Ahlfors meta review PhL 8.21.09

Comments on the book: I used it in 1972 (2nd year grad student) in Math 185 at Berkeley taught by Behrens. In that semester course we only did Chapters 2 and 4. During S-matrix days, I must have read Chapter 3 about mappings and branch points and Riemann sheets.

The author, Lars Ahlfors (1907-1996), was Finnish, did U of Helsinki for BA and PhD (1930), taught there after 3 years somewhere else, then got the Fields medal in 1936 at age 29. Prof at UH in 1938. Then WWII came along, he did not like Suisse, then Harvard snapped him up in 1945 and HU was home till 1977. This book Complex Analysis 1953, 1966, 1979 served many people. He has a separate book on Riemann Surfaces and that was an area of his research.

I have the 1966 2nd edition. In the raw notes, one sees that very few changes were made going to the third and last edition in 1979. Sticker on 3rd Ed is a whopping $150, all reviewers complain!

Ahlfors seemed intent on saying as much as possible in as few pages as possible, so figures are omitted, explanations very terse, the book is for the "advanced" reader. Minimal space is given to text organizational tags -- topics and examples just tumble one after another with no break in the text. Somewhat like lecture notes perhaps.

My motivation for looking at this book in Aug 2009 was simply that Stakgold got to the point of conformal mapping in his potential theory chapter, so that was a good excuse to do this "back and fill" pass through at least the first 3 chapters of Ahlfors. I am glad I did this, in retrospect.

CHAPTER 1. THE BASICS OF COMPLEX NUMBERS 2

CHAPTER 2. COMPLEX FUNCTIONS (21) 3
1. Introduction to the concept of "analytic function". 3
1.1. Limits and Continuity (22). 3
1.2 Analytic Functions (24). 3
1.2 Polynomials (28). 4
1.3 Rational Functions (30). 4
2. Elementary theory of power series (33) 4
2.1 Sequences 4
2.2 Series 4
2.3 Uniform Convergence 5
2.4 Power Series (38) 5
2.5 Abel's Limit Theorem (42) 6
3. The Exponential and Trigonometric Functions (43) 6
3.1 The Exponential. 6
3.2 The trig functions (44) 7
3.3. The Periodicity (45) 7
3.4 The Logarithm (46) 7

CHAPTER 3. ANALYTIC FUNCTIONS AS MAPPINGS (49) 7
1. Elementary Point Set Topology. 7
1.1 Sets and Elements (50) 7
1.2 Metric spaces (51). 7
1.3 Connectedness (54). 8
1.4 Compactness (59) 8

1.5 Continuous Functions (64). 9
1.6 Topological spaces (67). 10
2. Conformality (68) 11
2.1 Arcs and closed curves (68). 11
2.2 Analytic functions in regions (69). 11
2.3 Conformal Mapping (73). 12
3. Linear Transformations (76) 13
3.1 The linear group (76). 13
(a) transformations on the projective line 13
(b) transformations on the projective plane and beyond 15
(c) notes on this book section 16
3.2 The Cross Ratio (78) 16
Theorem 12: the cross ratio is invariant under linear transformations. 17
Question: How would you find a linear transform that takes z_1, z_2, z_3 to w_1, w_2, w_3 ? 18
Theorem 13. Claim that (z_1,z_2,z_3,z_4) is real $\Leftrightarrow$ the four points z_i lie on a Circle 20
Theorem 14: Linear transformation maps Circles into Circles. 20
3.3 Symmetry (80) 20
3.4 Oriented Circles (83) 22
3.4 Families of Circles (84) 24
4. Elementary Conformal Mapping (89) 26
4.2 A survey of elementary mappings (93) 27
4.3 Elementary Riemann Surfaces (97) 28

Chapter 1. The Basics of Complex Numbers

He has a section on inequalities which got me to write a separate "absolute value.doc" where I list off various complex and real-only inequalities. Here are some of these inequalities where a_i and b_i are arbitrary complex numbers,

$| \Sigma_i a_i| \leq \Sigma_i | a_i|$ Triangle officially when has sum has 2 terms, but OK

$| \Sigma_i a_i b_i | \leq |\Sigma_i a_i| | \Sigma_i b_i | \leq (\Sigma_i | a_i|) (\Sigma_i | b_i|)$
CSI Triangle

He introduces the graphic z plane, but does not call it an Argand diagram. Ahlfors holds off on using phasor notation because we don't have an expo function yet. He is trying to be axiomatic in this chapter. One can prove, without knowledge of the expo function, this theorem: (A calls it the binomial equation)

$(\cos\varphi + i \sin\varphi)^n = \cos(n\varphi) + i\sin(n\varphi)$ de Moivre's formula 1707/1722 (1667-1754)
(Euler expo formula was 1749)

Exercise 2 page 17 asks us to express a parabola in complex notation, see later in the raw notes, but I am not happy with my conclusion. This subject is called **analytic geometry** (the study of 2D curves and the algebraic equations that go with them, would be called algebraic geometry in modern use.)

We then get the **Riemann sphere** and its various equations and regularization of ∞. Circle on the sphere maps into circle or line on the plane. You could use this method to make a map of the northern hemisphere, but it would put the north pole area out in the perimeter of the map, and you need an infinite planar map. This is not what we usually see!

Chapter 2. Complex Functions (21)

1. Introduction to the concept of "analytic function".

1.1. Limits and Continuity (22). The derivative is a limit, so limits get comments here, and limits are used to talk about continuity. Remember that "the limit" is an essential element of "calculus".

1.2 Analytic Functions (24). An analytic (= holomorphic) function f(z) is one for which the derivative function f '(z) is well defined, finite, and the same regardless in what direction in the C plane you take your differential limit. That is, the derivative f '(z) = lim [f(z+Δz)-f(z)]Δz is the same for any complex Δz. This suggests that f(z) should be defined on an open set so there is room to consider all such approaches. You can write f '(z) = $\hat{\mathbf{n}}\bullet\nabla_2 f(x,y)$ where $\hat{\mathbf{n}}$ can be in any direction. This means f(z) is "continuous in all directions of approach". For real functions, this means only the left and right directions. An analytic function is continuous in this sense, trivial to show. The requirement of analyticity is pretty stringent, requiring this f '(z) to exist and be the same in all 2D directions.

By considering the approach in real and imaginary directions, one can show some amazing facts about the real and imaginary parts of f(z) = u(z) + iv(z).

First are the **Cauchy Riemann equations**

$\partial_x u = \partial_y v$ and $\partial_y u = -\partial_x v$ u and v are "harmonic conjugates"

$\partial_x u = \partial_y v$ and $\partial_y u = -\partial_x v$ $\Leftrightarrow$ f(z) is analytic

Given u, it is possible to find v or f and vice versa, several methods are given. One way involves thinking of $f(z,\bar{z}) = f(z)$ "only" if f(z) is an analytic function. f(z) = 2u(x/2,z/2i) - u(0,0) is an example p 27. Another way is to just integrate the C-R equations.

Second, it is trivial to show that $\nabla^2 u = 0$ and same for v, meaning u and v are **harmonic functions** meaning they must satisfy Laplace's equation! But then f(z) itself must also be **harmonic**! It is pretty clear in polar notation that $f(z) = z^n = r^n e^{in\theta}$ satisfies Laplace's equation in polar coordinates (Stakgold II p 93), and so does $\mathrm{Re}f(z) = r^n\cos(n\theta)$ etc. Easy to show $\nabla^2 e^z = 0$. So, an analytic function is one that solves the Laplace equation $\nabla^2 f(z) = 0$ in 2D space. This still seems amazing to me. Of course if you can write a convergent power series for some f(z), then it is essentially a polynomial and f(z) solves Laplace.

Question: consider a Legendre polynomial. It satisfies the Legendre ODE meaning in a single variable z. But it must also satisfy the Laplace PDE which involves ∇^2 in x,y where z = x+iy. Thus, we know that $\nabla^2 P_\ell(z) = 0$.

1.2 Polynomials (28). The sum or product of two analytic functions is analytic. Starting with 1 and z, you quickly conclude that a polynomial is analytic. He discusses the "fundy theorem of algebra" which says a poly has at least one root (Argand proved this first!), and he shows how this idea used repeatedly says that you can factor any polynomial to expose its n roots, some of which may be the same. Using this stuff he proves:

Theorem 1 (Lucas's Theorem) the zeros of P'(z) are in the same half-plane as the zeros of P(z).

In fact, the zeros of P'(z) are in the same convex polygon as those of P(z). This fact was never used in Chapters 2 or 3, but may come up later in the book.

1.3 Rational Functions (30).

The slight misnomer "rational function" refers to the ratio of two polynomials $R = P_n/Q_m$ where letters show degree of each polynomial. By factoring, you can at once see that R has n finite zeros and m finite poles. What you don't see are the zeros and poles of R at $z = \infty$. The rule is this: if you include the point ∞, the rational function R has # zeros = # poles = max(m,n) ≡ p = "the order of R". R and R-a have the same order. Rational functions of order p = 1 are called linear transformations, another misnomer. These are the R(z) = (az+b)/(cz+d) things we shall study a lot. Easy to solve R(z) = w for z = lin trans(w).

A spends a lot of time showing you can write R(z) as a sum of functions $R = G + \Sigma_i G_i$ where function G_i contains the pole of R at $z = \beta_i$ (including $\beta_i = \infty$) and where G has no poles. Thus you can *isolate* the poles into separate terms (which may contain a sum of poles of different orders at some β_i.) I think concept will be used in the integration Chapter 4.

2. Elementary theory of power series (33)

2.1 Sequences

A major **theorem** says this: (think of this as a "proofing tool" -- a tool you can use to prove things)

(a sequence converges) $\Leftrightarrow$ (sequence is a Cauchy sequence) " Cauchy's Condition"
$d(x_n,x) \to 0 \quad \Leftrightarrow \quad d(x_n,x_m) \to 0$

The $\Rightarrow$ direction (necessity) is easy to prove, but the $\Leftarrow$ direction (sufficiency) is harder to prove, and A spends 1.5 pages on this, getting us all tangled up in limes superior stuff about which I did a separate doc. This was not such a big deal in Stakgold where complete vector space allowed a general proof.

2.2 Series

Consider $s_n = a_1 + a_2 + \ldots a_n$ which is the famous "**partial sum sequence**". If s_n converges, we know

$d(s_n, s_m) \to 0$ by the above theorem, which means all "tail sequences" $\to 0$, including tails of length 1, from which we conclude that $a_n \to 0$. A major proofing tool is the idea of comparing two partial sum sequences. If $d(b_m, b_n) \leq d(a_m, a_n)$, then the b thing converges if the a thing does, and b is a **contraction** of a. Application: if $|a_k|$ converges (**absolute convergence**) so does a_k.

