/2) + i\sin(\theta/2)\right)\right)^2 = |z|(\cos\theta + i\sin\theta) = z$$

and so the two complex numbers $\pm\sqrt{|z|}(\cos(\theta/2) + i\sin(\theta/2))$ are the two square roots of z. For example, suppose we want to find the two square roots of i. In polar form, $i = \cos\pi/2 + i\sin\pi/2$ and $|i| = 1^2 = 1$. Thus our formula gives

$$\sqrt{i} = \pm(\cos\pi/4 + i\sin\pi/4)$$

and as one can calculate using Pythagoras' theorem, $\cos\pi/4 = \sin\pi/4 = 1/\sqrt{2}$. Thus the square roots of i are $\pm(1 + i)/\sqrt{2}$.

Similar ideas allow you to find any root of any complex number, for example you might like to try finding the *three* cube roots of i. In fact, as explained in Note 2.3, it is a famous result that there is no need to introduce any 'super-complex' numbers to solve any polynomial equation created with any complex number coefficients at all.

28

Computing with complex numbers

From now on, almost every computation we make will use complex numbers. To do anything, therefore, we must set up a method for programming them. As is inevitable with computers, this is easy in some setups and harder in others. In the high level mathematics packages, in FORTRAN and in some releases of C++, complex number data types and complex arithmetic are built in. In most other environments, you will have to write your own library of complex number routines. An example for the language C can be found in the very useful book *Numerical Recipes in C*.[1]

Writing a complex arithmetic library works differently in traditional and object-oriented languages. In the traditional language C, for example, you first define a new data type `complex` (called a 'structure') so that a complex number `z` is an array of two real numbers. In some languages, the real and imaginary parts of a complex data structure `z` are obtained by functions such as `re(z)` and `im(z)` (or `real(z)` and `imag(z)`; many variations of the names occur). In C, the definition of the `complex` structure associates names to the real and imaginary parts, as in:

```
typedef struct double re, im; complex;
```

With this convention, the real and imaginary parts of a complex structure `z` are defined as double precision real numbers, and they are henceforth referenced as `z.re` and `z.im`.

[1] W. Press, S. Teukolsky, W. Vetterling and B. Flannery, *Numerical Recipes in C, The Art of Scientific Computing*, 2nd edition, Cambridge University Press, 1992.

Note 2.3: **Solving polynomial equations**

A **polynomial equation of degree** n is an equation of the form $ax^n + bx^{n-1} + cx^{n-2} + \cdots + fx + g = 0$, where n is a positive integer and $a, b, c, \ldots, f, g$ are any numbers, real or complex. For example, $x^3 + 2x - 5 = 0$ is a polynomial equation of degree 3, otherwise known as a **cubic equation**. A special case is the quadratic equation $ax^2 + bx + c = 0$. As we have seen, in complex arithmetic we can always find square roots even of negative or complex numbers. So the formula $x = (-b \pm \sqrt{b^2 - 4ac})/(2a)$ always yields a solution or **root** of this equation. The *Fundamental theorem of algebra* proved by d'Alembert in 1746 (with some gaps filled by Argand in 1814) says that a polynomial of degree n with complex coefficients always has a solution which is a complex number, and indeed that all possible solutions are either real or complex. For example the polynomial equation $x^5 + 4i\,x^4 - 9x^3 - 12i\,x^2 + 9x - 2 + 4i = 0$ has a complex root. (Polynomials like this will resurface in Chapter 9.) In other words, if you start in the complex numbers, you never have to go outside to find solutions to any polynomial equations: the complex numbers are, algebraically speaking, a closed system.

1

29 #23

Complex multiplication and mapping the plane

In Chapter 1 we discussed in detail how a transformation T of the plane can be viewed as a rule which assigns to each point P in the plane a new point Q. The operation of multiplication by a fixed complex number provides us with just such a rule. Fix a complex number a. Remembering that each point P can be represented by a complex number z, our rule can be written very simply:

$$T(z) = az.$$

The effect of such a T may be guessed at from the results of the experiment we carried out in Figure 2.1. Since maps like this are going to be very important to us, let's take as a first simple example the value $a = 2$. Applying T to the complex number $z = x + iy$ we find $T(z) = 2z = 2x + 2iy$, in other words the point with coordinates (x, y) maps to the point with coordinates $(2x, 2y)$. Now suppose that F is some shape in the plane. If we multiply the coordinates of all the points in F by 2, we obtain a similar figure $T(F)$ in which all the linear dimensions have been expanded by the factor 2. In fact T is a transformation of the plane which expands out from the origin O by a factor of 2 in all directions. Multiplying by $a = 1/2$ gives the inverse of T: the map $z \mapsto z/2$ pulls every point directly towards the origin, halving the distance from O and shrinking every object by half. You can see the effect of multiplication when $a = 1/2$ and $a = 3/2$ on the red fox in the top right quadrant of the left frame of Figure 2.3.

What would it mean to multiply by a negative number? If we try with -1, then the point $x + iy$ is transformed into the point $-x - iy$, otherwise said, (x, y) moves to $(-x, -y)$. This is the transformation which moves every point P across the origin O to the point Q on the opposite side at the same distance, in other words, reflection across the origin. We can get multiplication by -2 in two steps: first transform $x+iy$ into $2x+2iy$, which causes expansion in all directions by a factor of 2, and then multiply by -1 which flips across the origin. This is what has happened to the foxes in the bottom left quadrant on the left. Notice that we could first shrink and then flip, or first flip and then shrink, getting the same answer either way. In mathematical language, multiplication is **commutative**. When we start combining different kinds of operations, for example translations and expansions, they may no longer commute.

So far, we have not done anything which required complex numbers; all we have said could have been done with vectors, and in more than two dimensions. The really interesting step is to multiply by a number

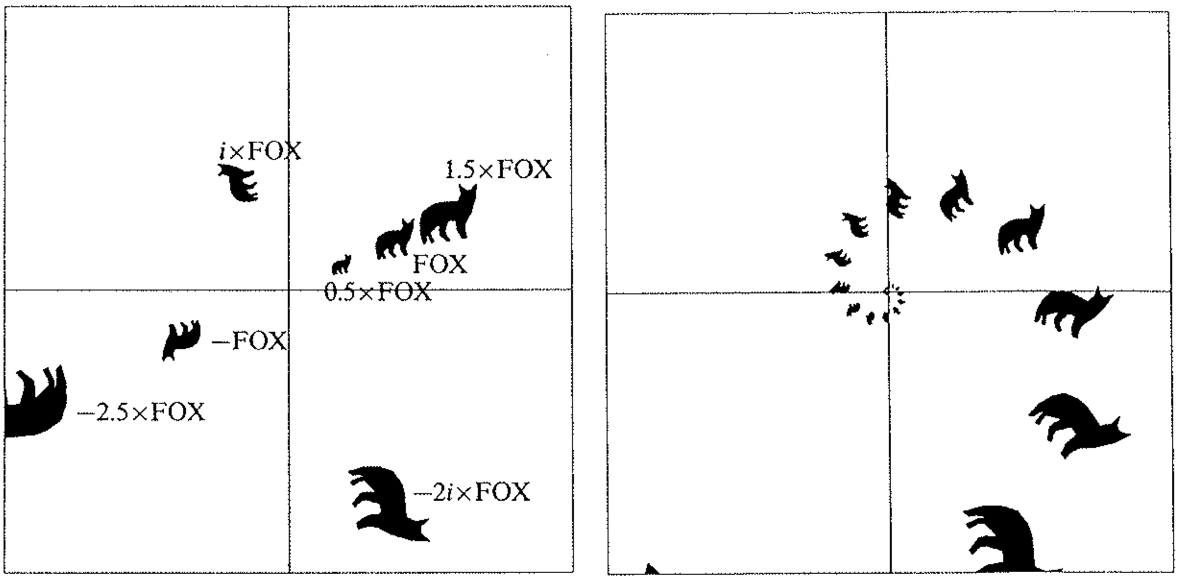


Figure 2.3. Left: The original red fox in the upper right quadrant is transported by the transformations $z \mapsto 1.5z$, $z/2$, $-z$, $-2.5z$, iz and $-2iz$. Right: The fox is transported by the transformations $z \mapsto \ldots, a^{-2}z, a^{-1}z, az, a^{1}z, a^{2}z, \ldots$ and so on, where $a = 0.8 \times \left(\frac{\sqrt{3}}{2} + \frac{1}{2}i\right)$. Positive powers of a make him spiral into the origin; negative powers make him spiral away.

which is truly complex. Just as we investigated the effect of multiplying by -2 in two stages, first a stretch and then a flip, so we can split up the question of understanding multiplication by a into two parts. Write a in its polar form $a = A(\cos\theta + i\sin\theta)$, where $A = |a|, \theta = \arg a$. Since A is real and positive, we already know that multiplication by A is expansion or contraction by this factor. How about multiplication by a complex number $\cos\theta + i\sin\theta$ of modulus one? From the formulas in Box 5, we see that for any z, the product $(\cos\theta + i\sin\theta)z$ has the same modulus as z, while its argument is $\theta + \arg z$. Thus $(\cos\theta + i\sin\theta)z$ is the point you get by rotating z about the origin anticlockwise through the angle θ. The combined operations of expansion (or contraction) and rotation cause spiralling in or out from O, as on the right in Figure 2.3.

This explains what happened in our experiment in Figure 2.1 when we multiplied the 3×3 grid of blue dots by $1 + i$. In polar coordinates, $1 + i = \sqrt{2}(\cos 45° + i\sin 45°)$, so the modulus of $1 + i$ is $\sqrt{2}$ and its argument is 45°, confirming our conclusion that the grid turned through 45° and expanded by $\sqrt{2}$.

In the right frame of Figure 2.3 the multiplication factor is $a = \frac{4}{5}(\cos 30° + i\sin 30°)$. Denoting the fox by F, we have drawn his orbit F, aF, a^2F, a^3F, a^4F and so on. These image foxes make angles $0, 30°, 60°, 90°, 120°$ with the real axis and are at distances 1, 0.8, 0.64, 0.512, 0.4096 from the origin. The orbit is spiralling into the origin, and the angle $\theta = 30°$ determines the rate of turning. If θ were negative, the spiral would turn backwards. The factor A, in this case $4/5$, is the rate of contraction. If A were greater than one, the foxes would spiral away. This is exactly what happens if we *divide* instead of multiply by a. Dividing by powers of a is the same as multiplying by powers of $1/a$; thus $T^{-1} : z \mapsto z/a$ is the inverse of T. Applying T^{-1} causes the fox to spiral backwards, away from the origin, which you can also see in the figure.

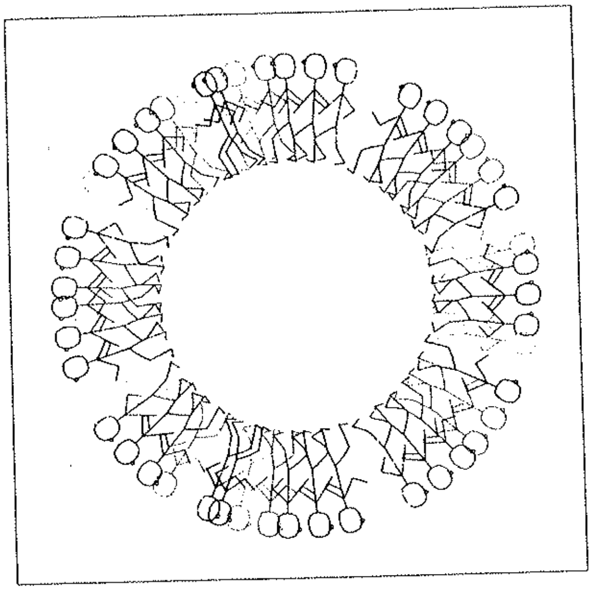


Figure 2.4. Dr. Stickler in the centrifuge. Dr. Stickler whirls around at angles equal to an irrational multiple of π, in this case $\pi\frac{\sqrt{5}-1}{2}$ radians. Each time he rotates by this angle, he changes colour. (Wouldn't you?) Eventually all the copies of Dr. Stickler would fill out a blurry ring. This was difficult for Dr. Stickler to endure, but after all it was in the interests of science.

Multiplication by powers of points on the unit circle gives an interesting exception. If $|a| = 1$, then there is no expansion or contraction, only rotation through an angle θ. All the points in the orbit of an initial point z_0 remain on the circle of radius $|z_0|$ and centre O, moving round and round with equally spaced jumps. There are two possibilities: either some orbit point eventually lands back on z_0, or not. In the first case, we start retracing our steps over the same path, as happened for example in Figure 1.6 (with $\theta = 72°$). In the second case, shown in Figure 2.4, orbit points (or rather Dr. Sticklers) keep piling up more and more thickly all the way round the circle.

Our pictures of foxes illustrate the **dynamics** of $z \mapsto az$. No matter what the value of the complex number a, the origin O is always a **fixed point**, because $T(O) = O$. If we start at some point other than O and **iterate**, we create an orbit which usually spirals into or out from the fixed point. The only exceptions are when either $|a| = 1$ and the orbits encircle O in concentric circular paths, or when a is real-valued, in which case the orbits head straight into or out from the origin with no spiralling. Whether they spiral or not, if the orbits head in towards O, (the case $|a| < 1$) we say that O is an **attracting fixed point** or **sink**, and if they head out away from O (when $|a| > 1$) we say that O is a **repelling fixed point** or **source**.

To gain more flexibility in constructing our pictures, we shall want to create maps which spiral into or out from fixed points other than O. To change the origin, we **conjugate** (not complex conjugate!), in the manner described in Chapter 1, by a translation $S : z \mapsto z + b$ which carries O to b. This means we need to study the conjugated map $STS^{-1}(z)$ which, as we easily calculate, is given by the formula

$$\hat{T}(z) = STS^{-1}(z) = ST(z - b) = S(a(z - b)) = a(z - b) + b.$$

It is easy to verify that, as expected, this maps fixes the point b. The dynamics of $\hat{T}$ is exactly the dynamics of T, translated by the amount b. Thus the orbit of any other point z under $\hat{T}$ spirals into or out from b, the direction, speed and twist of the spiral depending on a.

The map $\hat{T}$ has the form $z \mapsto Az + B$ where A and B are complex numbers. Such maps are sometimes called **affine**. In Project 2.4 we explore the effects of conjugating and composing affine maps in more detail. It turns out that, in the language of Chapter 1, the set of affine maps with $A \neq 0$ forms a **group**.

