PHI TECHNICAL MONOGRAPHS

Volume VII

Hyperbolic Geometry and Holomorphic Contraction

Geometry, analysis and applications



PHI research collective

Preface

A holomorphic map can stretch Euclidean lengths without bound and nevertheless contract the correct intrinsic metric. This observation connects Schwarz's lemma to the geometry of surfaces, to the stability of analytic iteration and to interpolation constraints in engineering. The purpose of this volume is to make those connections precise, with the constants, normalizations and limiting cases kept visible throughout the argument.

Volume V developed uniformization and geometric analysis; Volume VI studied extremal length and variational geometry. Here the principal source sequence is Section 7, "Hyperbolic metric and Schwarz Lemma," of Mikhail Lyubich's *Conformal Geometry and Dynamics of Quadratic Polynomials.* The supplied 702-page manuscript places that section on pages 133–145. We develop its themes into a sustained exposition rather than follow the economy of its research-level presentation. Local analysis leads to disk geometry; disk geometry leads to boundary estimates and quotient surfaces; normal families then explain what persists under limits. The final applications use the same mathematical mechanisms in several settings with different interpretations.

The text includes explicit proofs of the elementary analytic tools, the Schwarz and curvature routes to contraction, area and distortion estimates for univalent functions, the planar Riemann mapping theorem, the pointed kernel theorem, and the principal dynamical consequences developed here. General surface uniformization, the basic lifting theorems of covering spaces, and the named global surface theorems are classical inputs where stated. They have the same role as standard prerequisites in the preceding volumes: a named imported theorem is distinguished from an argument given in this book. This is an expository monograph, and no priority for the classical results or for the elementary calculations is asserted.

Some literal statements in the supplied source need a normalization or a quantifier corrected. We retain enough source correspondence to make those changes checkable, prove the replacement statements used here, and give exact counterexamples when a missing hypothesis matters. A strict inequality for each map is not a uniform inequality over all maps; a pointed limit needs a controlled basepoint; a model calculation does not establish its own physical applicability. These distinctions are part of the mathematics, not additional qualifications attached after the calculation.

The public and internal readings share one mathematical text. The internal material supplies teaching routes and discussion prompts. The light and dark PDFs contain the same text and numbering. Proof-assistant developments and computer-algebra certification are a separate subsequent programme; neither is represented as completed by the prose proofs or the illustrations in this volume. No physical measurements or hardware-performance results are reported.

How to read the volume

Readers approaching hyperbolic geometry for the first time should follow the analytic foundations and disk chapters in order. The univalence chapter is needed for boundary-distance comparisons and for compactness of normalized maps. The curvature chapter supplies another proof mechanism

and a bridge to geometric analysis. The symmetric-domain and boundary chapters explain which forms of contraction survive without a round model. The quotient-surface chapters introduce closed geodesics, cusps, collars and area. Convergence and dynamics then bring these estimates into families and iteration.

The normal-family chapter is followed by a complete existence proof for the normalized planar conformal map. Its local primitives, global logarithms, square-root construction and extremal derivative argument explain why the map exists before the kernel theorem studies its dependence on a changing domain. The quantitative cusp and depth estimates distinguish exact geometric statements from asymptotic approximations, with their error terms and extremizing examples given explicitly.

The applications can be read independently after disk contraction and the normal-family tools. Interpolation concerns the existence of one holomorphic function meeting several constraints. Passive-system examples interpret a Schur function as a reflection response. Iteration concerns a sequence of successive evaluations of a fixed map. Relativity uses the hyperbolic metric on a space of velocities; the physical velocity is not the same coordinate as the Poincaré disk radius. The worked investigations invite the reader to move between these viewpoints while retaining the assumptions of each model.