2.3 Uniform Convergence

A then extends the notion of convergence from a_k to $a_k(x)$ where x is in some set E. This is of course the idea of "uniform convergence" and he presents a set of 4 small theorems on this subject:

Theorem 1: If $f_n(x)$ converges uniformly to f(x), then if the $f_n(x)$ are continuous, so is f(x).

Theorem 2: Cauchy uniform convergence $\Leftrightarrow$ regular uniform convergence

Theorem 3: If $f_n(x)$ is contraction of a_n for all x in E, and if $a_n \to a$, then $f_n(x) \to$ some f(x).

Theorem 4: **Weierstrass M Test** : Find M and a_k such that $|f_k(x)| \leq M\, a_k$. If $a_k \to a$, then $f_k(x) \to f(x)$. The series compared here are a_k = the **majorant**, and $f_k(x)$ = the **minorant**. Notice that this is more general than comparing two series in R, which was the contraction thing above. Here we are comparing the norm of a sequence of vectors to a sequence in R. I suppose you might also consider comparing the norms of two sequences of vectors, but he does not do this here.

Again, these are more proofing tools. A likes proofs and rolls these things out when needed.

2.4 Power Series (38)

Here we have a blockbuster multi-part theorem which applies to any series with *constant complex coefficients*. He states the theorem then proves all or most of its claims. I have reviewed the proofs in the raw notes, and here just state the claims (in my own terms): [Abel was Norwegian, died at age 27 of TB]

Abel's Theorem on Series Convergence which has several parts:

There exists a radius of convergence R for |z|. R could be 0 or finite or ∞. In other words, if there is a region of convergence, that region is a *disk*, not some region of some other shape. Many things are true:

(1a) for $|z| < R$, the series converges.
(1b) it also converges absolutely since we can get another series with the same radius (as we shall see) by just changing the sign of all negative coefficients.
(1c) Convergence is uniform for $|z| \leq \rho < R$

(2a) If $|z| > R$, the series diverges.
(2b) The terms of the series keep growing so that $|a_n z^n| >$ any C for n large enough. Terms are unbounded. There is no bound C such that $|a_n z^n| \leq C$ for all n.

(3a) For $|z| < R$, the sum of the series not only converges, but is in fact an *analytic function*,

(3b) You can compute the derivative of the series termwise.
(3c) this derivative series has the same radius of convergence R.

(4) Nothing is claimed about convergence or divergence for a value of z such that |z| = R.

(5) There is a formula for R: $1/R = \limsup |a_n|^{1/n}$ = limes superior of sequence $|a_n|^{1/n}$ (Hadamard)

Notice that even if the sequence of *coefficients* cycles through multiple accumulation points, there will still be some very-large-n least upper bound on the accumulation points which is the limes superior. Obviously if $a_n = 1$, this formula gives R = 1.

I comment on the two people Abel and Hadamard in the notes. Abel's 1827 work was for *real* series, and Hadamard in 1892 found the complex series radius formula, not knowing that Cauchy had found it maybe 50 years earlier. By fiddling with the derivatives as in (3c) above, people got the idea that maybe ANY reasonable function could be represented at least in some z region as a power series, and thus we have the idea of Maclaurin (z=0) and Taylor (z=a) series in the complex function world. The two guys Maclaurin and Taylor did their work for real series way back around 1725. The whole flow of complex function theory is not at all presented by A, not his job I guess.

2.5 Abel's Limit Theorem (42)

This theorem concerns the possibility of convergence of a series at a point on the circle |z| = R. Suppose the series is $f(z) = \Sigma a_n (z/R)^n$ and suppose $\Sigma a_n = B$, some finite number. We would like to say that $\lim_{z \to R} f(z)$ = "f(R)" = B. This Abel limit theorem just clarifies the fact that this hope is realized as long as you approach the convergence point on a line from inside the circle which is not tangent to the circle! A guy Stolz in 1875 got this theorem into its final form and I quote it in the text. Note that this theorem does not say that there are any points on the circle where a series converges, there might or might not be. The simple series $\Sigma (z/R)^n$ in fact diverges everywhere on the circle.

3. The Exponential and Trigonometric Functions (43)

I spent a lot of time on this, perhaps too much, because it interested me. I wrote "more about e" while doing my meta notes for this section.

3.1 The Exponential.

Question: What is the "reader" supposed to "know" before reading this section? The reader is probably used to the notion of "a number raised to a power" such as z^n where n is an integer. The meaning is then z*z*..z. Also the reader knows about $z^{1/n}$ which can be discussed in the same sense. Then I suppose you could talk about $z^{n/m}$ with a rational exponent: $y = z^{n/m}$ means $y^m = z^n$ which means y*y*y... = z*z*z... . So we have handled any rational exponent z^r .

Does the reader know this: $z^{n1/m1}\ z^{n2/m2} = z^{n1/m1+n2/m2}$? I suspect this could be figured out. But more generally, what about $z^a z^b = z^{a+b}$ where z, a and b are all real numbers? Or $(z^a)^b = z^{ab}$? How would you prove these well known facts? Both are clear for a and b both being integers, but we want to

have an interpolated clean meaning between such integers. I suspect that you can NOT prove this without introducing the exp(z) and ln(z) functions.

[That suspicion is correct. I had to spend a *lot of time* writing "more about e.doc" today to clarify this stuff. The whole meaning of the notation a^x for non-integer x requires the notion of a certain interpolating function with certain assumptions on it. This leads to $a^x \equiv \exp(x \ln(a))$ as the interpolating function. A special case is then $e^x = \exp(x)$ since ln(e) = 0. So there is no "big jump" in going from the notation exp(x) to the notation e^x. Really exp(x) is the "meaning of e^x " so writing e^x is no different than writing exp(x).]

So back to Ahlfors. He chooses to define his exp(z) as the ODE solution. This is where I was thrown off when he also write it as e^z. He gets the power series solution for e^z from the ODE. $R = \infty$ for convergence. A then derives $\exp(x)\exp(y) = \exp(x+y)$ in the method I quote in my about e doc, he calls this "the addition theorem". He also derives various other properties for complex exponents.

3.2 The trig functions (44)

These are defined using the e^x function and there is no mention of angles or geometry at all!!

3.3. The Periodicity (45)

Again, without any notion of angles running through 2π, he shows that the trig functions are periodic and he then defines the period to be 2π. So this is his definition of the number π. And of course e is the sum of the power series with x = 1. So e and π are related to each other, and of course $e^{i\pi} = -1$. You could take the imaginary part and say $\sin(\pi) = 0$ and compute π from the sine power series, but this does not let you "compute π from e" or vice versa.

3.4 The Logarithm (46)

He defines this as the inverse to the function exp(x). [I defined exp(x) as the inverse of ln(x) in my notes about e.] He shows that $\ln(w) = \ln|w| + 2\pi ni$ for n = all integers, hence the "sheets". He then ends this section talking about things like $\cos^{-1}w$.

Chapter 3. Analytic Functions as Mappings (49)

1. Elementary Point Set Topology.

My raw notes give a reference to a low-cost book on this general subject

1.1 Sets and Elements (50).

You can have sets without having a metric or vector space. Ie, don't need notion of distance between set elements, nor do you need a way to add elements. Days of the week form a set.

1.2 Metric spaces (51).

The **δ-neighborhood** of a point is I think an open ball of radius δ>0 around the point. A **neighborhood** of a point is any set including that point which contains at least one **δ-neighborhood** (for some $\delta > 0$). These definitions seem a little tedious. **Isolated point** is alone in its neighborhood, **accumulation point** is the usual thing. Metric is written d(x,y) where x and y are metric space elements.

1.3 Connectedness (54).

A (**topology** T on a set of elements X) is a family of subsets of X that is "reasonable" in that it has three very reasonable properties:

1. Both the empty set and **X** are elements of ***T***.
2. Any union of arbitrarily many elements of ***T*** is an element of ***T***.
3. Any intersection of finitely many elements of ***T*** is an element of ***T***.

A topology is purely a set concept, no need for distance or addition. You can talk about a topology on a space that is not a metric space, for example (and not a vector space either). Thus, a topological "space" is somehow a broader thing than, say, a metric space. In Section 1.6 below we shall see the significance of this "broader thing" idea.

The question of interest to us now is this: given a set X, what are the "***connected* subsets**" of X. For the real line, any **interval** is a connected subset. The union of two connected intervals need not be connected, so I would say the connected subsets of R do not form a topology on R due to requirement 2 above. In a plane (set = points in plane), a rule for determining whether a set (a region) is connected is given: an in-set **poly line** connection exists between any pair of points. If a set is *not* connected, it can be **partitioned** into a set of subsets which are each connected and those sets are called the **components** and the partition is unique. The components might be **locally connected** (every point has a connected neighborhood). Number of components is countable. If space S has a **dense** subset E, then S is **separable**.

A is throwing out lots of fancy definitions here, just giving his student a little whiff of things.

1.4 Compactness (59).

A metric space is **complete** if every Cauchy sequence converges.

definition of compact: A set X within metric space S is **compact** if X has the **Heine-Borel Property**: every open covering of X has a finite subcovering. **(Def 6)**

Theorem X: If X is compact, then it is complete.

X is bounded: $d(x,y) \leq M$ for all x,y in X

This section is pretty dense, see raw note comments, here are some of the major results:

1. A Heine-Borel compact space is complete, meaning all Cauchy sequences converge. (Theorem X)

2. A compact set is bounded.

3. If you can cover a set X with a finite number of $N_{\varepsilon}(y)$ neighborhoods, set is "**totally bounded**". **(Def 7)**

6. (set is compact) ⇔ (set is complete and totally bounded) **[Theorem 6]**
   6a. (set is compact) ⇒ (set is complete) // same as Theorem X
   6b. (set is compact) ⇒ (set is totally bounded)
   6c (set is complete and totally bounded) ⇒(set is compact)

7. (A subset of R or C is compact) ⇔ (it is closed and bounded)

On page 62, he claims to have three distinct "characterizations" of compactness:

1. The Heine-Borel definition of compactness : any open cover has a finite subcover

2. The Theorem 6 idea that set is compact if it is complete and totally bounded.

3. The idea that every subsequence must have a convergent subsequence (limit point). [Theorem 7].
This is the **Bolzano-Weierstrass** property if you make this be your definition of compactness, or it is the BW Theorem if you use the Heine-Borel definition.

I note that 3 is the historical definition, and 1 is the modern definition. I found some good low-cost Amazon books on this subject.

1.5 Continuous Functions (64).

In Section 1.4 above we first did generic sets, then we did metric space sets, and we worried about the connectedness and the compactness of subsets of metric spaces. Here for the first time in this chapter we talk about **functions** which map between two such metric space sets, so we have f:S→S' and a possible inverse mapping to go with it. Metrics are d and d'.