Complex exponentials

There is another perspective on the polar representation of complex numbers which is at the same time very beautiful and very useful. The fact that multiplying two quantities corresponds to adding their arguments may remind you of another very important bit of mathematics. Before the days of calculators, in order to multiply two large numbers one would take their logarithms and add. This is the principle of the slide rule: the method is based on the power rule

$$a^x a^y = a^{x+y}.$$

In other words, if you want to multiply powers you should add the exponents, as in $3^4 \times 3^2 = 3^6$. One of the most far reaching properties of complex numbers is that the rule relating multiplying numbers to adding arguments is in fact just a complex version of the same equation! The relation between complex numbers and exponentiation is expressed in an wonderful formula of Euler[1] which has good claim to be called one of the most important formulas in mathematics:

$$e^{i\theta} = \cos\theta + i \sin\theta.$$

[1] Leonhard Euler (1707–1783) was a master of calculations using complex numbers, which he brought to new heights. He did not, however, get as far as representing complex numbers geometrically in the plane.

33

Here $e = 2.71828\cdots$ is the base of 'natural' or Naperian logarithms[1] and θ is measured in radians.

Whether this formula is considered to be a *definition* of exponentiation for purely imaginary values, or is a *theorem* based on requiring exponentiation to have certain basic properties, depends on one's point of view. The point is that it makes the basic power rule $a^x a^y = a^{x+y}$ consistent with the complex multiplication rule $\arg(zw) = \arg(z) + \arg(w)$, because according to our new formula, the equation

$$(\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi) = \cos(\theta+\phi) + i\sin(\theta+\phi)$$

can be rewritten

$$e^{i\theta} \times e^{i\phi} = e^{i(\theta+\phi)}.$$

Let us try substituting some special values. For example, at 360° or 2π radians we get $e^{2i\pi} = \cos 2\pi + i\sin 2\pi = 1$, at 180° we get $e^{i\pi} = \cos\pi + i\sin\pi = -1$ and finally at 90°, the formula reads $e^{i\pi/2} = \cos\pi/2 + i\sin\pi/2 = i$. The formula $e^{i\pi} = -1$ has been described as one of the most beautiful in elementary mathematics, involving as it does all the fundamental quantities e, π, i, -1 in such a simple yet profound way.

[1] The number e can be defined as how much money you have after one year if you deposit one dollar in a savings account giving you 100% per annum interest *continuously compounded*. If they credited you with interest every month, you would have $2.613\cdots = (1+\frac{1}{12})^{12}$ dollars after 12 months; if they credited you with interest every day, you would have $2.714\cdots = (1+\frac{1}{365})^{365}$ dollars after 365 days; if they did so every hour, you get $2.71812\cdots$ dollars. And so on. If the interest is 'continuous', you get e dollars!

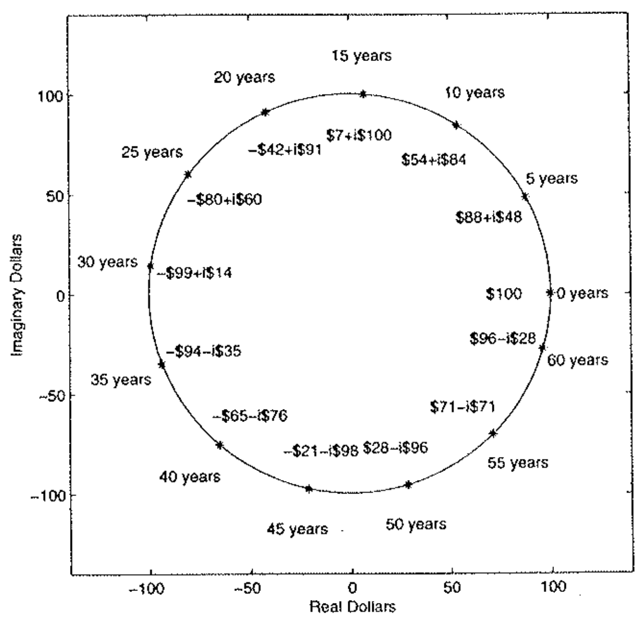


Figure 2.5. The history of Joe Bloggs' investment in the bank of imaginary interest. Each year, the point measuring his bank balance moves round by 0.1 radians or $18/\pi \doteq 5.73$ degrees. Because this number is irrational, no matter how long you go on, the starred point will never land back on the initial point marked 0 years.

There is an amusing and intuitive way to explain this family of esoteric looking formulas which goes like this. Suppose an imaginative and enterprising banker decides to offer an exciting new type of savings

§ 36. Stereographic Projection and Circle-Preserving Transformations. Poincaré's Model of the Hyperbolic Plane

Consider a sphere resting on a horizontal plane (Fig. 239). Let us project the sphere onto the plane from the highest point N ("north pole") of the sphere. The map of the sphere ($P' \to P$ in

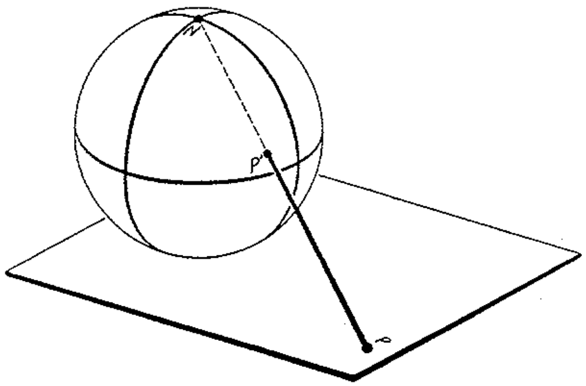


FIG. 239

Fig. 239) produced in this way is called a *stereographic projection.* The entire surface of the sphere with the exception of the point N is mapped onto the entire plane. The image plane is parallel to the plane n tangent to the sphere at N. Furthermore, if p' is the plane tangent to the sphere at P' (Fig. 240), then it follows, from the perfect symmetry of the sphere, that the two planes n and p' form equal angles with the straight line NP' joining their points of contact, and that the line of intersection of n and p' is perpendicular to NP. Since n is parallel to the image plane,

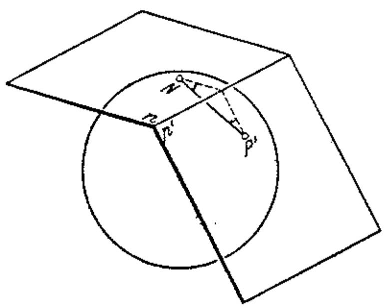


FIG. 240

the image plane also forms the same angle as does p' with the projecting ray PP', and it intersects p' in a straight line perpendicular to PP'. This gives rise to several visually evident properties of stereographic projection. First, if r' is a tangent to the sphere at P' (see Fig. 241) and r is the image of r', then r and r' form equal angles

with PP'. For, r is obtained as the intersection of the image plane with the plane containing r' and NP'; but if two straight lines r and r' are the respective intersections of a plane e with two planes

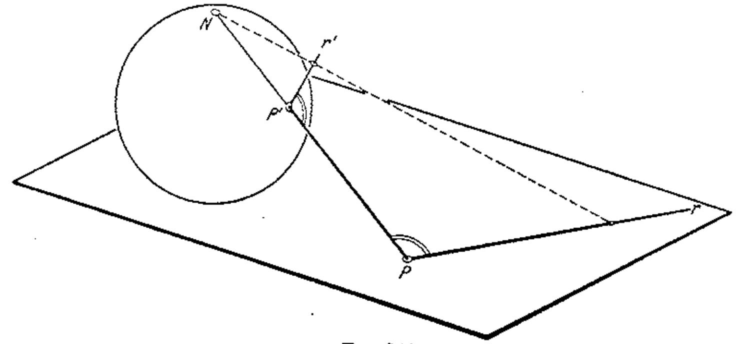


FIG. 241

p and p' where e contains PP' and where p and p' form equal angles with the straight line PP' and intersect in a straight line perpendicular to PP' (see Fig. 242), then r and r' also form equal angles with PP'. The same consideration of symmetry gives us the following additional result: If s' is another tangent to the sphere at P' and if s is its image, then the angle formed by r and s is equal to the angle formed by r' and s'. *Thus stereographic projection reproduces the angles on the sphere without distortion.* For this reason, the mapping is called *angle-preserving*, or *isogonal*; another term is *conformal.*

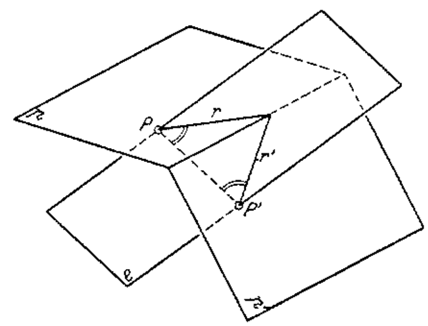


FIG. 242

Now let k' be an arbitrary circle lying on the sphere and not passing through N (Fig. 243). The planes tangent to the sphere at the points of k' envelop a circular cone, whose vertex we shall call S. Since k' does not pass through N, NS is not tangent to the sphere at N and is therefore not parallel to the image plane; let M be the point at which NS intersects the image plane. We shall prove that the curve k that is the image of k', is a circle with M as center.

The proof is apparent from Fig. 243. If P' is an arbitrary point of k' and P is its image, then $P'S$ is tangent to the sphere at P' and PM is the image of $P'S$. Hence $\angle PP'S = \angle P'PM$. Through S we draw the line parallel to PM; let P'' be the point at which it intersects NP. Then either P'' coincides with P', or the triangle $P'P''S$ has equal angles at P' and P'' and is thus isosceles: $SP' = SP''$. But now $PM/P'S = PM/P''S = MN/SN$, so that $PM = P'S \cdot MN/SN$. $P'S$ is constant, because S has the same distance from all the points of K'. Hence it follows from the last formula that PM is constant too. In other words, k is a circle with center M.

Thus the stereographic projection of the sphere onto the plane

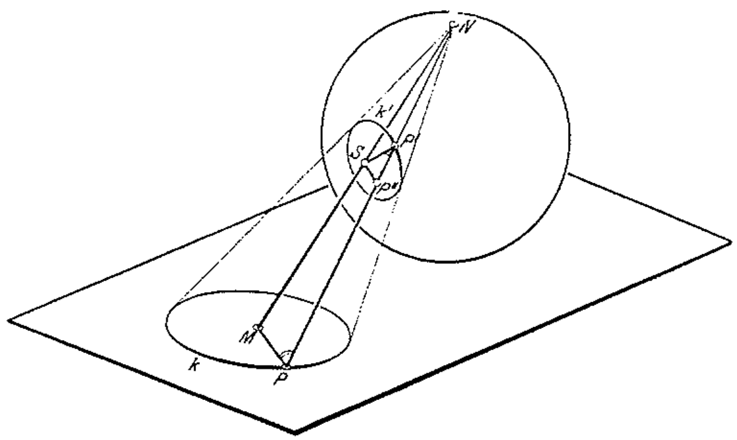


FIG. 243

maps those circles on the sphere that do not pass through N onto circles in the plane; and by reversing the preceding argument we can see that every circle in the image plane is the image of a circle on the sphere. If a circle that is free to move on the sphere approaches a circle passing through N, then NS approaches a tangent to the sphere at N, and thus M recedes to infinity. It follows that the images of the circles on the sphere that pass through N are straight lines of the image plane. This fact is obvious even without the limiting process, since the projecting rays of a circle that lies on the sphere and passes through N lie in the plane of the circle, so that the *straight line* forming the intersection of this plane with the image plane is the image of the circle. Thus we see that

under stereographic projection the set of all circles on the sphere corresponds to the set of all circles and straight lines in the plane. Stereographic projection is *circle-preserving.*

Now consider any mapping α' of the sphere onto itself that maps all the circles of the sphere into circles; for example, α' may be a rotation of the sphere about any diameter (not necessarily one that passes through N). As a result of the stereographic projection, the mapping α' gives rise to a mapping α of the image plane into itself which maps the set of all circles and straight lines into itself. Any such map of the plane into itself is called a *circle-preserving transformation.*

In the Euclidean plane, the circle-preserving transformations are not in general one-to-one, since under stereographic projection no point of the plane corresponds to the point N of the sphere. Now, the mapping α' of the sphere will not in general leave the point N fixed but will transform some other point P', whose stereographic image we shall call P, into N. Then the point P of the plane has no image under the circle-preserving transformation α corresponding to α'. In order to avoid having to make exceptions under the mapping process, we proceed as in projective geometry, by making an abstract extension of the Euclidean plane. But while the projective plane was constructed by supplementing the Euclidean plane with a whole family of "infinitely distant" points, the extension made in the theory of circle-preserving transformations consists in supplementing the Euclidean plane by a single "infinitely distant" point U which is regarded as the image of N under the stereographic projection. As a result of this extension, the relation between the plane and the surface of the sphere becomes one-to-one and continuous, and the circle-preserving transformations become one-to-one mappings; in the example given above, the point P is mapped by the circle-preserving transformation α into U. The corresponding mapping α' of the sphere into itself obviously transforms the circles passing through P' into the circles passing through N; hence α maps the circles passing through P into the straight lines of the plane. Accordingly, it is found expedient to regard the straight lines as "circles passing through the infinitely distant point." The images of parallel straight lines under a circle-preserving transformation are either themselves parallel straight lines, or mutually tangent circles.

36

We have some trivial examples of circle-preserving transformations in the rigid motions, reflections, and similarity transformations of the plane; these transformations map the Euclidean plane one-to-one into itself; hence if we enlarge the plane by adjoining U, we may say that these transformations are circle-preserving transformations leaving U fixed. But it may also be proved, conversely, that the three types of plane transformation just mentioned are the only circle-preserving transformations leaving U fixed. On the basis of this theorem, we may readily obtain an exhaustive description of all the circle-preserving transformations of the plane. Consider a given circle-preserving transformation a_0. Let P be

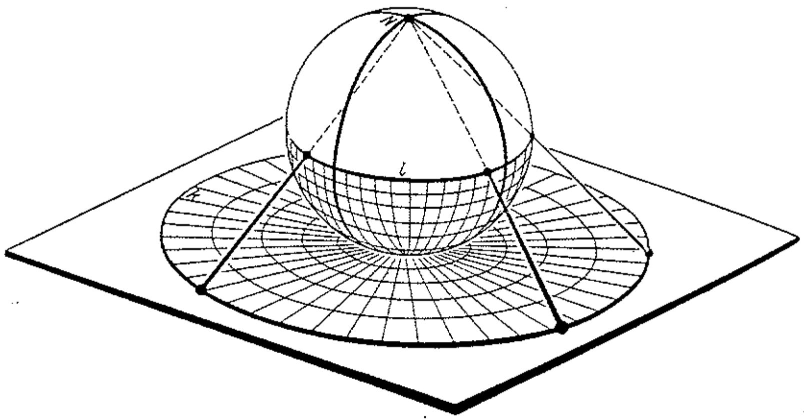


FIG. 244a

the point of the plane whose image under a_0 is U, and let P be the stereographic image of the point P' of the sphere. We now subject the sphere to a rotation a' transforming P' into N. There corresponds to the rotation a' a certain circle-preserving transformation, and the properties of this transformation are connected with the properties of a' in a way that is easy to describe in graphical terms. Like a, the given transformation a_0 moves P to U, so that the transformations a_0 and a can differ only by a circle-preserving transformation that leaves U fixed. It follows, by the theorem we have just cited, that a_0 is identical with a except for a possible rigid motion, reflection, or similarity transformation.