We are especially interested in functions which are **continuous** which means within ε ball in S' means within δ ball in S, notice the reverse sense of this definition. I show in a picture how this fails for a function of one variable which has a piece with infinite slope, and in that same example, I show how open sets in the range do NOT map to open sets in the domain (which really is the same as the definition of continuous). In fact we have

Theorem Y: (f continuous) ⇔ (the *inverse* image of any open set is an open set)

and you can replace the words open with the words closed and theorem is still true. We now get a series of theorems which I will just quote:

Theorem 8: If f is continuous, it maps any compact set into a compact set. (see raw note comments)

Theorem 9: If f is continuous, it maps any connected set into a connected set.

Topological Mapping: If f is **one-to-one** and uses up all the range (is **onto**), then you can invert any point in the range (image) and therefore $\mathbf{f}^{-1}$ **exists** and everything is nice. IF in addition both f and f^{-1} are continuous, then such a mapping is called a **topological mapping** (= **homeomorphism** -- "the isomorphism of the topological world"). Such a mapping preserves certain topological properties. Two

such properties that are preserved are compactness and connectedness, according to our theorems 8 and 9. Openness is not such a property.

Uniform Continuity. Let's review our definition of "regular" continuity above: a function $f:x_1 \to x_1'$ is continuous at point x_1 if ε ball around x_1' implies δ ball around x_1. That is to say: $|x' - x_1'| < \varepsilon$ can find $|x - x_1| < \delta$. Can also write this as $d(x',x_1') < \varepsilon$ can find $d(x,x_1) < \delta$. Here the points x and x' are close to x_1 and x_1', but we regard these latter two points as fixed, and we are talking about f being continuous *at* x_1.
Now, if this is true for all x_1 in some region R, then f is said to be **uniformly continuous** *on R*. So the word uniform here means uniformly over some region R. We are reminded of the notion of something converging $f_n \to f$, whereas uniform convergence means $f_n(x) \to f(x)$ for x in some region R. In both cases, some "parameter" ranges over some region and for each parameter value you must have the property of interest. Remember that in both cases the "δ" cannot be a function of the parameter, the same δ must work throughout the region. I give some examples of functions which are and are not uniformly continuous: $f(z) = z$ is uniformly continuous *on the entire C plane*, but $f(z) = z^2$ is not. With this definition we are now ready for the next theorem:

Theorem 10: If continuous f acts on a compact set, then f is uniformly continuous on that set.

1.6 Topological spaces (67).

I am quoting my raw notes verbatim here.

This is very strange again. He wants us to get rid of the notion of "distance" which we have used as our main tool! He wants us to replace this tool with the tool of "open sets". He defines the topo space in the standard manner in Def 8, I already quoted this above. It is not a metric space with a metric (distance). It is a topological space with a "topology", namely, with a set of open sets which have the usual properties.

Example: how would you talk about a **neighborhood** without talking about distance? Here it is:

N is a neighborhood of x if we can find open set U such that $x \in U \subset N$.

Distance was never mentioned. This approach is discussed here:

http://en.wikipedia.org/wiki/Neighbourhood_(mathematics)

So this is the "thing that happened" in math since the early days! There has been a conversion from metric space thinking (with its distance between points) to topological space thinking (with its open sets). I wonder WHEN this happened. [My wild guess: old way 1860, new way 1920]

To the extent our earlier ideas and theorems involved only open sets (such as the Heine Borel definition of compactness), they are true for topological spaces as well as for metric spaces. But theorems that involve convergence, with its strong distance requirement, have problems. These are fixed somehow if we add the magic **Hausdorff Property** (1914) to our topo space:

Hausdorff Property: you can surround any two distinct points in topo space with disjoint open sets

This reminds me of developing geometry "axiomatically" and see what is the minimum you have to assume, your minimal axiom set. This Hausdorff thing, he claims, prevents a sequence from converging to two different points. This is pretty technical. What is the meaning of "convergence" in this topo space world in the first place?

> $x_n \rightarrow x$ in topo space means every "neighborhood" (defined above) of x contains all x_n except a finite number (which would be the head of the sequence).

So no matter how you place your open set U around x, U contains a tail of the sequence. Suppose there were two tail convergence points x and y. Well, I need help here, but I sort of see the idea. He says it is "obvious" that the Hausdorff condition makes the tail converge to a unique point. It is not obvious to the beginning student reader who has not been in the field for 35 years! Suppose there are two limit points. Suppose you can separate them with two open sets that are disjoint. Suppose the sequence just jumps back and forth between x and y, so each open set has an infinite number of points. Well I guess the entire tail has to fit in one open set. A good text on this subject would have lots of examples, but here he is just mentioning this extra "axiom" so we will have heard about it once in our lives.

A topo space with this property is a **Hausdorff Space** and that is the only kind of space we care about.

Complaint about Ahlfors' book: <u>There is not a single reference anywhere</u>! If you want to read more on some subject, he has zero advice for you! I wonder why he chose that path? Goldstein, Jackson, Stakgold all have references at the end of each chapter. BD have footnote references.

2. Conformality (68)

2.1 Arcs and closed curves (68).

A finite **arc** γ is closed and bounded in C, hence compact (our first application of the previous section), and can be **parameterized** with some real t, hence the arc can be written as some function z(t) which maps a finite interval of R to C. The **tangent** to the arc at some point is z'(t). If this avoids being 0, arc is **regular**. If z'(t) exists all along the arc and is a continuous function, arc is **differentiable** as for a regular function. If arc does not self-intersect it is **simple** or **Jordan**. (not sure this is in domain or range or either). Arc could be **closed**. A non-self-intersecting closed arc is a **Jordan curve**.

2.2 Analytic functions in regions (69).

Recall that a **region** is a non-empty connected set in C (see p 57).
Def 10 states our definition of f(z) being **analytic on a region Ω** .
Def 11 says f(z) is **analytic on a point set A** if it is analytic in a region containing A.

Ahlfors then digresses on the notion of **branches** or **sheets** of functions. He does $z^{1/2}$ and ln(z) as the first two examples, giving both a branch cut on the negative real axis, using $-\pi,\pi$ for the "angle". He also talks about $\cos^{-1}(z)$ which can be written in terms of the $z^{1/2}$ and ln(z) functions p 72. Section ends with this battery of similar theorems:

Theorem 11: If $f'(z) = 0$ for an analytic function $f(z)$ everywhere in a region Ω, then $f(z) =$ constant.
Theorem 11A: If $\arg(f(z)) =$ constant everywhere in Ω, then $f(z) =$ constant.
Theorem 11B: If $\mathrm{Re}((z)) =$ constant everywhere in Ω, then $f(z) =$ constant.
Theorem 11C: If $\mathrm{Im}((z)) =$ constant everywhere in Ω, then $f(z) =$ constant.
Theorem 11D: If $|f(z)| =$ constant everywhere in Ω, then $f(z) =$ constant.

2.3 Conformal Mapping (73).

Ahlfors first shows this fact: the rotation of the tangent to an arc through point z_0 as you go through mapping f is a function only of $f'(z_0)$. Thus, if you have two domain arcs γ and γ' that both pass through z_0, both tangents are rotated the same amount. This at once means that relative angles are preserved. Of course we have to assume that $f(z)$ is analytic in some tiny region Ω about z_0 so this derivative exists. Here is the proof that tangent rotation depends only on $f'(z_0)$:

$w(t) = f(z(t))$ mapping of arc $z(t)$ into image arc $w(t)$
$w'(t) = f'(z(t))\, z'(t)$ chain rule then relates domain and image tangents
$\arg [w'(t_0)] = \arg [f'(z_0)] + \arg [z'(t_0)]$ $z_0 = z(t_0)$ imaginary parts of both sides
angle of image tangent = angle of domain tangent + $\arg [f'(z_0)]$

So: Any analytic mapping f(z) **preserves angles** of intersecting arcs -- the "**conformal** property"

Similarly he shows that the mapping **preserves scale**. For example, if a 1 mm line segment ending on z_0 maps into a 3 mm line segment ending on $f(z_0)$, then as you rotate your initial segment, the final segment will always be 3 mm as it rotates in the image (ie, true to the extent that a mm is "small" relative to the mapping warp). If either angles or scale are preserved, you can show that the mapping must be analytic.

Now we add $f'(z) \neq 0$ to the situation. Without proof at this point, A shows that this implies that f(z) must have an inverse at least in some small Ω and $f(\Omega)$ region, which means f(z) is 1-to-1 and the inverse is analytic, which means that f(z) is then a **topological mapping** as defined earlier. If $f'(z) = 0$, $f^{-1}(z') = \infty$ and we lose the analyticity of the inverse function, so that is why we require $f'(z) \neq 0$. As the picture on page 75 shows, the mapping might be "locally topological" but not globally so.

So to summarize: *(verbatim from raw notes)* ANY analytic function f(z) provides a mapping from domain to image plane which preserves both angles and scale and is thus a "conformal map". This f(z) does not have to be a rational function or anything particular, just as long as it is analytic. This seems to mean that for a tiny piece of the domain near z, we map into a tiny piece of image where things are merely rotated and scaled by some amount. There is no distortion! If you put a tiny pixel image in this domain region, the image region would show the image merely rotated and scaled, but a circle would remain a circle. Of course for *large* pixel image, this would not be the case, because conformality is a local property. I would add that this means if you define a local x,y orthogonal coordinate system in the domain, and a similar local one in the range at the corresponding point, then the image x',y' coordinate system is merely rotated and scaled relative to x,y. It seems to me this fact ought to be useful.

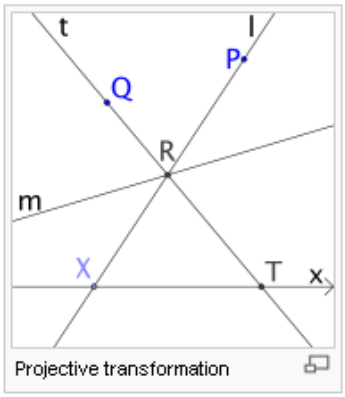
3. Linear Transformations (76)

3.1 The linear group (76).

Before we can even start this section, we shall digress on a mathematical side subject known as projective transformations, and I devoted a lot of raw notes to this subject. This digression is not really necessary, only if you want to understand what Ahlfors is saying on page 77. Still, I include some digressive notes here in this meta review (probably way to much for a meta review, really "second round" notes).

(a) transformations on the projective line

The subject opens with a sort of 2D "playhouse" at which two audience members P and Q look at a real 1D play taking pace on a sloped line m, the stage. Behind this sloped stage is a 1D matte painting that is not sloped which provides a background scene for the play. When the two audience members view the point R on the stage, just behind this point they see points X and T on the matte painting. Whether the stage (objective line) is sloped or the background (subjective line) is sloped is all relative, they are sloped relative to each other. So here is the picture:



You can think of the audience seeing the stage actors in effect projected onto the background. Of course in this 2D version, the actor's world and the matte painting are 1D affairs, but no matter. Now, if we put an (x,y) origin on the background screen to the left of the point X such that X = (x,0) and T = (t,0), we can ask this question: Given x, what is t ? If viewer P sees R at X on the matte painting, where does viewer Q see R on the matte painting (ie, where is T)? Here is the answer:

$$t(x) = \frac{\alpha x + \beta}{\gamma x + \delta}$$

where

$$\alpha = mQ_xP_y - Q_yP_y + bQ_y,$$
$$\beta = bQ_xP_y - bQ_yP_x,$$
$$\gamma = m(P_y - Q_y),$$
$$\delta = mP_xQ_y + P_y(b - Q_y).$$

To get this formula, you can put point R at (r,mr+b) and then write an equation for the line through Q,R and T and another equation for the line through P,R and X. Then eliminate r between these two equations and you get the answer shown. [I will eventually show in an external document how this calculation works out in the general case.]

Notice that the formula on the left has the form of what we would call a "fractional linear transformation" (although it is clearly non-linear in the sense $t(x+y) \neq t(x)+t(y)$). As expected, the answer depends on the coordinates of viewers P and Q and on the equation of the stage line which is $y = mx+b$. We could write a similar formula for $x(t) = (at +b)/(ct+d)$ where the new parameters come from P ↔ Q. This thing would be the inverse of t(x), and of course $x(t(x)) = x$. So this transformation always has an inverse. There is an obvious identity transformation as well, $\alpha = 1$ others $= 0$. This transformation appears to have 4 free parameters, but there are really only 3 since scaling all four parameters the same results in the same t(x).

As stated here, this transformation has many different names all of which mean the same thing:

- linear transformation (Ahlfors)
- transformation on the projective (real) line (at least this goes with the picture)
- bilinear fractional transformation (I would vote for this name, used by wiki site)
- Mobius transformation (in honor of Mobius who invented the homo coordinates for proj geom, below)
- homographic transformation (wiki)

These names usually imply that all quantities are complex, though our construction to get the transformation gives all real quantities, but let's stay real for the moment.

NOW: Suppose we chain two of these transformations together. Each is really a non-linear operator acting on the reals. We could say that

$x' = t(x; \alpha_1, \beta_1, \gamma_1, \delta_1) = T_1(x)$
$x'' = t(x'; \alpha_2, \beta_2, \gamma_2, \delta_2) = T_2(x') = T_2T_1(x) \equiv ???$

It is a simple fact of algebra that, if you multiply everything out as shown in the raw notes, you find that the resulting transformation can be written as $T_3(x)$ where parameters are a certain α_3 etc which are functions of the 1 and 2 parameter sets! In other words, these "linear transformations" form a **group**, because concatenation of two linear transformations gives another linear transformation! The various other group properties such as identity and inverse are also met, as noted above. This group has several names, one is P(1; R) where R is for reals, P for projective group, 1 for the line.

The formulas for the new 3-parameters are these

$$\begin{bmatrix} \alpha 3 & \beta 3 \\ \gamma 3 & \delta 3 \end{bmatrix} = \begin{bmatrix} \alpha_2 & \beta_2 \\ \gamma_2 & \delta_2 \end{bmatrix} \begin{bmatrix} \alpha_1 & \beta_1 \\ \gamma_1 & \delta_1 \end{bmatrix} = \begin{bmatrix} \alpha_2\alpha_1 + \beta_2\gamma_1 & \alpha_2\beta_1 + \beta_2\delta_1 \\ \gamma_2\alpha_1 + \delta_2\gamma_1 & \gamma_2\beta_1 + \delta_2\delta_1 \end{bmatrix}$$

This shows that our projective group is isomorphic to the matrix group SL(2,R). Here we have added the condition that det(params) = 1 to remove the scaling degree of freedom (this is what S = Special means). So we might say P(1; R) ~ SL(2,R) where ~ means "is isomorphic to". Now every one knows that SL(2,R) has elements which really are linear transformations. So the word "linear" gets into our soup perhaps from two different sources: we did a transformation on the "line", and this isomorphic matrix group has elements which are "linear" transformations of vectors in some 2D space.

If we allow everything to "go complex", all our equations look exactly the same, but of course our movie theatre model is harder to draw. In this case, if we scale all parameters by $Re^{i\theta}$ for arbitrary R and

θ, the transformation stays the same, so there are then only 8-2 = 6 real parameters. And of course we then say that P(1; C) ~ SL(2,C) where C now means complex instead of real.

(b) transformations on the projective plane and beyond

Now we have a 3D "playhouse" with a 2D stage on which we have 2D actors acting on a planar objective screen behind which is a 2D matte painting. You might think of the 2D actors on the 2D tilted planar screen as a movie of 3D actors, and the screen on which this movie appears is partially transparent so you can see the matte painting behind it. In any event, our picture is the same as above, more or less. The stage is a 2D plane with equation z = mx+ny+b where now z is the axis perp to the matte painting subjective screen. We now denote our subjective screen points as X = (x,y) and T = (T_x, T_y) and we ask the same question as before: how do we express the coordinates of T in terms of those of X ? The result is shown in the raw notes, of which I now quote some of the results:

$$T_x = \frac{\alpha x + \beta y + \gamma}{\delta x + \epsilon y + \zeta}. \qquad T_y = \frac{\eta x + \theta y + \kappa}{\delta x + \epsilon y + \zeta}.$$

Our wiki site calls these trilinear fractional transformations and we have 9 real parameters for our real playhouse, see raw notes for the formulas. If we concatenate to of these trilinear transforms, you just insert things and rationalize and you get another trilinear transform, so again we get a group! The parameters of a concatenated transform are given as follows (that is $T_3 = T_2T_1$) :

$$\begin{bmatrix} \alpha_2 & \beta_2 & \gamma_2 \\ \eta_2 & \theta_2 & \kappa_2 \\ \delta_2 & \epsilon_2 & \zeta_2 \end{bmatrix} \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \eta_1 & \theta_1 & \kappa_1 \\ \delta_1 & \epsilon_1 & \zeta_1 \end{bmatrix} = \begin{bmatrix} \alpha_3 & \beta_3 & \gamma_3 \\ \eta_3 & \theta_3 & \kappa_3 \\ \delta_3 & \epsilon_3 & \zeta_3 \end{bmatrix}.$$

So in this case we get P(2,R) ~ SL(3,R) . If we let all go complex, we get P(2,C) ~ SL(3,C) . Notice that these 3x3 matrices must operate on some kind of 3-vectors.

An interesting fact about these transformations: if you apply them to the equation of a conic section (meaning general quadratic = 0) , you get the equation of another conic section.

We can continue our progression with these projective transformations and get to a general case which would be this: P(n,C) ~ SL(n+1,C) where we did n=1 and n=2 above, but any larger integer n is also possible. Here are the results for n = 3, the quadrilinear fractional transformations:

$$T_x = \frac{\alpha x + \beta y + \gamma z + \delta}{\epsilon x + \zeta y + \eta z + \theta},$$
$$T_y = \frac{\iota x + \kappa y + \lambda z + \mu}{\epsilon x + \zeta y + \eta z + \theta},$$
$$T_z = \frac{\nu x + \xi y + o z + \rho}{\epsilon x + \zeta y + \eta z + \theta}.$$

$$M_T = \begin{bmatrix} \alpha & \beta & \gamma & \delta \\ \iota & \kappa & \lambda & \mu \\ \nu & \xi & o & \rho \\ \epsilon & \zeta & \eta & \theta \end{bmatrix}$$

Here, the linear transformations map F(x,y,z) → (T_x,T_y,T_z) and we have P(3,C) ~ SL(4,C) so we have 4x4 matrices which act on some mysterious 4-vectors. Notice that the projective transformation is non-linear in this sense: F($\mathbf{r_1}$ + $\mathbf{r_2}$) ≠ F($\mathbf{r_1}$) + F($\mathbf{r_2}$) where the $\mathbf{r_i}$ are 3-vectors. However, the matrix transformations acting on their mysterious (so far) 4-vectors are linear in that M ($\mathbf{A_1}$ + $\mathbf{A_2}$) = M($\mathbf{A_1}$) + M($\mathbf{A_2}$) where the A_i are these 4-vectors.

Comment: I got interested in this subject and have just written a long doc on the subject whose title is *Projective and Fractional Linear Transformations.doc* . The general case is treated and homogeneous coordinates are explained (which appear in Ahlfors at the bottom of page 76 for the case n=1). The various mysterious comments Ahlfors makes at the top of page 77 are explained in full.

(c) notes on this book section

He writes the general (n=1) linear transformation as p 76 (5) and gives the inverse in p 76 A. If you want this inverse to exist, you need det M ≠ 0 where M is the obvious 2x2 matrix. He suggests that we then set detM = 1 to set the scale of M (since scaling all four constants leaves the LT unchanged). He then introduces homogeneous coordinates in a very *ad hoc* manner that left me pretty mystified, and claims that you can then do 2x2 matrix multiplication to get the results of combining two LT's. I found this to be non-obvious, and that led me to write a separate document on this whole subject, as noted in the Comment above.

[77] So, the main useful result is as shown top page 77 where you use 2x2 matrices to compute the results for concatenating two linear transformations. At the end of this section, he looks at certain special cases of the 2x2 matrix (and the corresponding linear transformation).

$\begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$ w = z + α **parallel translation**

$\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}$ w = kz **scaling**

There are two special cases of scaling. If $k = e^{i\theta}$, it is a **rotation**. If k is real and k>0, it is called a **homothetic transformation**, which is a pretty fancy term for scaling by a positive real number.

$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ w = 1/z **inversion**

He finally shows that the most general LT can be written as a concatenation of the above primitive elements.

3.2 The Cross Ratio (78)

The opening act here, shown in (7), is writing a particular LT w = S(z_2,z_3,z_4)z [ie, S:z→w] that does this mapping:

$z_2 \to 1$, $z_3 \to 0$ and $z_4 \to \infty$

which is to say, an *arbitrary* triplet of C plane points z_2,z_3,z_4 is mapped to these three points on the real line. The fact that (7) does this is obvious by inspection, though A manages to confuse the issue by using his horrible ratio symbol ":" , suggestive of the German "double ratio" term, see below.