We have mentioned earlier that stereographic projection is angle-preserving. Also, the rotation a' is an angle-preserving transforma-

tion of the sphere, and since a is obtained from a' by a stereographic projection, a must therefore be an angle-preserving transformation of the plane. Since a_0 and a differ at most by an angle-preserving transformation, it follows that *all circle-preserving transformations are angle-preserving*.

Figs. 244a and b elucidate the relation between the maps a and a' by exhibiting prominently a circle k in the plane that passes through P and is the stereographic image of a great circle l of the sphere. Under a', l is transformed into a great circle n that passes through N and has the straight line g as its image. Thus a transforms k into g. From the figures it is plain, moreover, that the interior and

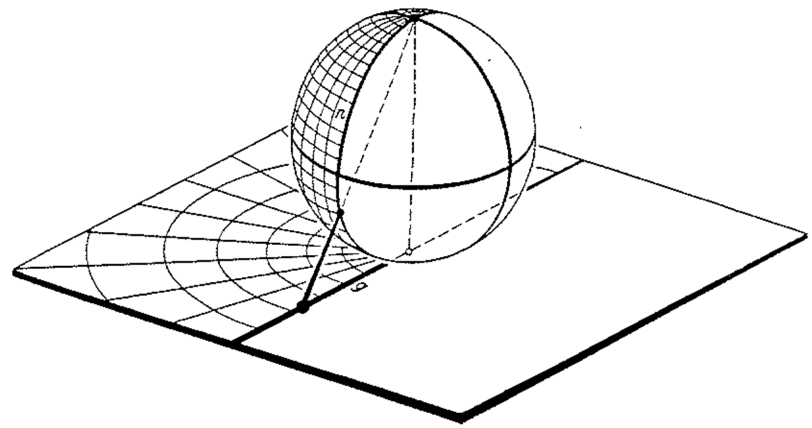


FIG. 244b

the exterior of k are transformed respectively into the two half-planes bounded by g, which is in any case evident from considerations of continuity.

The reflection u of the plane in g is a circle-preserving transformation. Hence the mapping $i = aua^{-1}$ is a circle-preserving transformation that leaves every point of the circumference of k fixed and interchanges the interior with the exterior of the circle. The map i is called an *inversion* in the circle k, or a *reflection* in the circle k, or a plane transformation by reciprocal radii. This transformation is particularly important, and we shall therefore discuss it in some detail.

Let h be a circle intersecting k at right angles at a point R (see Fig. 245). Then h and k have a second point S of intersection at which they are also perpendicular. Then the tangents to h at R

62

and S are radii of k intersecting at the center M of k, which is therefore exterior to h. The inversion i transforms h into a circle h', and this circle must also pass through R and S, because R and S remain fixed. Since the inversion is angle-preserving, h' intersects the circle k at R and S at right angles. But this is possible only if h' is identical with h. Hence every circle h that intersects k at right angles is transformed by i into itself. Since the interior and exterior of k are interchanged, the two arcs into which k divides h must also be interchanged.

Consider a straight line l passing through M, for example the straight line RM. Let its second point of intersection with k be R' (see Fig. 245). Then l must be transformed into a circle or straight line l' meeting k at right angles at R and R'. This is possible only if l' is identical with l. Accordingly, the inversion transforms each diameter of k into itself. Since the only points that these straight lines have in common in the enlarged plane are M and the infinitely distant point U, it follows that the inversion interchanges M with U. The totality of straight lines not passing through M is therefore interchanged with the totality of circles passing through M.

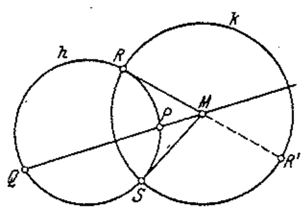


Fig. 245

Now let P be a point of h different from R and S. The image of P under the inversion i can only be at the second point Q of intersection of the straight line MP with h, because MP and h are each mapped into themselves. By the elementary theorem about intersecting chords of a circle, we have $MP \cdot MQ = MR^2$. Q is called the *inverse point* of P with respect to the circle k. Thus we have found a method of determining the inverse of any point P with respect to k without the use of the auxiliary circle h: If r is the radius of k, the inverse Q of the point P is that point on the ray MP for which $MP \cdot MQ = r^2$.

It may be proved that every circle-preserving transformation can be expressed as the resultant of at most three inversions. We shall consider, in particular, the totality of circle-preserving transformations that transform a given circle k into itself and also the interior of k into itself. Evidently these maps constitute a group, which we shall call H. If n is a circle cutting k at right angles, the inversion

in n certainly belongs to H. It can be proved that every map of the group H can be produced by three inversions in circles perpendicular

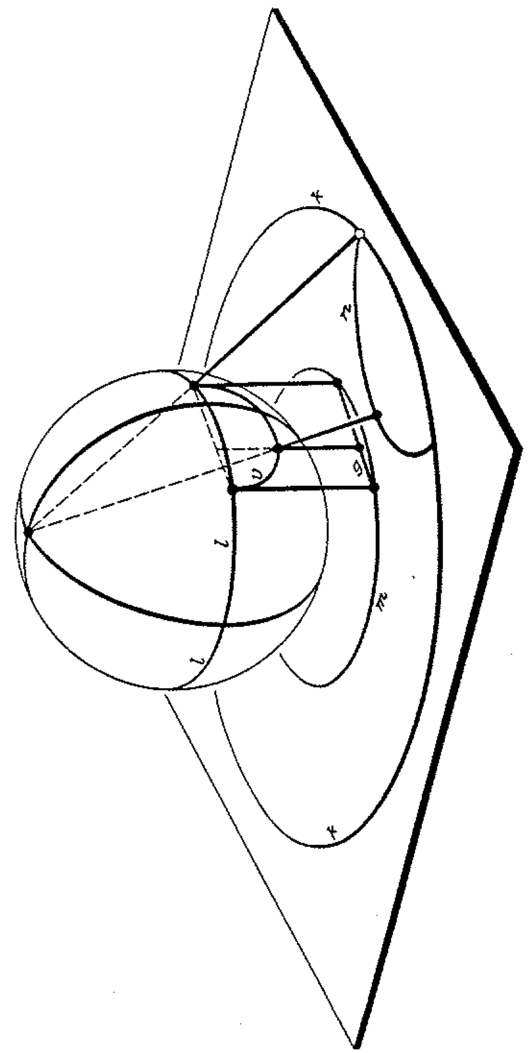


Fig. 246

to k, that is, by three inversions which themselves belong to H.

Stereo projection.

We will define a map from the sphere minus a point ONTO the plane which is one-to-one. This will show that we can think of the plane plus a "point at infinity" as equal to the sphere.:

$$S^2 = \mathbb{R}^2 \cup \{\infty\},$$

where S^2 denotes the sphere, and $\mathbb{R}^2$ the Euclidean plane. The map enjoys the additional property of taking circles (small or great) to circles or lines, and preserving all angles.

Call the center of the sphere O, and the point we subtract from the sphere 'N' (for North). Place the plane so that it does not pass through N and so that ray ON is normal to it. We think of N as a lighthouse, with beams of light shining out from it. We now define the stereographic projection map

$$F : S^2 \setminus \{N\} \to \mathbb{R}^2$$

Take a point P on the sphere. Draw line NP through N and P. Think of it as the beam of light from N to P. It intersects the sphere at a unique point P'. This intersection point P' is $F(P)$. Write $W = F(P)$. To define the inverse map, join W to N by line NW. This intersects the sphere in a unique point, and that point is our old P.

Question: Light rays beaming out of N hit the plane and thus define our map. What happens with the rays parallel to our plane? They never hit. This helps explain why we should extend our map by sending N to ∞.

Formulae. To obtain formulae for F we need to place the plane somewhere, and choose coordinates. We will place the plane passing through the center of the sphere O, and mark it up with xy coordinates. We put Cartesian coordinate (x, y, z) on space, with $0 = (0,0,0)$ and $N = (1,0,0)$, so that the sphere S^2 is defined by $x^2 + y^2 + z^2 = 1$, and our plane is obtained by setting $z = 0$. The computation of F depends on similar triangles and splitting space up into the xy-plane times the z-axis. We use notation indicating this splitting. Write $(x, y) = \vec{r}$, and $r = |\vec{r}| = \sqrt{x^2 + y^2}$. Then

$$P = (\vec{r}, z) \qquad r^2 + z^2 = 1$$

is a point lying on the sphere, and

$$F(P) = (\vec{w}, 0)$$

is a point on the plane, and our problem is to find a formula $\vec{w} = F(\vec{r}, z)$ and to invert this formula.

Set $\hat{u} = \frac{1}{r}\vec{r}$. This is a unit vector pointing in the same direction as $\vec{r}$. And r is the distance from the z-axis to P. The line NP lies in the plane spanned by $(0,0,1)$ and $(\hat{u}, 0)$. So we can write F in the form:

$$F(r\hat{u}, z) = (\rho\hat{u}, 0)$$

where we have set

$$\rho = |\vec{w}|.$$

It remains to relate r, z and ρ. This is done using similar triangles. (See figure.) and the fact that (r, z) lie on a circle – the guts of the computation is a planar geometry problem

CLAIM:

$$\rho = r/1 - z$$

,

$$z = (\rho^2 - 1)/\rho^2 + 1$$

$$r = 2\rho/(1 + \rho^2)$$

Proof. By similar triangles, $\rho : 1 = r : (1 - z)$ See figure. This is the first relation. Squaring: $\rho^2 = r^2/(1 - z)^2$. But since (r, z) lie on the unit circle we have $r^2 = 1 - z^2 = (1 - z)(1 + z)$ which yields $\rho^2 = (1 + z)/(1 - z)$. Now solve for z in terms of ρ^2 from this: $(1-z)\rho^2 = (1+z)$, or $\rho^2 - 1 = z(\rho^2+1)$ which yields the second result upon division. Finally, compute $r^2 = 1 - z^2$ using the second relation to obtain $r^2 = \frac{(\rho^2+1)^2-(\rho^2-1)^2}{(\rho^2+1)^2} = \frac{4\rho^2}{(\rho^2+1)^2}$ which is the third relation, upon taking the square root.

Putting these together we see that

$$F(\vec{x}, z) = \frac{1}{1 - z}\vec{x}$$

and that

$$F^{-1}(\vec{w}) = (\frac{2}{1 + |\vec{w}|^2}\vec{w}, \frac{|\vec{w}|^2 - 1}{|\vec{w}|^2 + 1})$$

Here we have dropped the extraneous 0 from the point $(\vec{w}, 0)$ in the xy plane.

Application 1. Pythagorean triples. A Pythagorean triple is a triple of integers a, b, c which form the sides of right triangle. Thus $a^2 + b^2 = c^2$. The simplest of these is $(3, 4, 5)$. Multiplying a pythagorean triple by an integer k trivially yields another , since if $a^2 + b^2 = c^2$ then $k^2a^2 + k^2b^2 = k^2c^2$. We will think of these as the same.

Question: How many distinct pythagorean triples are there?

Answer: as many as there are rational real numbers.

To see this, first divide by c^2 to get $(a/c)^2 + (b/c)^2 = 1$. Thus $(x, y) = ((a/c), (b/c))$ is a point on the unit circle whose coordinates are rational numbers. Conversely, given any such point, if we take c to be the least common multiple of the denominators of x and y we get a Pythagorean triple. Now use stereo projection (F^{-1} above), but in one-dimension less:

$$(x, y) = (\frac{2w}{1 + w^2}, \frac{w^2 - 1}{w^2 + 1}),$$

$w \in I\!R$ being a point on the number line. If w is rational then so are x, y. This gives a map taking all rational numbers to Pythagorean triples. Using F instead of F^{-1} we see that this correspondence is one-to-one and onto.

Eg: $(4, 3, 5)$ above yields $(x, y) = (4/5, 3/5)$ which corresponds to $w = 2$ above.

4 #7

Inversion through a circle.

Fix a circle with center C and radius ρ in the Euclidean plane. Define a map f from the plane minus C to the plane minus C as follows. If $P \neq C$ is in the plane, then $f(P)$ lies on the ray joining C to P and $CP \cdot Cf(P) = \rho^2$. (Here we write CX for the distance between points C and X.)

Prove:

i) If the circle is the unit circle $x^2 + y^2 = 1$, then $f(x, y) = (x, y)/(x^2 + y^2)$.

Prove that if we write $z = x + iy$ then the inverseion is given by

$$f(z) = 1/\bar{z}$$

Now return to the case of a general circle, not necessarily centered at the origin, or of radius 1.

ii) Prove if P lies on the circle, then $f(P) = P$.

iii) Prove that for any $P = C$ we have $f(f(P)) = P$.

iii) Show that as P tends to C, $f(P)$ tends to infinity.

Because of iii) we extend f to the whole plane by declaring $f(C) = \infty$ where ∞ is thought of as an extra point added to the plane. And $f(\infty) = C$.