Next, he defines **the cross ratio** by yet another strange symbol "(z,z_2,z_3,z_4)" $= S(z_2,z_3,z_4)z$, making use of the transformation he just defined. He sets $z = z_1$ to have it look nice, so we then get the traditional result,

$$(z_1,z_2,z_3,z_4) = S(z_2,z_3,z_4)z_1 = [\,(z_1 - z_3)/(z_1 - z_4)\,] * [(z_2 - z_4)/(z_2 - z_3)] \qquad \text{// p 78}$$

which I rewrite in the following form to be used below

$$(z,z_1,z_2,z_3) = S(z_1,z_2,z_3) = (z - z_2)/(z - z_3) * [(z_1 - z_3)/(z_1 - z_2)] = (z - z_2)/(z - z_3) * (1/Z_{123})$$

$$\text{where } Z_{123} = [(z_1 - z_2)/(z_1 - z_3)] = 1/Z_{132}$$

Terminology and history

The English term "cross-ratio" was introduced by Clifford. German geometers of the 19th century called it *das Doppelverhältnis* (Ger: double ratio) and French mathematicians after Chasles used the term *le rapport anharmonique* (Fr: anharmonic ratio).

In a geometric context, the notion of a cross-ratio goes back to antiquity. A theorem on the anharmonic ratio of lines appeared in the work of Pappus, but Michel Chasles, who devoted considerable efforts to reconstructing lost works of Euclid, asserted that it had earlier appeared in his book *Porisms*

The cross-ratio is preserved by the fractional linear transformations. It plays a prominent role in projective geometry because it is the only projective invariant of an ordered quadruple of points on a projective line.

This last comment jumps the gun a bit, but suppose we do our projective thing to get $t_i = f(x_i; P,Q,\sigma)$ for an arbitrary set of 4 points on the subjective line, i = 1,2,3,4. This is an FLT of n=1 and we will show below that it preserves the cross ratio, hence $(t_1,t_2,t_3,t_4) = (x_1,x_2,x_3,x_4)$ so we would say that this cross thing is "a projective invariant of an ordered quadruple of points on a projective line". The order must be the same on both sides. The web comments suggests there are no other invariants. So now we claim:

Theorem 12: the cross ratio is invariant under linear transformations.

I give here verbatim my proof from the raw notes.

Proof. His proof is very efficient -- I would have written everything out to get a big mess. As usual, A manages to make his proof be hard to follow, so I will expand his ultra-densepack cryptic statements to a human-understandable proof.

Warmup Problem: Suppose someone tells you this fact about an LT called R:

$R\,[z_2,z_3,z_4] = [1,0,\infty\,]$ // ie, it does this mapping on three points

You would conclude that $R = S(z_2,z_3,z_4)$, since $S(z_2,z_3,z_4)$ does this same mapping, and because the LT that does this is unique (could prove that). You would go on to say $(z_1,z_2,z_3,z_4) \equiv S(z_2,z_3,z_4)z_1 = Rz_1$. So your conclusions would be:

(a) $R = S(z_2,z_3,z_4)$

(b) $(w,z_2,z_3,z_4) = Rw$ for any point w

Now onto the real problem:

(1) When S appears with no arguments, it is an abbreviation for $S(z_2,z_3,z_4)$. T = some LT.

(2) Question: what does ST^{-1} do to the points Tz_2,Tz_3,Tz_4 ? Answer:

$ST^{-1}[Tz_2,Tz_3,Tz_4] = S[z_2,z_3,z_4] = [1,0,\infty]$ // last = by definition of S !

(3) Applying the results of our warmup problem, we would conclude that (think $ST^{-1} = R$)

(a) $ST^{-1} = S(Tz_2,Tz_3,Tz_4)$

(b) $(w,Tz_2,Tz_3,Tz_4) = (ST^{-1})\,w$ for any point w

(4) Now select $w = Tz_1$. Then from (b) we can write

$(Tz_1,Tz_2,Tz_3,Tz_4) = (ST^{-1})\,Tz_1 = Sz_1 = S(z_2,z_3,z_4)z_1 = (z_1,z_2,z_3,z_4)$

and therefore the cross ratio is invariant under any LT like T.

Question: How would you find a linear transform that takes z_1, z_2, z_3 to w_1, w_2, w_3 ?

Here is a mapping, with z as variable, which maps z_1, z_2, z_3 to 1,0,∞:
$y = Az = S(z_1,z_2,z_3)z = (z,z_1,z_2,z_3)$

Here is a mapping, with w as variable, which maps w_1, w_2, w_3 to 1,0,∞:
$y = Bw = S(w_1,w_2,w_3)w = (w,w_1,w_2,w_3)$

The mapping that maps z_1, z_2, z_3 to w_1, w_2, w_3 must be this
$y = Az$ $w = B^{-1}y \Rightarrow w = B^{-1}(Az) \equiv Cz$

This says $Bw = Az$ which from above says $(w,w_1,w_2,w_3) = (z,z_1,z_2,z_3)$. If we solve this equation for w in terms of z and the 6 parameters, we get the following messy LT (done in the raw notes)

$w = Cz = (az + b)/(cz + d)$ (*)

$a = \alpha w_2 - w_3$ $\alpha = Z_{123}/W_{123}$
$b = w_3z_2 - \alpha w_2z_3$

$c = \alpha - 1$ $\qquad\qquad$ $Z_{123} = [(z_1 - z_2)/(z_1 - z_3)]$

$d = z_2 - \alpha z_3$ $\qquad\qquad$ $W_{123} = [(w_1 - w_2)/(w_1 - w_3)]$

Exercise for Meta Notes: Repeat this same calculation using the matrix method:

Start with this version of the cross ratio stated earlier and derived in the raw notes

$$(z,z_1,z_2,z_3) = S(z_1,z_2,z_3) = (z - z_2)/(z - z_3) * [(z_1 - z_3)/(z_1 - z_2)] = (z - z_2)/(z - z_3) * (1/Z_{123})$$

$$= (z - z_2)/(Z_{123}z - Z_{123}z_3) \qquad \text{where } Z_{123} = [(z_1 - z_2)/(z_1 - z_3)] = 1/Z_{132}$$

Now construct the matrices,

$$S(z_1,z_2,z_3)z \sim \begin{pmatrix} 1 & -z_2 \\ Z_{123} & -Z_{123}z_3 \end{pmatrix} = A = \begin{pmatrix} 1 & -z_2 \\ Z & -Zz_3 \end{pmatrix}$$

$$S(w_1,w_2,w_3)w \sim \begin{pmatrix} 1 & -w_2 \\ W_{123} & -W_{123}w_3 \end{pmatrix} = B = \begin{pmatrix} 1 & -w_2 \\ W & -Ww_3 \end{pmatrix}$$

You would then compute

$$B^{-1} = k\begin{pmatrix} -Ww_3 & w_2 \\ -W & 1 \end{pmatrix} \qquad k = (1/(Ww_2 - Ww_3))$$

We then find that

$$C = B^{-1}A = k\begin{pmatrix} -Ww_3 & w_2 \\ -W & 1 \end{pmatrix}\begin{pmatrix} 1 & -z_2 \\ Z & -Zz_3 \end{pmatrix} = k\begin{pmatrix} -Ww_3+Zw_2 & Ww_3z_2 - Zw_2z_3 \\ Z-W & Wz_2-Zz_3 \end{pmatrix}$$

$$= kW\begin{pmatrix} -w_3+\alpha w_2 & w_3z_2 - \alpha w_2z_3 \\ \alpha-1 & z_2-\alpha z_3 \end{pmatrix} \qquad \alpha \equiv Z_{123}/W_{123}$$

We can dispense with the overall constant when constructing the corresponding FLT, so

$$C \approx \begin{pmatrix} -w_3+\alpha w_2 & w_3z_2 - \alpha w_2z_3 \\ \alpha-1 & z_2-\alpha z_3 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and this exactly duplicates the results shown above in (*)/

Note: I wasted about 4 hours because I made this stupid mistake while doing the above matrix stuff:

$$(z - z_2)/(z - z_3) * (1/Z_{123}) \sim \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & -z_2 \\ 1 & -z_3 \end{pmatrix} * (1/Z_{123}) \qquad \text{WRONG!}$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & -z_2 \\ Z_{123} & -Z_{123}z_3 \end{pmatrix} \qquad \text{RIGHT!}$$

For some reason I could not detect this trivial error despite vigorous debugging efforts. Very scary.

We now come to our next claim:

Theorem 13. Claim that (z_1,z_2,z_3,z_4) is real $\Leftrightarrow$ the four points z_i lie on a Circle

where Circle means circle or line. I prove this in the raw notes "in my own way" which is not great but I think it works. His way required the usual frustrating and time-consuming decoding so I thought my way would be faster. At least it was interesting, but it certainly was not a good or fast way. I could probably find some better way, and it would probably come out being his way. In any event, this Theorem 13 is key for what follows.

Theorem 14: Linear transformation maps Circles into Circles.

The proof is trivial given Theorem 13. Start with a Circle and pick four points on this circle z_i . The cross is real. The LT takes these points to four w_i. Since the cross is invariant under LT, the new cross is also real in the w_i . But then Thm 13 says we have a Circle in w.

Summary to this point: Any analytic f(z) transformation preserves angles and scale. Linear transformations in addition preserve Circles.

3.3 Symmetry (80)

I have a lot of raw notes here, but the ideas can be compactly presented I think. In these notes, I use the notation $\bar{z}$ to refer to the complex conjugate of point z, and I use the notation z* to refer to a point which is the "symmetry point" of point z reflected in some Circle C. This thing z* needs to defined.

(1) Consider the real axis and the points z and $\bar{z}$. We well know what this looks like, the points z and $\bar{z}$ both lie on a line which bisects the real axis at x = Re(z). The distances from the real axis are the same, namely |Im(z)|, one is above and the other below.

(2) The question is this: what happens to the points z and $\bar{z}$ if you run them both through a LT T? By definition (of w*), you get w = Tz and w* = T$\bar{z}$. We know that the LT T maps the real axis into some Circle C (which could be a circle or a line). It turns out that the set (w,w*,C) has an internal relationship. Given w and C, w* is fully determined. There might be lots of T's which map the real axis into C, we don't care. When the dust settles, here is what you find. Assume the circle C has its center at point a.

- the three points a, w and w* are collinear
- $|w^* - a|\,|w - a| = R^2$

terminology: w and w* are reflected in C

If point w lies inside the circle, then $|w-a| < R$, so $|w^*- a| > R$, which means w* lies outside the circle. So w and w* are on opposite sides of the circle. If the Circle is a line, then w and w* are equally spaced away from the line on another line which bisects the line. I show this explicitly in the raw notes by starting with a circle and taking the limit in which it becomes a line.

On page 81 Ahlfors shows a little geometric construction (which I have verified) showing a circle and a point z, and it shows how you locate z* : do a perp through z, then do tangents where that perp hits the circle, and they meet at z*. This picture shows that the symmetry point on the outside of the circle is always farther from the circle than its partner point inside. And as z moves toward the center of the circle, the partner point moves to ∞ and gets very far away indeed.