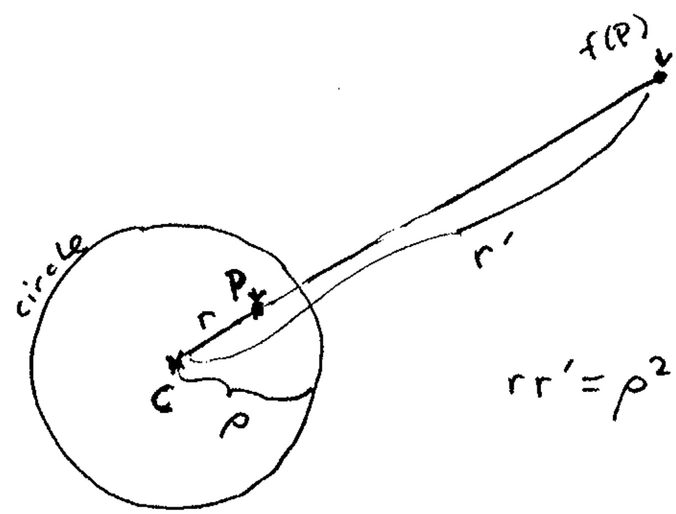
The plane with the added point at infinity is called the "extended plane" or the "Riemann sphere". It is (topologically) isomorphic to the geometric sphere and the isomorphism is given by stereographic projection.

iv) Prove that the image of any line ℓ not passing through C is a circle passing through C. You might want to do this in two steps:

iv. a) If ℓ does not intersect the circle then its image $f(\ell)$ lies entirely inside the circle.

iv.b) If ℓ does intersect the circle, then its image $f(\ell)$ intersects the circle at the two points where ℓ intersects the circle.

v) [Harder] Prove that f preserves angles: if two curves c_1, c_2 intersect at a point P and make an angle θ there, then their image curves $f(c_1), f(c_2)$ intersect at $f(P)$ and make the same angle θ there.



picture of inversion

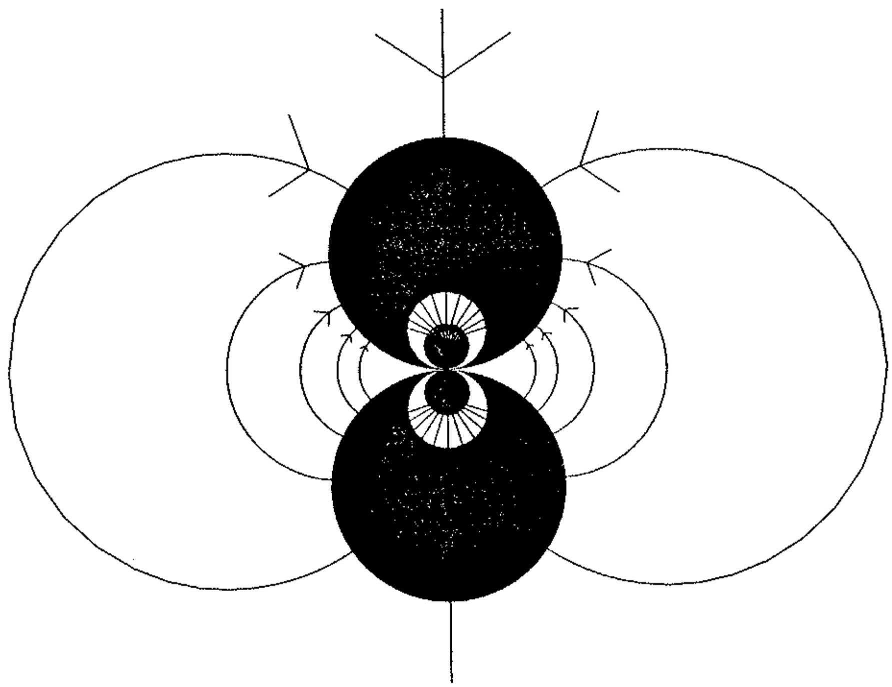


Figure 3.6. A conjugate of a translation. This illustrates the conjugation of a translation by the map $z \mapsto 1/z$. The origin is fixed and the blue circles are trajectories of the map. The colours show the successive powers of the map: the small bottom green region is expanded to the yellow region, then to the red, then to the whole outer white region (including ∞). This is then contracted first to the top red region, then to the top yellow region then to the green. This figure also results from stereographic projection of the map in Figure 3.5 it after turning that sphere upside down.

map the whole region outside one red disk so that it exactly covers the other. Running the map backwards reverses this process, mapping the outside of the second red disk to the inside of the first.

Iteration of the original map T on the plane pushes the left hand red half plane forward inside itself. Thus repeating the map $\hat{T}$ pushes the red disk forward to another tangent disk, shown in yellow, which is nested inside. Successive iterations of $\hat{T}$ push the disk further and further inside itself, which means that under forward iteration all orbits eventually get nearer and nearer to 0. Iterations of $\hat{T}^{-1}$ push the disks backwards so they again nest down to 0, this time from the opposite side.

The algebra of linear fractional transformations

The reader who has been checking the calculations in the last sections may rapidly begin to wish for a proper machine. Such a machine becomes an essential tool when we start making computer pictures which involve iterating many different maps many times. In the remainder of this chapter we shall develop a systematic theory of all maps of the

42 128

form

$$T : z \mapsto \frac{az+b}{cz+d}$$

where the coefficients a, b, c, d are arbitrary complex numbers subject to the one condition that $ad - bc \neq 0$. These are called either **linear fractional transformations** (because they are fractions of linear expressions) or **Möbius transformations**, after the nineteenth century German mathematician August Möbius, who first studied such transformations systematically, showing that they are the most general transformations of the extended complex plane which map circles to circles. (As usual, lines are included here as a special kind of circle.) All the special examples studied thus far are of this type.

There is a mechanical procedure for composing linear fractional transformations using the algebra of 2×2 matrices. Understanding this fully involves a hefty dose of algebra, and it's possible to understand the rest of the book qualitatively without working it through. However for the reader who actually wants to implement our programs or follow the technical derivations later, this section and the next more geometric one will be among the most important reference points of the book.

First of all, let's see that T makes sense for any z which is either a complex number or ∞. We can work out the value for $z = \infty$ as follows:

$$T(\infty) = \frac{a\infty + b}{c\infty + d} = \frac{a + b/\infty}{c + d/\infty} = \frac{a+0}{c+0} = \frac{a}{c}.$$

The other thing to worry about is what happens when the denominator is 0, that is when $z = -d/c$:

$$T(-d/c) = \frac{-ad/c + b}{-cd/c + d} = \frac{-ad + bc}{-cd + dc} = \frac{-ad + bc}{0} = \infty.$$

This works because the assumption that $ad - bc \neq 0$ rules out any difficulties with the illegal value $0/0$. We shall see shortly that exactly the same assumption ensures that T has an inverse. The quantity $ad - bc$ is called the **determinant** of T.

How do linear fractional maps compose? Remembering our convention that TT' means first do T' and then do T, we just calculate out: if

$$T(z) = \frac{az+b}{cz+d} \quad \text{and} \quad T'(z) = \frac{a'z+b'}{c'z+d'}$$

43 ~~39~~

This shows that the composition TT' is again a linear fractional transformation.

The rule for composition looks rather messy. However it is really quite simple if we introduce the basic algebraic idea of representing T with the **matrix**

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

formed from the coefficients of T. 'Matrix' is just a fancy term for a rectangular array of numbers, in this case a 2×2 square.

The algebra of 2×2 matrices is summarized in Box 7. The rule for multiplication is shown using arrows: you choose any row of the first matrix and any column of the second and, moving across, take products and add. The coefficients in our formula for the composition of two linear fractional transformations are exactly the same as those given in Box 7 for the matrix product! *Composing two linear fractional transformations is the same as multiplying the corresponding* 2×2 *matrices.*

Notice that if we multiply all the coefficients a, b, c, d in a linear fractional transformation by the same complex number t, the map T doesn't change. This useful fact just uses the identity

$$\frac{taz + tb}{tcz + td} = \frac{az + b}{cz + d}.$$

Let's try inverting the map $T : z \mapsto (az+b)/(cz+d)$. To do this, we set $T(z) = w$ and solve the equation $w = (az+b)/(cz+d)$ for z in terms of w. Multiplying up and rearranging we find that $z(cw - a) = -dw + b$ so that $T^{-1}(w) = (dw - b)/(-cw + a)$. So T^{-1} has the matrix $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. This is not quite the inverse *matrix* because

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = (ad - bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

The factor $D = ad - bc$ multiplies all entries inside the matrix on the right. The factor is the determinant of the matrix for T. The matrix inverse of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\begin{pmatrix} d/D & -b/D \\ -c/D & a/D \end{pmatrix}$. Because multiplying the matrix coefficients by a complex number doesn't change the transformation, the matrix inverse gives the same *map* T^{-1}.

In Chapter 1 we introduced the idea of a group of transformations of the plane. Remember that a collection of transformations is called a group provided that:

(1) if S and T are in the collection then so is ST;
(2) if S is in the collection, then so is S^{-1}.

Box 7: Matrix algebra and Möbius maps

Matrices are usually introduced as maps which act on vectors by the formula

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Such maps compose using matrix multiplication:

$$\begin{pmatrix} a \quad b \\ \xrightarrow{\hphantom{ab}} \\ c \quad d \end{pmatrix} \begin{pmatrix} a' & b' \\ c' & \downarrow d' \end{pmatrix} = \begin{pmatrix} aa' + bc' & ab' + bd' \\ ca' + dc' & cb' + dd' \end{pmatrix}.$$

The arrows have been put in only to indicate how you compute the product: each number in the matrix product is a 'dot product' of a *row* of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with a *column* of $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$. The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is known as the **identity matrix** and is often written I.

The determinant of the 2×2 matrix $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $D = ad - bc$. A matrix M has an **inverse** M^{-1} such that $MM^{-1} = M^{-1}M = I$ provided that $D \neq 0$, given by the formula

$$M^{-1} = \begin{pmatrix} d/D & -b/D \\ -c/D & a/D \end{pmatrix}.$$

A Möbius map is formed from the matrix M by setting

$$T(z) = \frac{az + b}{cz + d}.$$

One can compose and invert Möbius maps using the algebra of 2×2 matrices. Multiplying all coefficients of M by the same complex number does not affect the map. By dividing through by $\sqrt{D}$, one can always arrange that M has determinant 1.

We can replace 'the plane' in this definition by 'the extended complex plane' or Riemann sphere. So we have just shown that both of these two statements are true for the collection of all linear fractional transformations. In other words, we have just verified that *the set of linear fractional transformations forms a group.*

Here is a trick which very often makes inverting and other calculations easier. Because $\begin{pmatrix} d & -b \\ -c & d \end{pmatrix}$ and $\begin{pmatrix} td & -tb \\ -tc & td \end{pmatrix}$ define the same Möbius transformation, it is often handy to choose the number t so as to

45

make the determinant come out as 1. Start with a matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with determinant $D = ad - bc$. Change the coefficients by dividing by $\sqrt{D}$, replacing a by $a/\sqrt{D}$ and so on. We get the same map and the new determinant $(ad - bc)/D$ simplifies to 1. Thus if $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $ad - bc = 1$, then $T^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. From the formulas in Box 7, one can calculate that the product of matrices with determinant 1 also has determinant 1.[1] Very often, we shall just assume the matrix coefficients have been chosen at the outset with $ad - bc = 1$.

What we have just said shows that the subset of 2×2 complex matrices with determinant 1 itself forms a group, usually denoted in mathematical texts by the rather mysterious notation $SL(2, \mathbb{C})$. Decoding, L stands for linear, the 2 indicates their size and $\mathbb{C}$ shows that the entries are complex numbers. The 'S' stands for 'special' and indicates that $ad - bc = 1$.

Some linear fractional transformations are already familiar friends. First of all, taking $a = d = 1$ and $b = c = 0$, we get the map $T : z \mapsto (1 \cdot z + 0)/(0 \cdot z + 1)$, which is a complicated way of writing the identity map $z \mapsto z$. Secondly, it includes the affine maps $z \mapsto az + b$ coming from the arithmetic of complex numbers, which we studied in Chapter 2. These can be expressed in the form $T : z \mapsto (az+b)/(0 \cdot z+1)$, given by a 2×2 matrix with $c = 0$. Since $T(\infty) = a/c$, maps with $c = 0$ are exactly those such that $T(\infty) = \infty$. In other words, our old friends $z \mapsto az + b$ are those Möbius maps for which ∞ is a fixed point. Important subclasses are the **pure translations** $z \mapsto z + b$ and the **pure scalings** $z \mapsto az$ (for which we shall often replace the letter a by the letter k). The coefficient matrix for translation $z \mapsto z + b$ is just $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$, which is already normalized with determinant 1. To get determinant 1 for the scaling map $z \mapsto kz$, we have to write the matrix in the rather surprising form $\begin{pmatrix} \sqrt{k} & 0 \\ 0 & 1/\sqrt{k} \end{pmatrix}$.

Linear fractional transformations also include the map $J : z \mapsto 1/z$ which we have already found useful for interchanging the North and South Poles. The obvious way to write the coefficient matrix is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, but watch out! This matrix has determinant -1, so a correct normalized form would be either $\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix}$. Details like this will reappear in our algebra time and again.

How many Möbius maps are there? This is a pretty vague question

[1] This is done by checking that the determinant of the product of two matrices is the product of their determinants.

Math Horizons
MAA

46

"We have before us a world with many coincident parallels to a given line and another world with no parallels to a given line. After a quick look at each, we will see a surprisingly simple connection between these worlds."

The Hyperbolic Tourist

Michael McDaniel
Aquinas College

Our Tour Begins

Are you tired of strict adherence to Euclid's parallel postulate? How about a little adventure in places where judging by appearances is almost lethal? Well then, grab your compass, straightedge, and about twenty sheets of scrap paper: two non-Euclidean disks await your attention. As tourists, we don't get to see every detail; yet, there's a nifty connection between hyperbolic and elliptic geometry that you have to see.

About 100 years ago, there was no such tour. Only a few mathematicians like Bolyai and Lobachevsky in the 1850s had described non-Euclidean worlds and their publications were not popular. But when models of non-Euclidean geometries were created that had points, lines, angles, triangles, parallels, and distance, people went for the new stuff. The obstacle of visualizing these strange worlds had been overcome.

By the late 19th century, many mathematicians had embraced the revolutionary idea that a geometry was a set of axioms, some undefined terms and some definitions: an axiomatic system. An interpretation of undefined terms in which all the axioms are true is a model. Up until that time, Euclidean geometry was THE Geometry, largely due to the usefulness of the Euclidean interpretation of points, lines and planes. This model seemed to fit the way people build and measure things; the proofs were regarded as essential to developing logical thinking; the subject grew richer each generation. Any other geometry seemed erroneous in comparison.

But the truth is, hyperbolic and elliptic geometry have the same logical standing as Euclidean geometry and the disk models are the proof. Euclid's first four axioms are true in each disk, as well as a negation of the fifth axiom (through a point not on a given line, there is exactly one line parallel to the given line.) We have before us a world with many coincident parallels to a given line and another world with no parallels to a given line. After a quick look at each, we will see a surprisingly simple connection between these worlds.

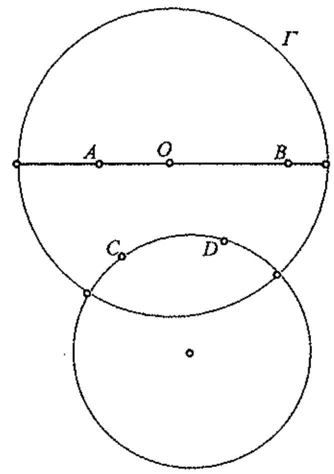


Figure 1. Hyperbolic lines $\overleftrightarrow{AB}$ and $\overleftrightarrow{CD}$.

The Hassle

Since the disk interpretations use Euclidean components, some math words have different meanings in the different geometries. In order to differentiate between which type of geometry is being employed, we will mention the type whenever it is not clear in context. Some truly clunky sentences lie ahead. Here is a preview: the hyperbolic center of a hyperbolic circle is Euclidean collinear with O and the Euclidean center of the hyperbolic circle. Remain calm: we have a constructive model, where we can actually see what we are talking about.

The Hyperbolic Interpretation

Here are the rules for the Poincaré disk. The points are Euclidean points inside a circle we will call Γ. Points on Γ, while not considered points of hyperbolic geometry, will be useful to us in many ways. The point in the center of Γ we will call O. The lines are diameters of Γ and the arcs of circles orthogonal to Γ. (See Figure 1.) This second type of line requires a bit more explanation. When two circles intersect in two points, we can measure the angles formed by their tangents at the points of intersection. When these tangents are perpendicular, we say the circles are orthogonal. The part of a circle orthogonal to Γ which is inside Γ is interpreted as a line. Angles are measured by Euclidean angles between tangents.

46

47

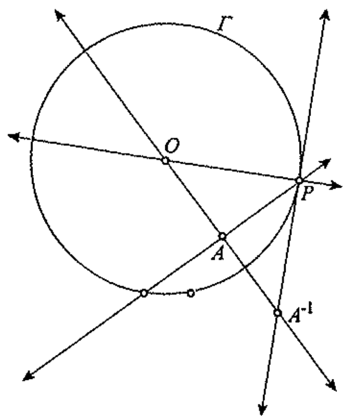


Figure 2. A point A and its inverse A^{-1} on the hyperbolic disk.

The First Axiom

Euclid's first axiom is that two points determine a line. With compass and straightedge in hand, we will see how this axiom is true in the disk. Let A and B be two points inside Γ. If A and B are Euclidean collinear with O, then the hyperbolic line through A and B is a diameter of Γ. This is also the case if one of the two points happens to be the same as O. In order to handle the case in which A and B are not on a diameter, we will need to find the inversion of A through Γ, which we will call A^{-1}. Place your straightedge across O and A and draw the ray from O through A and beyond. Construct the perpendicular to this ray through the point A. This perpendicular will intersect Γ at two places. Choose one of them and call it P. Extend the ray from O through P and then construct the perpendicular through P and extend it until it intersects the ray $\overrightarrow{OA}$. This point of intersection is A^{-1}. (See Figure 2.) **Any hyperbolic line through A also contains A^{-1}.** The center, the point and its inverse have the following relationship in Euclidean geometry: $|OA||OA^{-1}| = |OP|^2 = r^2$, where r is the radius of Γ.

This construction gives us three points on the hyperbolic line we seek; that is, A, B, and A^{-1} are on a Euclidean circle. The experienced reader will recognize the rest of the case. Consider two segments with pairs of these three points as endpoints: these segments are chords on the circle we seek. Simply construct the perpendicular bisectors of the two segments. They will intersect at the center of the Euclidean circle orthogonal to Γ through A and B.

The Tourist Must Try These!

1. Let A be any point in Γ except the center. Construct the one hyperbolic line through A perpendicular to $\overline{OA}$.

2. Even though points on Γ are not in the model, the Euclidean circle orthogonal to Γ through two points on Γ gives a hyperbolic line. Choose any two points on Γ, not collinear with O, and construct this hyperbolic line.

3. Construct at least three hyperbolic lines through a hyperbolic point A and verify that they all contain A^{-1}.

4. We can consider points on Γ as Euclidean points, at least. Show that, for every P on Γ, $P^{-1} = P$.

The Second Axiom

Euclid's second axiom is that any line may be extended indefinitely. A quick glance at our disk model would make any hyperbolic novice think that hyperbolic lines have finite length. This much is certainly true: hyperbolic lines have finite Euclidean length. It should be no surprise that the distance formula in the disk is non-Euclidean. Let A and B be points in the disk and let the hyperbolic line through A and B intersect Γ at C and D. For coordinates, let O be the center of the unit disk in the complex plane with the radius of Γ equal to 1. We will denote the length of $\overline{AB}$ as $|AB|$ where

$$|AB| = \left|\ln\left(\frac{A-C}{A-D} \div \frac{B-C}{B-D}\right)\right|.$$

Problem 6 in the next set of homework problems shows how the second axiom is satisfied in this model.

More Fun for the Tourist.

5. Let $A = a + Oi$. Show that $|OA| = \ln\frac{a+1}{1-a}$.

6. Show that $\lim_{a\to 1^-} |OA| = \infty$.

The Third Axiom

Both geometries under consideration have circles. Hyperbolic circles are, in fact, Euclidean circles. Their hyperbolic centers, however, are not the Euclidean centers unless the Euclidean center is O. The hyperbolic center of a hyperbolic circle is Euclidean collinear with O and the Euclidean center of the hyperbolic circle. As we have seen, hyperbolic length of a hyperbolic segment increases as one endpoint approaches Γ. So, when a hyperbolic circle of constant radius approaches Γ, its hyperbolic center also approaches Γ.

In order to find the hyperbolic center by construction, we will have to use the definition of the hyperbolic center: the intersection of all hyperbolic lines orthogonal to the hyperbolic circle. That is, any Euclidean circle orthogonal to Γ *and* to the hyperbolic circle must pass through the hyperbolic center. To find such a hyperbolic line, we must solve a case of double orthogonality. We start with Γ drawn big enough to accommodate a Euclidean circle Π drawn with its Euclidean center about halfway between O and Γ; make sure Π is entirely within Γ. Let the Euclidean center of Π be called E. Choose any

42

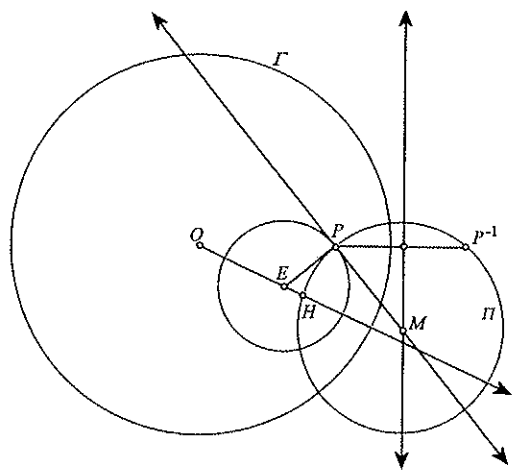


Figure 3. Circle with Euclidean center E and hyperbolic center H.

point on Π not collinear with O and E and call this point P. Find P^{-1}. If your choice of P leads to one part of your construction getting clogged with many lines and arcs, try a different choice.

At this moment, we seek a hyperbolic line (Euclidean circle) through P and P^{-1} that is orthogonal to Π. Getting the circle through P and P^{-1} will guarantee its orthogonality to Γ. We know the hyperbolic center is on $\overline{OE}$. All we need to do is make sure this hyperbolic line is also orthogonal to Π by finding the intersection of the tangent to Π at P with the perpendicular bisector of PP^{-1} (See Figure 3.)

The Fun Never Stops!

7. Draw a sequence of circles that vary in their proximity to Γ. Construct their hyperbolic centers. If you are not convinced that the hyperbolic center looks closer to Γ when the circle is closer to Γ, too bad. There's a bus waiting to take you back to Euclidean geometry.

The Fourth and Fifth Axioms

Euclid's fourth axiom, that all right angles are equal, follows from our interpretation of angles. All right angles are 90°, so that's that. In hyperbolic geometry, we follow a negation of Euclid's fifth axiom. In our model, through a point not on a line, there are many lines which do not intersect the given line.

Three More Fun Things to Do.

8. In the Euclidean plane, any two perpendicular lines have the property that any third line must intersect at least one of them. Explain why this property does not hold in the Poincaré disk.

9. In hyperbolic geometry, nonintersecting lines that do not meet on Γ have exactly one hyperbolic line that is mutually perpendicular. Construct the mutual perpendicular for the pair of lines pictured in Figure 1. One of the lines is a diameter.

10. Choose a hyperbolic line and a point not on this line. Construct at least two hyperbolic lines through this point that do not intersect the chosen line. (Helpful hint: do not choose a point very close to the chosen line.)

The Elliptic Interpretation

The parallel postulate has two different negations. In one form, the negation is taken that through a point not on a line there are more than one line parallel to the given line: hyperbolic geometry. The other form of the negation is that through a point not on a line, there is no lines parallel to the given line. This version, along with Euclid's other axioms and isometries gives us elliptic geometry.

The Model

Remember how strange it was to enter the world of hyperbolic geometry, yet after a while, we were able to reason our way through some interesting problems? The same thing happens with this nonorientable world of elliptic geometry. We have a model that is surprisingly similar to the Poincaré disk.

Our space begins with all the points in and on a Euclidean circle. The points on the circle are identified with their antipodes (points diametrically opposite on the circle), but no other points have antipodes. Lines are diameters and arcs of circles through antipodal points. Angles are again measured with tangents. The general construction of an elliptic line through two points depends on finding a third point that the line must go through, a point we will call the mirror-inverse of one of the given points.

Axiom 1 and the Highlight of Our Tour

Given the circle with center O (the thick blue circle given in Figure 4) and two points A and B inside the circle, we construct the points on the diameters through A and B that are reflected across O. The dotted circle in the drawing shows how to find the mirrors of A and B, which we will call Z and Y (reflected across the alphabet). Now we construct the inverse of these mirrors in exactly the same way we constructed inverses in hyperbolic geometry. The elliptic line we seek through A and B must contain Z^{-1} and Y^{-1}. (We'll see why in a moment.) The rest of the story goes as we expect: the perpendicular bisectors of $Y^{-1}Y$ and $\overline{AB}$ meet at the desired center. The elliptic line has been drawn as an arc, rather than as a circle, in the figure below. Through two points, there is one line. If the two points are Euclidean collinear with O, then the line is a diameter. The same is true if one of the two points is O. If the two points A and B are not Euclidean collinear with O, the construction above defines the desired line. For those

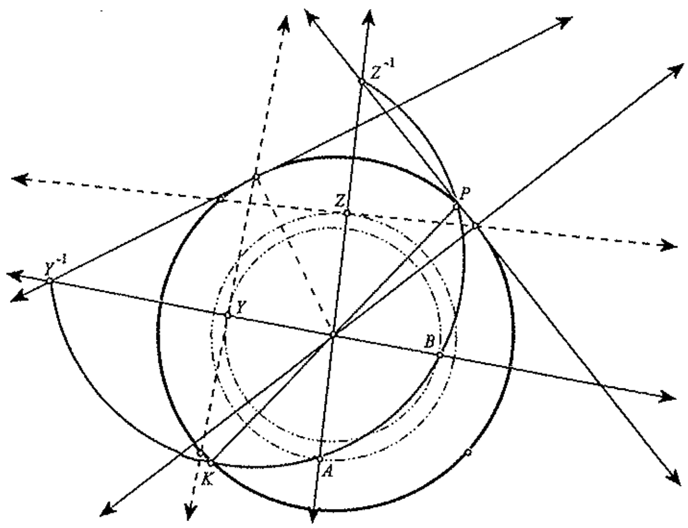


Figure 4. Elliptic line through *A* and *B*.

who successfully explored hyperbolic geometry, this construction should have been surprisingly familiar.

The proof will show that the elliptic line we seek goes through A, B, and the inverses of the reflections of A and B. In the figure below, we have the given points and their reflections, labeled Z and Y.) We have inverses of these points. It is a Euclidean geometry problem to show $\overline{AB}$ is parallel to $\overline{ZY}$. We can see that Z and Y and Z^{-1} and Y^{-1} form a cyclic quadrilateral because they are all on the hyperbolic line through Z and Y. To show ΔAOY^{-1} is similar to ΔBOZ^{-1} we use the definition of the inverse of a point (substituting for mirrored points): $|AO||OZ^{-1}| = r^2 = |OB||OY^{-1}|$.

The similar triangles are not marked in the drawing because it's getting pretty crowded down there. Using the corresponding angles, we can show that A and B and Z^{-1} and Y^{-1} form a cyclic quadrilateral by getting opposite angles supplementary. All that remains is to show that the circle we just found has the endpoints of a diameter of Γ. To do this, we consider a point P where the elliptic line candidate we have just constructed intersects the given circle. Let K be the reflection of P; our constructed line must contain K^{-1}. But $K^{-1} = K$. Therefore our elliptic line candidate contains antipodal points, which means we have an elliptic line.

In both geometries, we use the inverse of a point in order to construct the line through two points. In elliptic geometry, we take the inverse of a reflection of the given point; in hyperbolic, we skip the reflection. With the reflection, we get no parallels, without the reflection, we get infinite parallels.

That the remaining Euclidean axioms are true requires further exploration. But the cruise ship is getting ready to depart. Wipe the mango juice off your chin, jam the shirt into your bag and get a taxi to take you to a college geometry course. Each geometry you've seen has as many layers to explore as Euclidean geometry. It is exactly like finding two strange places where everything you usually do is different, yet tantalizingly familiar. Go and see the asymptotic hyperbolic triangles that have angles of size zero but maximal area. Check out the point where all elliptic lines perpendicular to the given line meet: it's called a pole and it acts just like the North Pole with the Equator as the given line. How about hyperbolic trigonometry? It uses sinh a and cosh b! Somebody help me! I'm beginning to sound like I was born here! ■

Further Reading

C. Goodman-Strauss, Compass and Straightedge in the Poincaré Disk, *American Mathematical Monthly*, 108 1, January 2001.

A. Baragar, *A Survey of Classical and Modern Geometries*, Prentice-Hall, 2001.

D. Hecker, Constructing a Poincaré Line with Straightedge and Compass, *College Math Journal*, November 2003.

We now proceed to tie this discussion in with the model of the hyperbolic plane constructed in the previous section. Let the hyperbolic plane be represented by the interior of a circle m lying in a horizontal plane. Let us place on the plane a sphere having the same radius as m and touching the plane at the center of m (see Fig. 246). We now project the circumference and the interior of m *by vertical parallel projection* onto the lower hemisphere bounded by the great circle l congruent to m. By virtue of this projection, the hemisphere has become a new model of the hyperbolic plane. Every chord g of m is projected into a semicircle v of the sphere meeting l at right angles, so that these semicircles now are to be considered as images of the hyperbolic straight lines. We now map the hemisphere back into the plane by *stereographic* projection. The image of the hemisphere under this projection covers the interior of a circle k, which thus becomes a new model of the hyperbolic plane. Because of the angle-preserving and circle-preserving nature of the stereographic projection, the semicircles v have in this model be- become circular arcs n perpendicular to the circle k. Here and in what follows we have to include the diameters of k as limiting cases in this class of circular arcs.