Theorem 15: if T takes C_1 to C_2 , it takes a sym pair (w_1, w_1*; C_1) to a sym pair (w_2,w_2*; C_2)

Proof: We know that a linear transform carries circle C_1 into circle C_2. We could of course write the transform as a product of two transforms one that takes C_1 into the real line, and then one that takes that real line into C_2. If we start with a symmetry point pair w_1, w_1* relative to C_1 , the first mapping would take this into a symmetry point pair z, $\bar{z}$ relative to the real axis, and then the second mapping would take these symmetry point pair w_2, w_2* relative to C_2.

Interesting Theorem p 82: Suppose you want to find a T that maps C_1 to C_2. Take some z_1 on C_1 and some z_1' *not* on C_1 . These map into some w_1 on C_2 and some w_1' *not* on C_2. Since z_1' is not on C_1, we know it has some distinct symmetry point z_1'* which maps under T to w_1'*. So we know 3 points of the mapping: T [z_1, z_1', z_1'*] → [w_1, w_1', w_1'*] . Reviewing what we did: we picked some z_1 and some z_1' arbitrarily and this determined z_1'* relative to C_1. (z_1 on C_1, z_1' off C_1). We did the same thing on the C_2 side, namely, we picked some w_1 and some w_1' arbitrarily and this determined w_1'* relative to C_2. (w_1 on C_2, z_2' off C_2) . We know that

$$(z,z_1,z_1', z_1'^*) = (Tz,Tz_1,Tz_1', Tz_1'^*) = (w,w_1,w_1', w_1'^*)$$

Equating the outer two cross ratios let's us solve explicitly for w = Tz, and this then is a mapping that carries C_1 into C_2 . (there are an infinite number of such mappings, and we just found 1 by our method).

Definition of "reflection" [p 80 location A] . Consider a 2D figure on one side of a circle. Each point of this figure has a symmetry point on the other side of the circle. The locus of these symmetry points is the **reflection** of the figure on the other side of the circle. Obviously the reflected figure will be some distorted version of the figure. A linear transformation T that will map a figure into the reflected figure is this:

$$z^* = a + R^2/(z-a) = [a(z-a) + R^2] /(z-a) = [\, az + (R^2-a^2)] / (z-a) = Tz \qquad M = \begin{pmatrix} a & R^2-a^2 \\ 1 & -a \end{pmatrix}$$

and easy to show that $M^2 = R^2 *$ identity matrix, so $T^2 = I$ for this particular linear transformation.

At this point, I solved a few "homework problems" I gave myself:

Question: what is the most general linear transformation that maps a unit circle into itself?

Answer: $w(z) = e^{-i\sigma}\ \alpha\ (z/\alpha - e^{i\theta}) / (\alpha\ z - e^{i\theta})$ where $z = e^{i\varphi}$ on unit circle

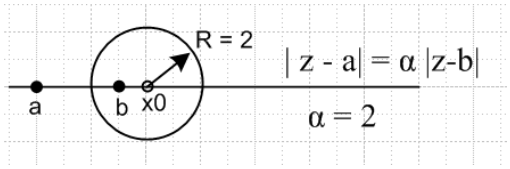
σ = arbitrary angle
θ = arbitrary angle
α = arbitrary positive real number

Question: what do conic sections look like in abs value notation?

$\|z-a\| = d$	**circle** of radius d
$\|z-a\| + \|z-b\| = d$	**ellipse**, obviously $d > 0$
$\|z-a\| - \|z-b\| = d$	**hyperbola** half, "left" half if d>0, else "right" half, bisecting line if d=0
$\|z-a\| / \|z-b\| = d$	another kind of **circle**, again $d > 0$ (see below) " Apollonius"
$\|z-a\|\ \|z-b\| = d$	a **quartic**, not a conic section

$|z-f| = |z-D(z)|$ = where D(z) is a point on the directrix. **parabola**

The parabola does not fit very well with the others. The "another kind of circle" is something I examine in a document called "a geometry problem.doc", here is a picture where I have d = 2.

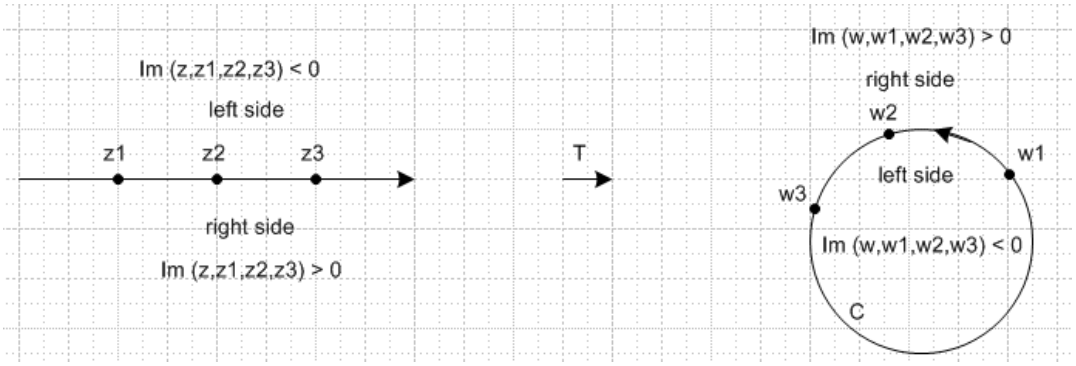


If $d < 1$, the circle is around point a.

3.4 Oriented Circles (83)

My notes indicate this section needed heavy "decoding", ie, rewriting by me. The main points seem not too complicated.

First consider this particular mapping of the real axis to a circle:

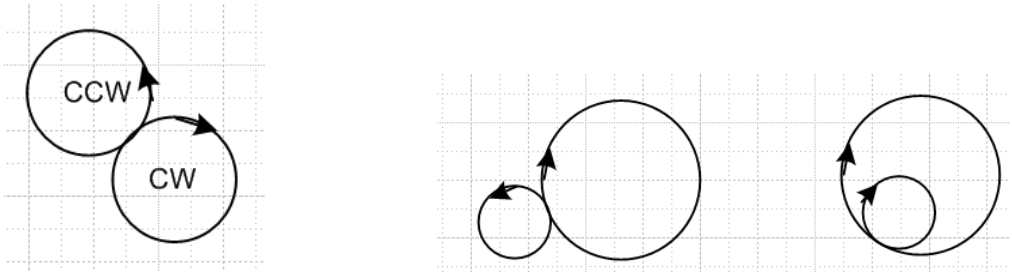


We know that on the line and on the circle, the cross ratio is real if we take z (w) as our forth point (not shown). This was Theorem 13 above. It seems pretty reasonable that one of the half planes on the left will map into the interior of the circle, and the other to the exterior. The question is which is which. I show only one of four possibilities in the above drawing. On page 83 A shows that

$$\mathrm{Im}(z, z_1, z_2, z_3) = (ad-bc)/|cz+d|^2 \,\mathrm{Im}(z)$$

where a,b,c,d are the constants which appear in the cross ratio written as a LT. This shows that there are really two labeling possibilities for the left half of the above picture (I have picked one). I think, however, that once you know which half plane is which in the left side, it always comes out as shown on the right side of the above picture in terms of the arrows. In other words, assuming the left picture is as shown, and assuming the arrows in the right picture as shown, then it is the interior of the right picture that corresponds to the upper half plane, and you think of this as the "left side" if you are walking along the arrow direction. So in some sense, left side maps into left side, whatever sign Im(cross) is for that side. And of course the point is that you have this same concept if you map a circle to a circle.

A then worries about tangent circles (because these are going to come up soon), and I find that drawing a picture makes it all clear. Here is my picture:



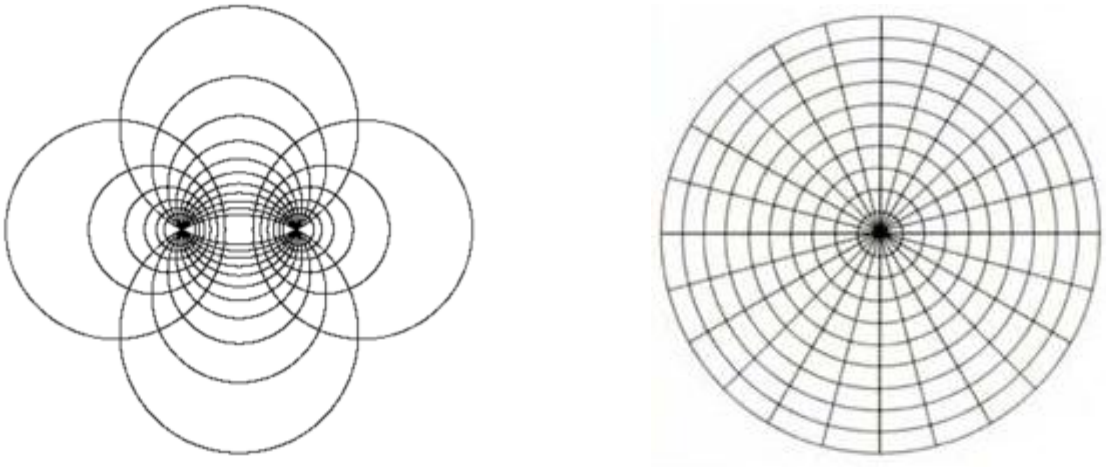
Here we have two tangent circles on the left. We pick a triplet of points on each circle and look at the mapped results and then draw all our arrows. Given the arrows on the left, which "align" at the tangent point, this must also be true in the mapped picture, however the arrows end up. Just think of three closely spaced points near the tangent point which are in effect common to both circles -- they can have only one "direction". Another argument is that "conformality" when applied to a small region around the tangent point says that the situation is merely rotated and scaled, so the arrows have to be as shown.

3.4 Families of Circles (84)

The most general LT can be written in the form w = k [z - a] / [z - b] . This shows the 3 (complex) parameters of freedom you really have, when you take out the overall scale factor of the usual A,B,C,D way of writing a LT. This form w = k [z - a] / [z - b] is very useful if you want to visualize how pieces of the domain map into pieces of the range.

The situation is well summarized in the picture on page 85 which also appears as half of a composite picture on the dust-jacket of Ahlfors book, and here is a little web shot with no rotation: (real k,a,b) which shows the level curves of the domain that we are talking about, a and b are the two foci:

domain (z) w = k [z - a] / [z - b] range (w)



Radial lines in the range (w space) map back into circles in the domain (z space) which pass through the two points a and b which I have just called **"thru" circles** like C_1 (vertical set in clip above). Concentric circles in the range map back into the **Circles of Apollonius** in the domain (like C_2 or the horizontal ones above) which are those circles I studied in "geometric problem.doc " which have the form |z-a| / |z-b| = d as was just mentioned in the last section above. Intersections of nearby C_1 circles (lunes) in the domain map into thin angular sectors in the range. As with any analytic mapping, angles are preserved and in this case, in either domain or range, the intersections of the two types of **"level curves"** are at right angles. The overall rotation of the domain picture is controlled by the phase of constant k. The set of circles shown in the domain are called **Steiner Circles** after Jakob Steiner (1796-1863), but are now referred to more usually as **bipolar coordinates**, as for example page 189 of my M&M book! Someday I will see how these separate the problem of the magnetic field of two parallel lines. I never really heard of them before this reading of Ahlfors! These are NOT parabolic coordinates which separate the hydrogen problem in QM.

Now suppose we map some Steiner circles into the concentric, then we do an inverse map on the concentric to get some different Steiner circles. The combination of these two LT's then maps one Steiner set into another. [Also, the inverse of the above shown transformation would map some Steiner set into a concentric range.] On bottom page 85 A shows a way to write a map that takes domain Steiner(a,b) into a range Steiner(a', b'). In this case, the constant k (not same as before) somehow affects the way the Steiner sets map into each other (the "flow"), easy to imagine. Here is that map:

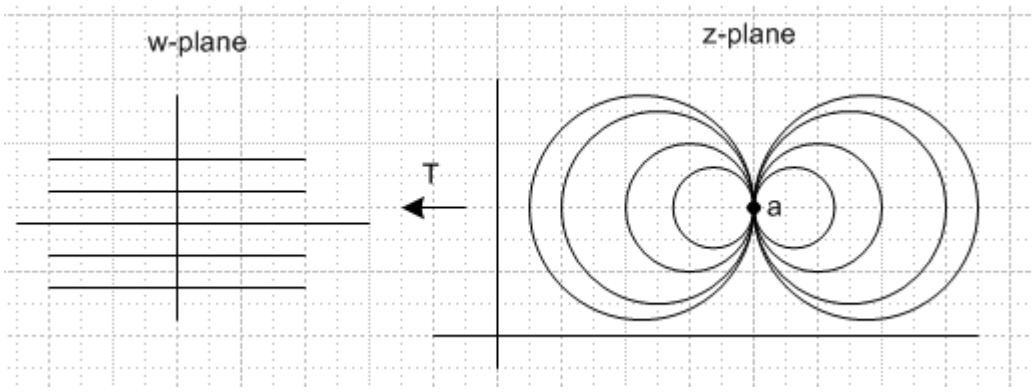
w = Tz such that (w-a')/(w-b') = k (z-a)/(z-b).

Notice that z = a maps into w = a' and the same for the other pair. In other words, a' = Ta and b' = Tb. If you *choose* a'=a and b'=b, then both these numbers a and b are "fixed points" of T, since then a = Ta and b = Tb. In this case, the two Steiner circle sets "look the same", they have the same foci. But of course k then affects the relative "flow" between them -- the way you would put parameter labels along the circles.

It turns out that if k > 0, then the C_1 circles map exactly, but they have different flows, and he calls this the **hyperbolic** situation, reason unknown. If |k| = 1, on the other hand, then the C_2's align, but they then have relative different flow, and he calls this the **elliptic** case.

Any LT has one pair of fixed points, because z = Tz is a quadratic equation. If it happens that a = b, you have the **parabolic** case.

He then does yet another LT form which is this: w = z/(z-a) + c . This is not a general form, it is a restricted form since only two parameters. Any *line* in the range hits w=∞ and must thus pass through z = a in the domain. Two parallel in the range meet *only* at ∞, so they meet only at z = a, and this means that two parallel lines map back into two circles tangent at z = a, very simple. We then get this kind of situation (I drew domain on the right this time)



So obviously a set of vertical lines gives a set of vertical circles -- what you would obtain by rotating the right picture 90 degrees. And so you get the picture on page 87 where there is only one "focus".

This single-focus circle set is the **degenerate Steiner circles**. As before, you can make a transformation that takes on degen-Steiner set to another one, one with a and the other with a'. And you can arrange to have a = a' and have the set map onto itself but with "flow" variation, just as in the regular Steiner case.

Comments on the above section: *Any* given LT can be written in various different ways, each of which suggests a way to look at the "level curves" in domain and range. We started with w = k [z - a] / [z - b] and this led to the two curve sets being Steiner and Concentric. But this same LT (if suitably restricted) can be written in the form w = z/(z-a) + c (different a) and they you can use Degenerate Steiner and Cartesian as your two curve sets. Any LT has one pair of fixed points. The two points could coincide, but I don't think this really relates to degen Steiner. It just means you have the parabolic case.

Notice that if you start with w = k [z - a] / [z - b] and take the limit a = b, you get w = k which is a single point in the w-plane, so somehow that entire concentric business shrinks to a point, not very useful. So this is not the right way to get the "degen Steiner" stuff.

This subject matter does not appear very useful to the passing student, because one needs to see a problem the requires this stuff. The notion of flow along the circles probably is important in certain situations, maybe improving accuracy. Ahlfors is just throwing out some fancy stuff so that the interested student who someday actually uses this stuff may remember something from this section. This reader is reminded of why "pure math" is sometimes not as interesting as "physics".

Using the form w = k [z - a] / [z - b], we know k>0 means **hyperbolic**, |k| = 1 means **elliptic**. If the two fixed points happen to be the same (they are not a and b!) , it is **parabolic**. If the fixed points are not the same and if k is neither on the positive real axis or the unit circle, the LT is said to be **loxodromic**. I think this might actually partition LT's into four bins, not quite sure.

At this point in my notes, I quote an Amazon reviewer's interesting comment about Ahlfors's book, and I quote the **table of contents of the 3rd edition** which is about 98% the same as my 2nd edition.

4. Elementary Conformal Mapping (89)

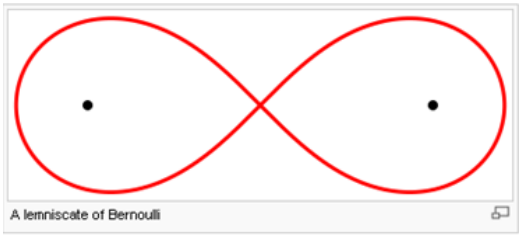
In the previous section we dealt only with mappings which were n=1 fractional linear transformations. With the usual condition on the coefficients, the mapping always had an inverse so was a 1-1 onto mapping of C into C with no confusion about branch points or any of that messy stuff. We looked at some examples of what I loosely called "level curves": how a domain grid of Circles relates to a range grid of Circles. We always had Circles (including lines) because LT's like circles.

{ Since a LT is a 1-1 onto map, you could say that it just rearranges the points on the complex plane C. An LT is thus an **automorphism of C**, and therefore **of the Riemann sphere** which is directly connected to the C plane. The set of LT's we already know forms a **group**, and you can think of this as Aut($\hat{C}$) which stands for **automorphism group** of the extended complex plane. So this is the thing (the group) that is isomorphic with GL(2,C). We also saw how this group has yet another name as the group of projections of the complex line. See wiki http://en.wikipedia.org/wiki/M%C3%B6bius_transformation for more. }

NOW we are going to look at analytic functions which are *not* linear transformations, such as $\mathbf{w = z}^2$. A whole new world of complexity arises now because we no longer have 1-1 and we no longer have Circles going into Circles.

In this section, the term **level curves** I think means you assume a Cartesian set of lines in the domain and examine the curves you get in the range (or vice versa). For example, with $w = z^2$ we know that $u = x^2 - y^2$ and $v = 2xy$ so Cartesian in the range w maps back to two different kinds of hyperbola shapes in the domain z, as shown page 91. As expected, all intersections are at right angles. On the other hand, if you reverse solve you get $u = -v^2/2x^2 + x^2$ and $u = +v^2/2y^2 - y^2$ so Cartesian in the domain gives parabolas in the range as shown page 92, and is our famous **parabolic coordinate system**.

Using the same function $\mathbf{w = z}^2$, we can consider a different set of "level curves" in my more general sense. If we put concentric circles centered at 1 in the range w, we get a family of weird **lemniscates** in the domain z which look like this: (it is a quartic, as shown p 91 A).



Then you would finish out a concentric system in the range using rays through w = 1, and those map into some set of hyperbolas and these of course will be orthogonal to the lemniscates. Hard to imagine how you would use this for something, but who knows! [could not find a web picture of this]

In the case of $w = z^2$, the upper half z plane maps into the entire w plane, so you need to sort of restrict your domain. For this restricted domain, $w = z^2$ is a single valued function, so we are happy. When $w = z^{1/2}$, however, a point in the domain maps into two places in the range, so no longer a function, and we have to start worrying about "sheets" of the function (in this case, domain sheets).

A then takes a look at $\mathbf{w = z}^3$ and page 91 and 92 pictures for z^2 now become those shown (in part) on page 93. The parabolic coordinates have become foliums of Descartes.

Finally A looks at $\mathbf{w = e}^z$. In this case, Cartesian in the domain maps into Concentric in the range about the origin, which is very easy to show. Write z = x+iy. If you vary only x, you are on a ray in the range. If you vary only y, you do a circle. If you take a strip of height π in the domain, it maps into a w half-plane, which we can then map into a circle by some LT. So combining e^z with LT let's you do conformal maps between strips and half planes or circles. This will be the game soon to come.

4.2 A survey of elementary mappings (93)

Now we start our study of mapping some region Ω_1 into another region Ω_2. The best approach is to find a mapping which takes each of these to a unit circle (or a half plane), then combine those mappings and you have your answer. So imagine a catalog of region shapes Ω_1 and how to take each shape into a disk or a half plane. The tools of the trade are the simple mappings we studied in the last section: e^z and ln(z) and powers and linear transformations. Since all are analytic and all do *some* Circles → *some* Circles, this method is generally only useful if Ω_1 is made up of pieces which are circular or straight, but we get some exceptions below.

Example 1: **Steiner lunes** map into angular sectors which can go to half plane by a power, so this gives a path from lunes to standard Ω using a LT then a power. Same for lens.

Example 2: **Boolean difference of two tangent nested circles** is a sort of crescent thing. But these are the things you get in degenerate Steiner, so we can map such a crescent to a strip and thence using expo into a half plane.

Example 3: Partial Steiner lune is a "**circular right triangle**", and we can get this to an angular sector out to some R, and then z^α can take that to a half plane.

Example 4: This is a combination of two LT's with $w = z^{1/2}$ which is summarized in (15) p 94, where it is shown in two stages. It happens that this mapping "maps the complement of a line segment into a circle", but that is not so important. What is important is that Concentric in range maps to **ellipses** and

hyperbolas in the domain, as shown bottom page 95. This lets us do some limited conic section mapping to standard shapes, so here is an example that takes you beyond things with piecewise circular boundaries.