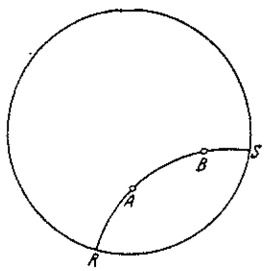


Fig. 247

This new model is due to Poincaré. Let us examine it in a little more detail. From our derivation it follows that there is a one-to-one correspondence between the set of all circular arcs perpendicular to k and the set of all chords of another circle m. Hence any two points A and B in the interior of k can be connected by one and only one circular arc perpendicular to k. If R and S are the points where this arc connecting A and B meets k (Fig. 247), the hyperbolic distance between A and B can be obtained from formula (1) of page 243. For if A', B', R', S' are the points of the original model that give rise to A, B, R, and S under the construction described above, then it can be deduced from theorems of projective geometry that the following relation holds:

$$\frac{AR \cdot BS}{BR \cdot AS} = \sqrt{\frac{A'R' \cdot B'S'}{B'R' \cdot A'S'}}.$$

This gives us the following formula for the hyperbolic distance s of A and B in our new model:

$$s = c\left|\log \frac{AR \cdot BS}{BR \cdot AS}\right|. \tag{2}$$

Now every rigid motion of the hyperbolic plane into itself must be associated with a mapping α of the interior of k into itself transforming into itself the set of circular arcs perpendicular to k. It is plausible, and it is easy to prove rigorously, that this mapping is a circle-preserving transformation, which means that it belongs to the group H considered above. In addition, it can be proved that the group H is even identical with the group of all hyperbolic rigid motions.[1] Being circle-preserving, the transformations belonging to H leave angles invariant; but at the same time they are hyperbolic rigid motions and therefore leave hyperbolic angles invariant. Consequently, the Euclidean angles in Poincaré's model are equal to the hyperbolic angles multiplied by a fixed proportionality factor, and since the angle 2π of a full rotation is obviously reproduced in the hyperbolic plane without change, the factor must be unity. Thus *Poincaré's model preserves angles.*

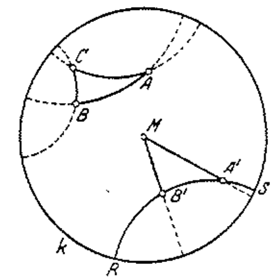


Fig. 248

By means of analytic methods, a formula can be set up by which an angle-preserving mapping can be effected directly from a given portion of a surface of constant negative curvature to a portion of the plane interior to k such that the geodesic lines are mapped into the circular arcs perpendicular to k.

We are now in a position to fill in the proof of the theorem stated on page 246, that in hyperbolic geometry the sum of the angles of any triangle is less than π. We begin with an arbitrary triangle ABC in Poincaré's model of the hyperbolic plane (Fig. 248). We know that the axioms of congruence are valid in the hyperbolic plane.

[1] Here the category of rigid motions is taken to include all maps of the hyperbolic plane that preserve distances, even if they cannot be effected continuously. A simple rigid motion that cannot be effected continuously, is represented by any inversion contained in H; this is a "reflection" of the hyperbolic plane in a straight line. By the remark on page 254, every hyperbolic rigid motion can be expressed as the resultant of at most three reflections.

According to these, we can draw a triangle $A'B'M$ congruent to ABC, in which the point corresponding to C is the center M of k. On page 253 we saw that every circle perpendicular to k that passes through M is bound to degenerate into a diameter of k and that M is exterior to all the other circles that meet k at right angles. In our model, therefore, the hyperbolic straight lines $A'M$ and $B'M$ are represented by Euclidean straight lines, while the hyperbolic straight line $A'B'$ is represented by a circular arc to which M is exterior. The *Euclidean* angles at A' and B' are therefore smaller in the triangle $A'B'M'$ formed by two straight lines and a circular arc than they are in the rectilinear triangle $A'B'M$, and it follows that the sum of the angles in the former triangle falls short of π. Since the model preserves angles, the same is true for the sum of the hyperbolic angles in the hyperbolic triangle $A'B'M$ and in the congruent triangle ABC.

In considering the hyperbolic rigid motions, it is natural to look for discontinuous groups of such motions. In the case of elliptic geometry we saw that the study of this problem boiled down to the study of the regular polyhedra and that there are only a few discontinuous groups in the elliptic plane. In Euclidean geometry it was already more difficult to obtain all the discontinuous groups. In the hyperbolic plane we find that the discontinuous groups are far more numerous than even in the Euclidean plane. All these discontinuous groups of hyperbolic rigid motions are represented in Poincaré's model by groups of circle-preserving transformations contained in H as subgroups.

These groups play a role in the theory of functions. Of special importance among them are the groups of "hyperbolic translations." By hyperbolic translation is meant any hyperbolic rigid motion that can be obtained continuously from the identity and that leaves no point fixed. In plane elliptic geometry there are no rigid motions analogous to this, since every rigid motion in the elliptic plane has a fixed point. In Euclidean geometry, the analogue of the hyperbolic translations are the ordinary translations. But the composition of hyperbolic translations does not follow any such simple law as does the composition of Euclidean translations, since uniqueness of parallels is lacking in the hyperbolic plane.

We shall limit our attention to those discontinuous groups of hyperbolic translations that have closed unit cells. Their Euclidean analogues are the translation groups having parallelograms as unit

cells. In a hyperbolic translation group with closed unit cell, the unit cell can never be a quadrangle. On the other hand, the number of corners of the cell can be any number divisible by four, except four itself. Fig. 249 indicates the tiling of the hyperbolic plane by unit cells in the case where they are octagonal, the hyperbolic plane being represented by Poincaré's model. Of course, we cannot depict the tiling of the whole plane, since the octagons composed of circular arcs get more and more crowded as we approach the boundary of the circle. In our unit cell, as in the case of the fundamental parallelogram of a Euclidean translation group, the sides come in pairs that are equal in length and equivalent; in Fig. 249 this division into pairs is indicated for one of the unit cells. The vertices of all the unit cells drawn in the figure have been numbered, corresponding vertices of different unit cells being identified by the same numbers. It is seen that we come across each of the numbers exactly once in going around any vertex. It follows that the sum of the angles of a unit cell must be 2π. In the representations of all the other groups too, the arrangement of the unit cells is analogous. Hence the sum of the angles of a unit cell is 2π in every case. Furthermore, the sides must be equal in pairs, the arrangement into pairs being made according to a certain rule which we shall not discuss here; in all other respects, the fundamental region may be formed arbitrarily. The fact that the angles always add up to 2π is the reason why the unit cells can never be quadrangles. For, the sum of the angles of a hyperbolic quadrangle is always less than 2π, as is easily seen by dividing the quadrangle into two triangles.

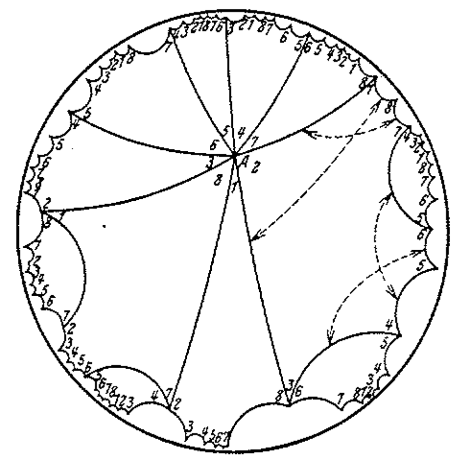
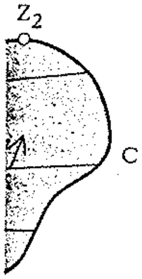


FIG. 249

A far greater variety yet is that of the groups of hyperbolic translations with open unit cell. One of these groups is made use of in the theory of the elliptic modular function.



·or $\overrightarrow{XY}$ is fixed to the
gh full 360^o angle; it

C).

81 ~~105~~

2
Spherical Geometry

2.1 Geodesics

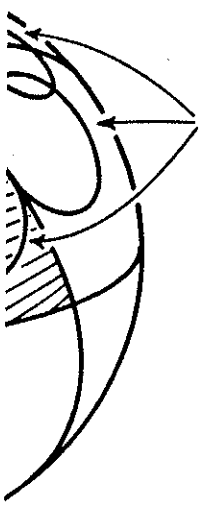
In any ~~geometrical~~ setting where it makes sense to talk about the distance between points, the most important curves are the geodesics.

Definition 2.1.1 Geodesics. The shortest curve connecting two points in a space is a *geodesic* in that space.

Example 2.1.1 A geodesic connecting two points on the globe can be found by stretching a piece of string across the globe between the points and pulling it tight. The geodesic connecting Los Angeles to London passes northeast through central Canada, turns east across the southern tip of Greenland, and arrives in London heading southeast. (Try it!) Ships and airliners save fuel by following such "great circle routes" when traveling long distances (fig. 2.1).

It should be emphasized that *one looks only at curves that lie entirely in the space* when searching for geodesics in a space S. The fact that there may be shorter curves outside of S is irrelevant – one treats S as if it were the entire universe. For instance one could find a shorter path from Los Angeles to London than the one in Example 2.1.1 by burrowing through the earth, but that does not matter since such a path would take one out of the "universe" which, in this case, is the surface of the globe.

There may be more than one geodesic connecting a given pair of points. For example, there are infinitely many geodesics connecting the north and south poles on the globe.



are those circles
pears straight to a
ser circle appears to
.

spherical triangle.
be a geodesic; that
n the sense that a
bending neither to
triangle is thus an

adians, to facilitate
nute you'll see how
t π radians = 180°,
otherwise, we will

94 ~~120~~

henceforth assume that all spheres are unit spheres, i.e. they all have radius one.

Exercise 9.1. For each spherical triangle in Figure 9.2 compute (1) the sum of its angles in radians, and (2) its area. To compute the areas, use the fact that the unit sphere has area 4π. For example, the first triangle shown occupies 1/8 of the sphere, so its area is $(4\pi)/8 = \pi/2$.

Find a formula relating a spherical triangle's angle-sum to its area. This formula appeared in 1629 in the section "De la mesure de la superfice des triangles et polygones sphericques, nouvellement inventee Par Albert Girard" of the book *Invention nouvelle en L'Algebre* by Albert Girard.

You should try to find the formula before reading on, because the following paragraphs give it away. □

Exercise 9.2. What is the area of a spherical triangle whose angles in radians are $\pi/2$, $\pi/3$ and $\pi/4$? What is the area of a spherical triangle with angles of 61°, 62° and 63°? □

Even though there is no overwhelming need for a proof of the formula you just discovered, I would like to include one anyhow because it is so simple and elegant. (It is not, however, the sort of thing you're likely to stumble onto on your own. I struggled for hours without being able to prove the formula at all.)

First we have to know how to compute the area of a "double lune". A **double lune** is a region on a sphere bounded by two

94

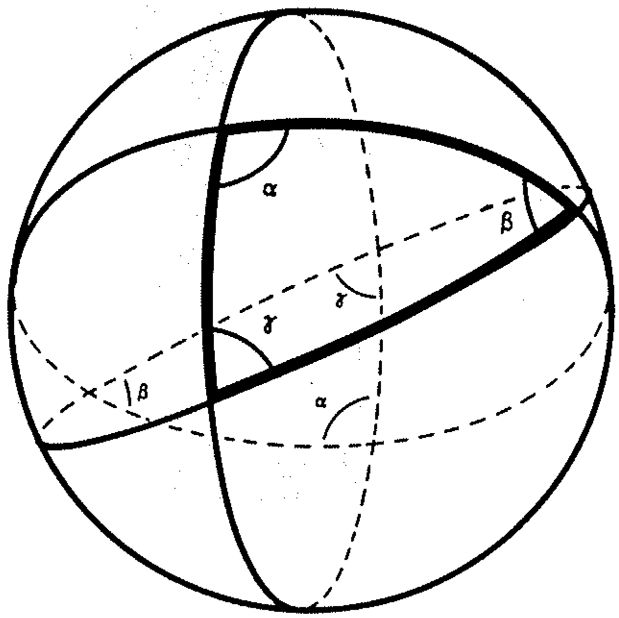


Figure 9.4: Extend the edges of the spherical triangle, and the resulting great circles will form an "antipodal triangle" on the back side of the sphere.

Figure 9.4. An "antipodal triangle", identical to the original, is formed on the back side of the sphere. Figure 9.5 shows three possible ways to shade in double lunes. These double lunes have respective angles α, β and γ, and therefore their areas are 4α, 4β and 4γ.

Now look what happens if we shade in all three double lunes simultaneously (Figure 9.6). All parts of the sphere get shaded in at least once, and the original and antipodal triangles each get shaded in three times (once for each double lune). So. . .

~~123~~

(area of first double lune) + (area of second double lune)

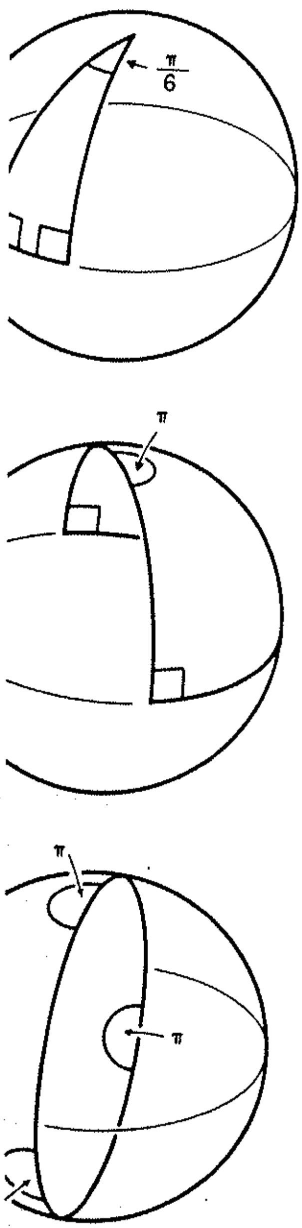
4α + 4

which is just w
says that the su
an amount equal

Exercise 9.3. T
on a sphere of
apply to triangl
three? Write d
radius r. □

Exercise 9.4.
radius is exactl
perfectly straig
carefully meası
area of the fi
(Bonus Questi
That, plus or

~~96~~ 122 93



es. Three of the
each has one or
cupies an entire
same great circle.