Example 5: This example uses w = **general cubic in z**, and we get more ellipses and hyperbolas. I did not bother to study this example. But, as part of this example, I show this fact in the raw notes:

$w = z^n + z^{-n}$ maps circles radius r into ellipses with $a = (r^n + r^{-n})$ and $b = (r^n - r^{-n})$

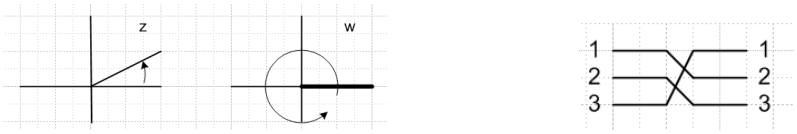
The 8 exercises ask the reader to find mappings from "some region" to a standard region.

4.3 Elementary Riemann Surfaces (97)

The "surface" is the combination of however many "sheets" there are (there could be an ∞ number). The domain and the range side each have a Riemann surface, though one usually is a clean single sheet.

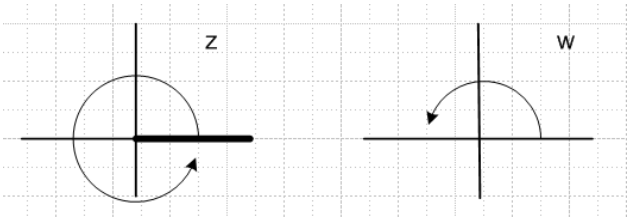
I am pretty good on this subject, so notes here will be minimal. Basically we just have a list of examples.

Example 1: $w = z^{12}$. Here a each 30 degree sector of domain maps into the range. The range then has a branch point at w = 0 and there are a total of 12 sheets and here is the mapping:



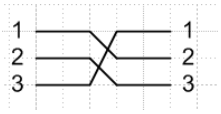
The angular sector shown maps into the entire w plane. It might be useful in the domain do demarcate each 30 degree sector with some heavy rays. Each such sector is a **fundamental region** since it is 1-1 with the entire w plane. I did not want to draw the "cross section" picture for n = 12, so I drew it instead for n = 3 (right side above). This is what the cut looks like viewed end-on (from the right, with the w plane tipped back 90 degrees from paper).

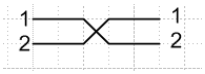
Example 1A. Consider $w = z^{1/n} = R^{1/n}\, e^{i\theta n/n}$. This is the same as $z = w^n$ so we just reverse the labels on the picture above. Now the branch cut is in the domain plane. Example $w = z^{1/2}$.



I think you would have to regard the fundamental region in this case as both sheets of the domain, since they are both needed to map into the entire w plane. One might define sheet 1 of the domain such that $\sqrt{4e^{i\varepsilon}} = +2$, so that z = 4 maps into w = +2. Then after we spin around once, we pick up $e^{i\pi} = -1$ and we then would say that $\sqrt{4e^{i\varepsilon}} = -2$ on this second sheet (which is near the arrow tip in the w plane).

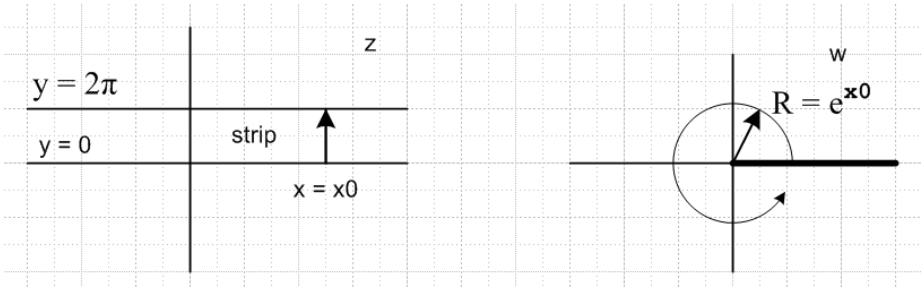
Example 1B: Consider $w = z^{2/3}$. In this case, you have a **branch cut in both planes**! If you cycle around 3 times in the domain, you go around 2 times in the range and it all starts over, and you can draw a nice picture showing this with some spirals. The domain and range cuts would have different cross sections:

domain = 

range = 

In the case $w = z^{\alpha}$, if α is a rational number such that α = P/Q, then it is like the $z^{2/3}$ case and the domain picture has Q layers, while the range one has P layers. On the other hand. $w = z^{\pi}$ has an infinite number of layers in both cross section pictures because you "never get back" to the starting point.

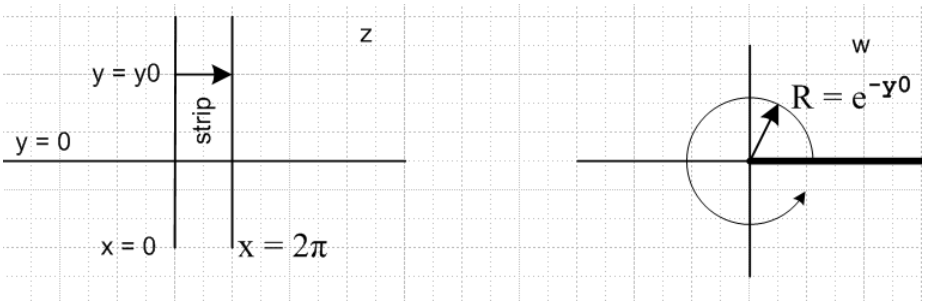
Example 2: $w = e^{z} = e^{x} e^{iy}$.



Here, each strip is a fundamental region. The x>0 half of the strip maps into circles R> 1.

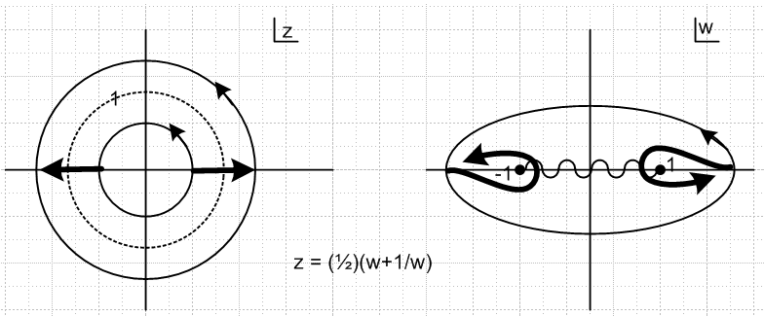
Example 2A: $w = w = e^{iz} = e^{ix} e^{-y}$

Same as above, but slightly different. The strips are now vertical strips. There are an infinite number of sheets on the Riemann Surface. Could demarcate fundamental regions with vertical bars in domain.

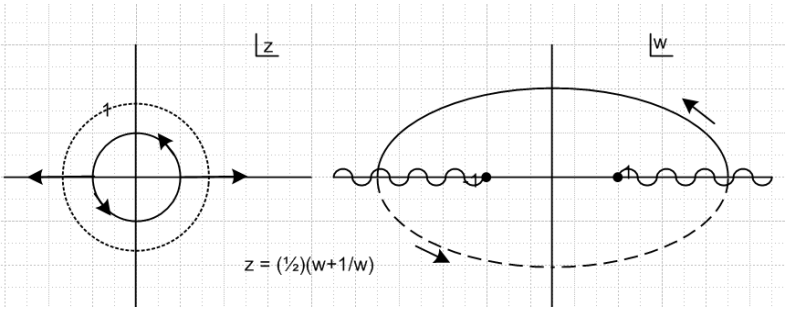


Example 3. $w = (1/2)(z + 1/z)$

This case is more complicated and you can read the raw notes. A nice picture is this:



which shows that two circles in z map into the same ellipse in w (but they are on different sheets in the range). The r<1 disk is a fundamental region of the domain, and r>1 is a second, so there are two sheets on the right. One easy way to see this is the inverse formula $z = w \pm \sqrt{w^2-1}$ for the two circles, and this shows the branch points directly. Raw notes show what happens if you distort the branch cut to its usual position. We get

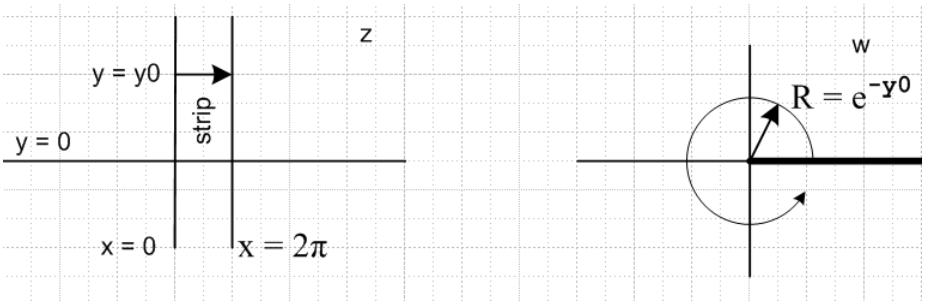


Example 4. $u = \cos(z) = 1/2 \left(e^{iz} + e^{-iz} \right)$

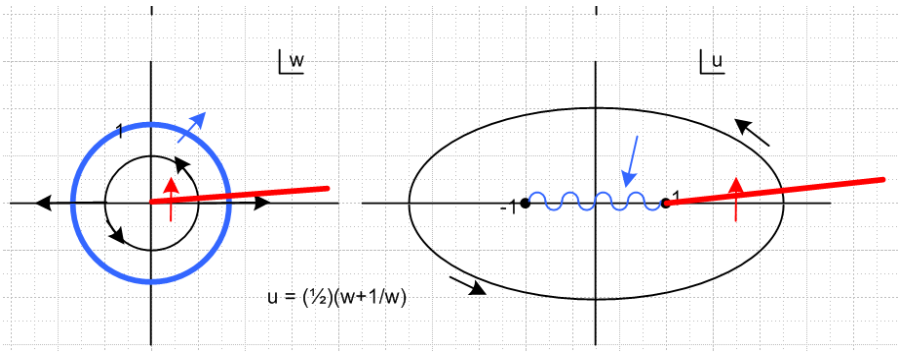
To understand this example, we need to consider two separate sequential transformations:

$u = (1/2) (w + 1/w)$ where $w = e^{iz}$

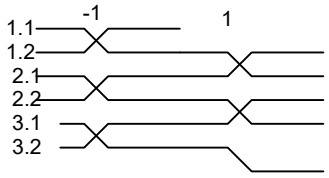
The first transform does this kind of thing'



and then the second does this



Overall, there are an infinite number of sheets, and the two branch points -1 and 1 have this combined cross section picture where the number before the dot is turns in the w plane.



Each vertical strip is a fundamental region. See raw notes for full details. Notice that this simple mapping takes circles into ellipses, so might be useful in potential theory.

You can see that Ahlfors has constructed his examples to have monotonically increasing complexity. He crafted his sequence very well, but he really needed 4x more pages to clearly present things! 2.5 pages was not enough for this subject.