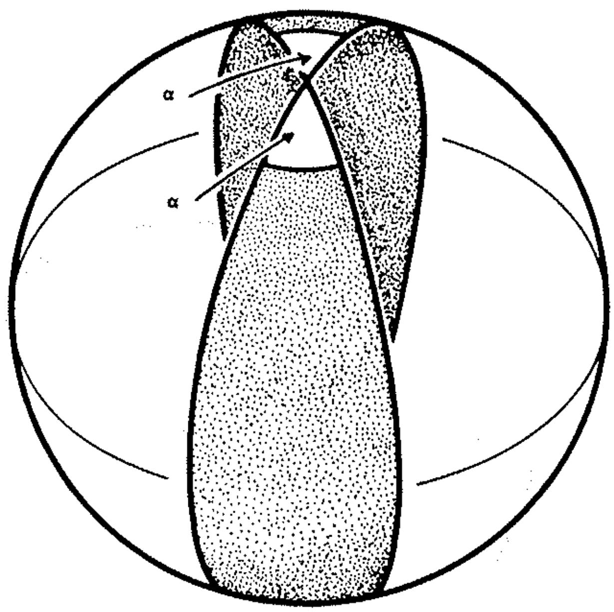
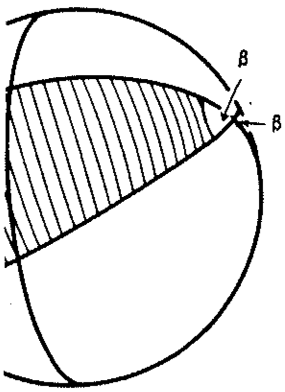


Figure 9.3: A double lune with angle α.

great circles, as shown in Figure 9.3. The largest the angle α can ever be is π, at which point the double lune fills up the entire sphere. So if α is, say, $\pi/3$, then we reason that since $\pi/3$ is 1/3 the greatest possible angle π, the double lune must fill up 1/3 the area of the entire sphere, namely $(1/3)(4\pi) = 4\pi/3$. Using the same reasoning, we get that the area of a double lune with angle α is $(\alpha/\pi)(4\pi) = 4\alpha$. You can check this formula for some special cases, e.g. $\alpha = \pi/2$ or $\alpha = \pi$.

Now we'll find a formula for the area of a spherical triangle with angles α, β and γ. First extend the sides of the triangle all the way around the sphere to form three great circles, as shown in



louble lunes.

on a sphere. They
ıles to be 60.0013°,
0.52 square meters.

ch of the following
estimate of the area

~~4~~ 126 96

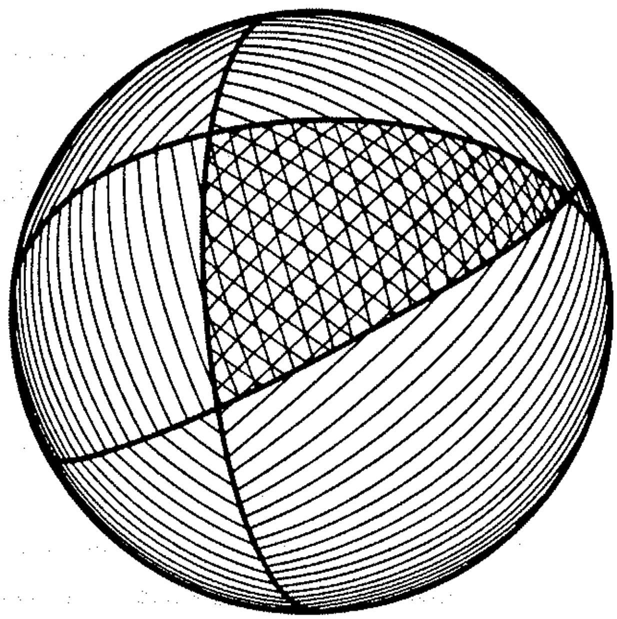


Figure 9.6: Look what happens when we shade in all three double lunes at once.

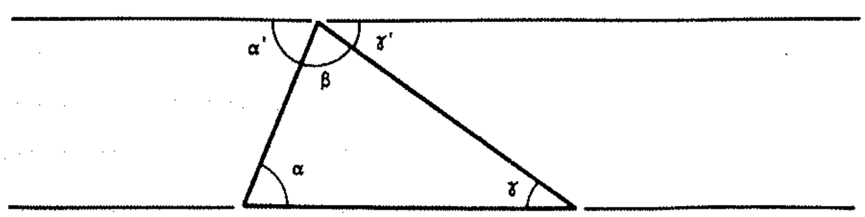
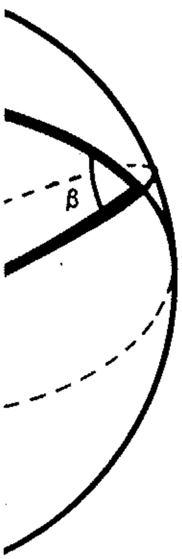


Figure 9.7: A quick proof that the sum of the angles in a Euclidean triangle is π: (1) $\alpha' + \beta + \gamma' = \pi$, (2) $\alpha = \alpha'$ and (3) $\gamma = \gamma'$, therefore (4) $\alpha + \beta + \gamma = \pi$.

908 1 24 95



:al triangle, and the
triangle" on the back

al to the original, is
iure 9.5 shows three
se double lunes have
heir areas are 4α, 4β

l three double lunes
sphere get shaded in
l triangles each get
e). So. . .

$$\begin{pmatrix}\text{area of}\\ \text{first}\\ \text{double}\\ \text{lune}\end{pmatrix} + \begin{pmatrix}\text{area of}\\ \text{second}\\ \text{double}\\ \text{lune}\end{pmatrix} + \begin{pmatrix}\text{area of}\\ \text{third}\\ \text{double}\\ \text{lune}\end{pmatrix} = \overset{\begin{bmatrix}\text{everything}\\ \text{was shaded}\\ \text{in once}\end{bmatrix}}{\begin{pmatrix}\text{area of}\\ \text{entire}\\ \text{sphere}\end{pmatrix}} \overbrace{+2\begin{pmatrix}\text{area of}\\ \text{original}\\ \text{triangle}\end{pmatrix} +2\begin{pmatrix}\text{area of}\\ \text{antipodal}\\ \text{triangle}\end{pmatrix}}^{\begin{bmatrix}\text{each triangle}\\ \text{was shaded}\\ \text{in two more}\\ \text{times}\end{bmatrix}}$$

$$4\alpha \quad + \quad 4\beta \quad + \quad 4\gamma \quad = \quad 4\pi \quad + \quad 2A \quad + \quad 2A$$

$$4(\alpha + \beta + \gamma) = 4(\pi + A)$$

$$\alpha + \beta + \gamma = \pi + A$$

$$(\alpha + \beta + \gamma) - \pi = A$$

which is just what we wanted to prove! In words, this formula says that the sum of the angles of a spherical triangle exceeds π by an amount equal to the triangle's area.

Exercise 9.3. The formula $(\alpha+\beta+\gamma) - \pi = A$ applies only to triangles on a sphere of radius one. How must you modify the formula to apply to triangles on a sphere of radius two? What about radius three? Write down a general formula for triangles on a sphere of radius r. □

Exercise 9.4. A society of Flatlanders lives on a sphere whose radius is exactly 1000 meters. A farmer has a triangular field with perfectly straight (i.e. geodesic) sides and angles which have been carefully measured as 43.624°, 85.123° and 51.270°. What is the area of the field? Don't forget to convert the angles to radians. (Bonus Question: How accurately do you know the field's area? That, plus or minus what percent?) □

from Weeks. 'Shape of Space' 9?

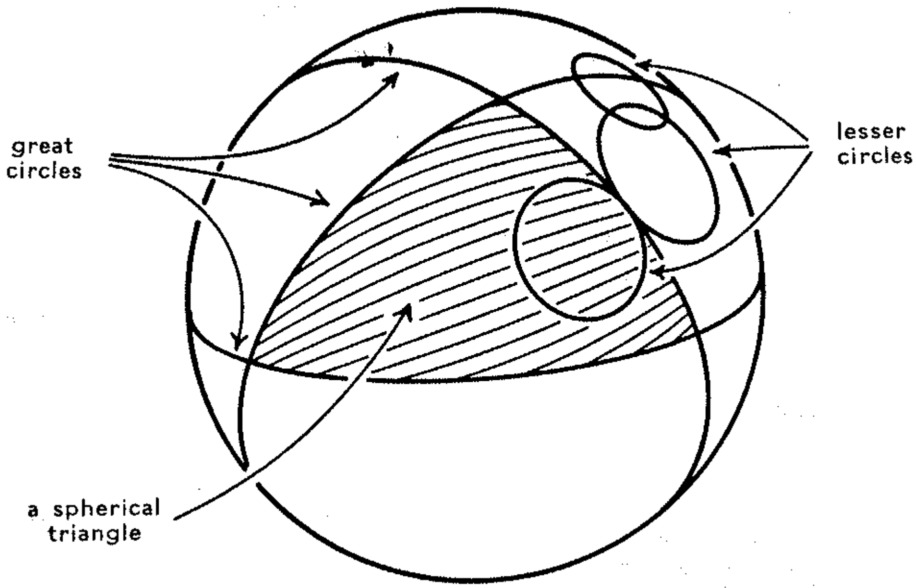


Figure 9.1: On any sphere, the **great circles** are those circles which are as big as possible. A great circle appears straight to a Flatlander on the sphere. By contrast, any lesser circle appears to bend to one side or the other.

A triangle drawn on a sphere is called a **spherical triangle.** Each side of a spherical triangle is required to be a geodesic; that is, it is required to be intrinsically straight in the sense that a Flatlander on the sphere would perceive it as bending neither to the left nor to the right. A side of a spherical triangle is thus an arc of a so-called great circle (see Figure 9.1).

From now on we will *measure all angles in radians*, to facilitate easier comparison of angles and areas (in a minute you'll see how and why we want to do this). Recall that π radians = 180°, $\pi/2$ radians = 90°, etc. Except when specified otherwise, we will a result would enable two-dimensional beings who lived inside the surface to discover whether it is saddle- or bowl- shaped, without ever going outside the surface, by adding up the angles of a geodesic triangle and comparing the sum with 180^o. A similar procedure would enable three-dimensional beings such as ourselves to discover whether or not their universe is curved without having to leave their universe to make the measurement. (Our own universe *is* curved, but you have to look at very large triangles to detect the curvature.)

The famous *Gauss-Bonnet Theorem* says that the conjectures in the previous paragraph are true, at least on surfaces that are sufficiently smooth; it is proved in courses in differential geometry.[1] We shall content ourselves with proving it in the special case where the surface is a sphere (Theorem 2.3.1 on page 51).

Exercise 2.1.1 Prove Equation 2.1 for surfaces built from sectors subtending an arbitrary angle $0^o < \theta < 540^o$. Show that if $\theta \geq 540^o$ then no geodesic triangle in the surface has the vertex of the sector in its interior. What is the sum of the angles of a geodesic triangle on these surfaces if the vertex of the surface is not in its interior?

2.2 Geodesics on Spheres

Definition 2.2.1 A *great circle* is the intersection of a sphere and a plane that passes through the center of the sphere. All other circles on the sphere are "small circles" (Fig. 2.5).

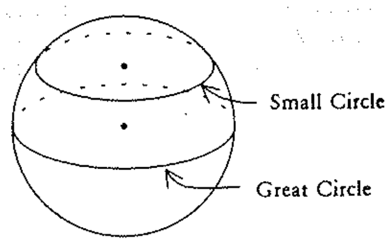


FIGURE 2.5.

Let A and B be two points on the sphere and let C be its center. The arc $\overparen{AB}$ subtended by the $\angle ACB$ is a segment of a great circle with length

$$\text{length}(\overparen{AB}) = R(\angle ACB)$$

where R is the radius of the sphere and $\angle ACB$ is measured in radians.

[1] See [17, Chap. 7 §8].

Sample solutions to Levi problems 1 & 2

1. ✓ 159 ~~126~~

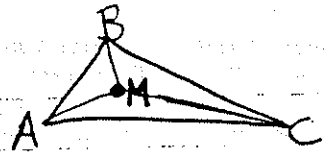
Daniel Perez
Good write up

H.W. #1

Math 128A

9/27/99

D. Perez 1)



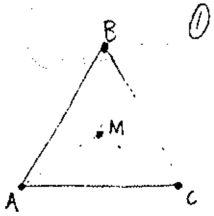
We are given triangle ABC w/ pt. M lying inside the triangle and are asked to prove that $\overline{MA}+\overline{MB}+\overline{MC} > \frac{1}{2}(\overline{BA}+\overline{AC}+\overline{CB})$, where $\overline{MA}$, $\overline{MB}$ & $\overline{MC}$ are constructed uniquely since any two pts. det. a line segment. First of all $\overline{MA}+\overline{MB} > \overline{BA}$, $\overline{MB}+\overline{MC} > \overline{BC}$, & $\overline{MC}+\overline{MA} > \overline{AC}$ since any two sides of a triangle are greater in length as a sum than the third side. Now by simple addition of line segments we find that $2\overline{MA}+2\overline{MB}+2\overline{MC} > (\overline{AB}+\overline{BC}+\overline{AC})$ and finally through division we get $\overline{MA}+\overline{MB}+\overline{MC} > \frac{1}{2}(\overline{AB}+\overline{BC}+\overline{AC})$. So the sum of the distances from a pt. inside of a triangle to the vertices of the triangle is greater than half the triangle's perimeter.

sharp Δ inequality

1. ✓ , 2 ✓

Good. writing

Fall 1999 Math 128A   Levi Problems 1 & 2   Maurice Turrieta 09/28/



① We are given ΔABC and a point within this triangle referred to as point M. We are asked to prove that $AM+BM+CM > \frac{1}{2}(AB+BC+AC)$. According to the triangle inequality, the following must be true: $AM+BM \geq AB$, $BM+CM \geq BC$, $CM+AM \geq AC$. Since point M forms ΔABM, ΔBCM, and ΔACM with the sides of ΔABC, it follows that $AM+BM > AB$, $BM+CM > BC$ and $CM+AM > AC$. Adding the three inequalities together we arrive at the inequality $2AM+2BM+2CM > AB+BC+AC$. Using the basic rules of arithmetic, we conclude that $AM+BM+CM > \frac{1}{2}(AB+BC+AC)$.

(This is the strict Δ ineq.)

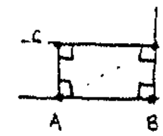
4+. 160 Sample Solution

Problem 1.
To prove: The perpendicular bisector of segment AB is the set of points X satisfying AX = XB.

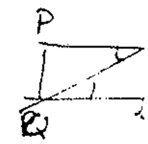
Noah Constant
math 128, hw4

1. Given a line through points A and B, we want to show that set of points equidistant from the line on one side is a parallel line. Construct the perpendicular to AB through A, and the perpendicular through B. Construct a circle about A, with radius equal to the distance chosen for the equidistant set. Mark where the circle meets the perpendicular through A, point C. Construct the perpendicular to AC, through C. Where AC meets the perpendicular to AB through B, mark point D.

By the parallel postulate, AB and CD are parallel, because the interior angles made by a line cutting across them sum to pi. By construction, C is distance AC from AB.



Now, pick an arbitrary point E. If E is on CD, then by the parallel postulate, a perpendicular to CD through E, is also perpendicular to AB. The distance between E and AB, along this perpendicular is the same as the distance AC, because they are opposite sides of a rectangle (to verify this, cut diagonal AE, and show triangles ACE and EBA are congruent by ASA, where the angles are the same because they are (alternate) interior angles of the diagonal cutting through parallels). If E is not on CD, drop a perpendicular from E to AB. Because AB and CD are parallel, the perpendicular must meet CD (mark point F). Because F is distance AC from AB, E must be either greater or less than this distance (unless it is exactly distance AC on the other side of the line. There are only two points on a line, a non-zero positive distance from a point on that line).

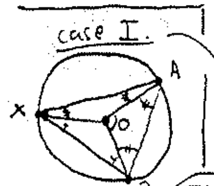


2. Note, for this problem AOX is shorthand for angle(AOX), unless specified.

a. Because the radii are equal, AOX and BOX are isosceles triangles with equal sides meeting at O. The base angles of isosceles triangles are equal, so OAX = OXA, and OBX = OXB. Because the sum of the interior angles of a triangle is pi, AOX = pi - 2 * OAX, and BOX = pi - 2 * OBX.

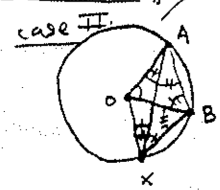
Now we break into two cases, depending on whether or not one of AX and BX intersects one of AO and BO. (in particular, AO will intersect BX, or BO will intersect AX, or neither).

case I.



Case I. The three angles AOX, BOX, and AOB sum to 2pi. Substituting earlier values for AOX and BOX, we get AOB = 2 * OAX + 2 * OBX. Because AXB = OAX + OBX, we see that AOB = 2 * AXB.

case II.



Case II. In this case, ABX equals either AXO - BXO or BXO - AXO. For this example, I am assuming it equals AXO - BXO. Because A and B are arbitrary, the names can be switched if this comes out wrong. AOB + BOX = AOX, so substituting with the valued we computed for BOX and AOX, we get AOB = 2 * OAX - 2 * OBX. Because AXB = AOX - BOX, we see that AOB = 2 * AXB.

b. By 2a, AXB = 1/2 AOB = AYB, so AXB = AYB.

c. If AB is a diameter, AOB = pi, so by 2a, AXB = pi/2.

3. If X is on the lesser of the two arcs AB, the first section of 2a still holds. By interior angles of triangles, AOB + OAB + OBA = pi = OXA + OXB + (OAX - OAB) + (OBX - OBA). Now substituting earlier values, after some algebra, we get AOB = pi - 2 * OAB, and AXB = (pi - 2 * OAB) / 2. So we conclude that in this case, AXB = 2 * AOB.

4. By 2c we know that all of C is in the locus of points X such that triangle(AXB) is right, with hypotenuse AB. Now we want to show that any point not on C is not in the locus.

Problem 2 with A, X, B on a circle with center O indicated, prove that

$\measuredangle AXB = \frac{1}{2} \measuredangle AOB$.

~~#49~~ 161

from Levi's class.

Collected problems. For practice.

Problem 1 A point M lies inside the triangle ABC. Prove that the sum of the distances from M to the vertices of the triangle is greater than half the triangle's perimeter.

Problem 2 Two points A and B in the plane are given. What is the set of all points M in the plane for which:

$$(a)\ |MA| < |MB|;\ (b)\ |MA| > |MB|;\ (c)\ |MA| = |MB|?$$

Problem 3 Prove that the sum of the diagonals of a convex quadrangle is greater than its half-perimeter but is less than the perimeter.

Problem 4 Prove that the largest possible area of a triangle inscribed in the circle of radius r is $\frac{3r^2\sqrt{3}}{4}$.

Problem 5 Prove that the largest of the areas of quadrangles inscribed in the circle of radius r equals $2r^2$.

Problem 6 Divide a given segment in the ratio of $1 : \sqrt{2}$. Hint: The diagonal of the unit square is $\sqrt{2}$.

Problem 7 Draw a line parallel to the base of a triangle so as to divide the area of the triangle in half.

Problem 8 Given two segments a and b. Construct the segment $\sqrt{ab}$.

Problem 9 Construct the square whose area equals the area of a given triangle.

Problem 10 Prove that if two similar polygons are related by the dilation factor k, then their areas are related by the factor k^2. In other words, if a figure is blown up k times, the area is enlarged k^2 times. State and prove the analogous fact for solids.

Problem 11 Prove that the area of a circle whose diameter is the hypotenuse of a right trangle equals the sum of areas of two circles whose diameters are the respective legs of the same triangle.

162 ~~150~~

Problem 12 In the plane two points A and B are given. What is the set of all projections of A onto straight lines passing through B?

Problem 13 What is the set of all congruent chords of a given circle?

Problem 14 What is the set of midpoints of all chords passing through a given point of a given circle?

Problem 15 Prove that the lines of the sides of the trapezoid and the line passing through the midpoints of the parallel sides of the trapezoid all intersect at one point.

Problem 16 A quadrangle $ABCD$ is called a "kite" if $AB = BC$ and $AD = DC$. Find the area of ~~the deltoid~~ a Kite in terms of the lengths d_1 and d_2 of its diagonals.

Problem 17 A triangle's sides are 3 and 4 in. What can its area be?

Problem 18 Construct the tangent to a given circle through a given point (lying outside the circle).

Problem 19 Two perpendicular chords are drawn from a point on a circle. The distance between the midpoints of the chords is a. Find the diameter of the circle.

Problem 20 Find the radius of a circle given that the distances from ends of a diameter to a tangent are a and b.

Problem 21 The base of an equilateral triangle coincides with a diameter of a circle. Find the angular measures of the arcs into which the circle is cut by the sides of the triangle.

Problem 22 Given segments a, b, c, construct the segment $x = \frac{ab}{c}$.

Problem 23 *State and prove the Pythagorean theorem.*

Problem 24 State and prove the theorem of cosines (using the Pythagorean theorem).

Problem 25 State and prove the law of sines.

163 15T

Problem 26 Prove the theorem: If two sides of one triangle are congruent to the corresponding two sides of another triangle and the angles lying between these sides are not equal, then opposite the larger angle lies the longer side.

Problem 27 Diagonal of a rectangle is m; the angle between the diagonals is α. For which α will the area of the rectangle be largest possible? The value m is fixed.

Problem 28 *Find the area of a triangle whose sides are a, b and c.*

Problem 29 *Construct a circle tangent to three straight lines, given that among the three none are parallel and that the lines do not pass through the same point.*

Problem 30 *Inside a given circle, inscribe the triangle similar to a given triangle.*

Problem 31 *Prove that the area of a circumscribed triangle equals half the product of the perimeter and the radius of the inscribed circle.*

Problem 32 *If the sum of opposite angles of a quadrangle is 180 degrees, then around such quadrangle there exists a circumscribed circle.*

Problem 33 *Prove that any rectangle with a symmetry axis not passing through its vertex admits a circuscribed circle.*

Problem 34 A triangle's sides are 3 and 4 in. What can its area be?

Problem 35 Construct the tangent to a given circle through a given point (lying outside the circle).

1 **Problem 36** Two perpendicular chords are drawn from a point on a circle. The distance between the midpoints of the chords is a. Find the diameter of the circle.

Problem 37 Find the radius of a circle given that the distances from ends of a diameter to a tangent are a and b.

Problem 38 The base of an equilateral triangle coincides with a diameter of a circle. Find the angular measures of the arcs into which the circle is cut by the sides of the triangle.

164 LS2

Problem 39 Prove that the three altitudes of a triangle meet at one point.

Problem 40 Find $(1+i)^{100}$.

Problem 41 The plane was rotated through $30°$ around the point $(3,6)$, and dilated by the factor of 3 with the center of dilation at the origin. Where did the point with coordinates $x = 1$, $y = 2$ end up?

Problem 42 Prove that the image of a circle under the inversion is a circle.

Problem 43 Two vertices of a triangle lie in the plane. Does the third vertex lie in the plane given that in the plane lies 1) the incircle; 2) the circumcircle.

Problem 44 Expain why the three legged stool is stable, while the four-legged one is not.

Problem 45 Given: $a \cap b = C$, $b \cap c = A$, $c \cap a = B$, $A \neq B$, $A_1 \in a$, $B_1 \in b$, $C_1 \in (A_1B_1)$. Prove: $C_1 \in (ABC)$.

Problem 46 Given: $a \cap b = A$, $a \subset \alpha$. True or false: 1) $A \in \alpha$; 2) $b \subset \alpha$?

Problem 47 Given a tetrahedron $ABCD$ and points M and N where $M \in [DC]$, $N \in [AB]$. Construct the intersection of the planes ABM and DCN.

Problem 48 Construct the section of the tetrahedron $ABCD$ with the plane passing throught the edge DC and the point of intersection of the medians of the face ACB. Find the area of the section if each edge of the tetrahedron is of length a.

Problem 49 Given a tetrahedron $ABCD$, constrict its section by the plane passing through the median DD_1 of the face DBC and the point M which lies on the face ADC but not on any of the edges of the tetrahedron.

Problem 50 Given two crossing lines m and n. Through the point P not lying on either of these lines, construct the line intersecting both given lines.

Problem 51 Four points: A, B, C and D don't lie in one plane. Prove that the midpoints of the segments AB, BC, CD and DA lie in the same plane. Identify the figure whose vertices are these midpoints.

Problem 52 Determine the union of all lines intersecting two crossing lines.

In a trihedral angle two acute angles are congruent. Prove that the projection of their shared edge onto the face of the other angle is the bisector of that angle.

Problem 53 In a tetrahedron $ABCD$ the edge DC is perpendicular to the plane ABC and the angle between the edge BD and the plane ABC is $60°$. The face ABC is an equilateral triangle with sides a. Find the area of the face ABD and the measure of the dihedral angle AB.

Problem 54 From a point on a face of an acute dihedral angle two lines are drawn, one of which is perpendicular to the edge and the other intersecting the edge at an angle different from $90°$. Prove that the angle between the perpendicular and the second face of the angle is greater than the angle between the skew line and the second face.

Problem 55 The slope of a hill is $30°$. Find the angle ϕ between the line of steepest descent and the road whose slope is $15°$.

Problem 56 Construct the section of the cube $ABCDA_1B_1C_1D_1$ with the plane passing through given points on the edges AB , BC and B_1C_1. A neat drawing is essential!

Problem 57 State and prove the optical property of the ellipse, of the parabola and the hyperbola.

Problem 58 Prove that the intersection of a circular cone with a plane is an ellipse if that intersection is a bounded curve.

Problem 59 Three mirrors form a trihedral angle in space, with each planar angle being $90°$. Prove that the ray of light sent into the corner so that it reflects off of each face will, after the three reflections, be parallel to its incoming direction. Note: this is how the bike and car night reflectors work: they send the rays of light back to their source, regardless of where the source is!

Problem 60 Consider an ellipse e with the foci F_1 and F_2. Let $P \in e$ be an arbitrary point on e, and let MN be the tangent line to e at P. Let F_2' be the mirror image of F_2 with respect to MN. Prove: as P sweeps out the ellipse e, the point F_2' sweeps out a circle.

Problem 61 Prove that curve given by the equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ in the plane is an ellipse. In other words, show that the sum of distances from each point P on the curve to certain two points is the same for all points P.

Problem 62 *Prove that the angle subtended by the diameter in a circle is a right angle.*

Solution. Consider an arbitrary diameter AB of the circle and an arbitrary point P on the circle ($P \neq A$ or B). We must prove that $\angle P = 90°$.

Let O be the center of the circle. The triangle POB is isoceles, and thus $\angle OPB = \alpha$, where $\alpha = \angle OBP$. Thus $\angle AOP = 2\alpha$ as the external angle (which must equal to the sum of the two non-adjacent angles). Since the triangle AOP is isoceles (two of its sides are the radii), $\angle APO = \frac{1}{2}(180° - 2\alpha)$. We have $\angle APB = \angle APO + \angle OPA = \frac{1}{2}(180° - 2\alpha) + \alpha = 90°$,
Q. E. D.

Problem 63 *With a compass and a straightedge, construct the line parallel to a given line and passing thought a given point.*

Solution. Denote the line by MN and the point by P. Place one needle of the compass on P and another on an arbitrary point A of MN. Draw the circle centered at P, and let A and B be the points of intersection of the circle with MN. On MN mark the point $C \neq A$ with $AP = BP$. Draw the circle of radius $BC = AB$ centered at B and another circle of the same radius centered at C. Let Q be the point of intersection of these two circles which lies in the same half-plane as P. CLAIM: the line PQ is parallel to MN.

Proof of the claim. It suffices to prove that the interior alternate angles for the lines PQ and AC are congruent: $\angle QPB = \angle PBA$. To that end we show that $\Delta ABP = \Delta PQB$. Both are isoceles with the same sides, and it remains to prove that $\Delta ABP = \Delta PQB$. Since $\Delta ABP = \Delta BCQ$ (all respective sides equal by construction), $\angle ABP = \angle CBQ \equiv \alpha$ and thus $\angle PBQ = 180 \circ -2\alpha$. But $\angle APB = 180 \circ -2\alpha$, and thus $\angle APB = \angle PBQ$ proving thus $\Delta ABP = \Delta PQB$.
Q. E. D.

Problem 64 Prove that the angle formed by a diameter and a chord meeting at a point on a circle equals half the angular measure of the subtending arc.

Solution. Let AB be a diameter and AC be a chord. Let O be the center of the circle and consider triangle AOC.

$$\angle COB = \angle CAO + \arg ACO \tag{1}$$

as the exterior angle. $\angle CAO = \arg ACO$ as the angles at the base of the isosceles triangle AOC (two of whose sides are the radii). From (1) we get $\angle COB = 2\angle CAO$, Q. E. D.

Problem 65 With a compass and a straightedge, divide a given segment in the ratio of $1 : \sqrt{5}$.