Conventions and prerequisites

Write $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, $\mathbb{H} = \{z : \operatorname{Im} z > 0\}$, and $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. A domain is a nonempty connected open set. Holomorphic means complex differentiable at every point of the domain; univalent means holomorphic and injective. Contours are piecewise C^1 unless a broader class is expressly specified. A conformal density $\lambda_U > 0$ defines lengths and distance by

$$L_U(\gamma) = \int \lambda_U(\gamma(t))|\gamma'(t)|\,dt, \qquad d_U(z,w) = \inf_{\gamma : z \rightsquigarrow w} L_U(\gamma).$$

All complete hyperbolic metrics have curvature -1:

$$\lambda_{\mathbb{D}}(z) = \frac{2}{1-|z|^2}, \qquad \lambda_{\mathbb{H}}(z) = \frac{1}{\operatorname{Im} z}, \qquad \|Df(z)\|_{U,V} = \frac{\lambda_V(f(z))|f'(z)|}{\lambda_U(z)}.$$

The norm is a norm between the indicated tangent metrics. An inclusion and a covering can have different norms if one changes the metric on their common underlying set. A density scaled by a factor c scales lengths by c and curvature by c^{-2}. We do not mix the curvature -1 and -4 conventions.

We use real completeness, compactness in finite-dimensional Euclidean spaces, the fundamental theorem of calculus, elementary measure integration and change of variables, the real inverse-function theorem, and the elementary topology of Jordan curves and covering spaces. Statements involving integration of a nonnegative series use monotone convergence; integration of uniformly convergent series on a compact contour uses uniform convergence. The Arzelà–Ascoli compactness theorem is used in the normal-family development. The surface-uniformization input says that a connected Riemann surface with disk universal cover inherits the disk metric through a holomorphic covering; the other simply connected conformal types are the plane and sphere. The spherical reference metric, when used, is the round unit-sphere metric.

The modulus of a round annulus $\{r < |z| < R\}$ is $\operatorname{mod} A = (2\pi)^{-1}\log(R/r)$. It is the extremal length of paths joining its boundary components. Its reciprocal is the extremal length of the core homotopy class. These two families must not be interchanged. A hyperbolic surface may be multiply connected, and its metric is defined by its universal covering, not by selecting an arbitrary planar chart.

Contents

Preface **i**

1 Analytic foundations for the disk arguments **1**
1.1 Triangle integrals from differentiability . . . 1
1.2 The integral formula and local series . . . 2
1.3 Maximum modulus, including its equality case . . . 4
1.4 Logarithms without assuming the desired contraction . . . 4
1.5 Removable singularities and counting zeros . . . 5
1.6 Local inversion with a complete zero-count argument . . . 6

2 Disk geometry and the Schwarz principle **7**
2.1 Automorphisms and an exact deficit . . . 7
2.2 Schwarz lemma with equality . . . 8
2.3 Distance and completeness . . . 8
2.4 Geodesics and half-plane coordinates . . . 9
2.5 Metric balls, area and exponential growth . . . 10
2.6 Equality, strictness and the lack of a universal deficit . . . 11
2.7 Higher-order zeros and quantitative rigidity . . . 12
2.8 Lifting to hyperbolic surfaces . . . 14

3 Univalent maps: area, distortion and conformal radius **15**
3.1 Why an injective holomorphic map has no critical point . . . 15
3.2 The exterior area theorem . . . 16
3.3 The second coefficient and the quarter theorem . . . 17
3.4 Changing the normalization point . . . 19
3.5 Compact disks, inverse estimates and normalization errors . . . 20
3.6 A useful link with extremal problems . . . 22

4 A second contraction proof from curvature **23**
4.1 Conventions and the differential calculation . . . 23
4.2 Comparison including critical points . . . 24
4.3 Contact rigidity by a circular-mean argument . . . 25
4.4 A finite-radius error estimate . . . 26
4.5 A model of negative-curvature target geometry . . . 27
4.6 Rescaling and an entire-map obstruction . . . 27

5 Symmetric maps: strictness and uniformity **28**
5.1 An explicit slit-plane coordinate . . . 28
5.2 Counterexample to a map-independent contraction gap . . . 29
5.3 A selected exact example . . . 30

6 Boundary geometry and definite contraction **31**
6.1 Normalization and the Koebe estimate . . . 31
6.2 A boundary asymptotic is not a global estimate . . . 31
6.3 Omitting one point gives an explicit sharp bound . . . 32
6.4 A distance buffer controls the metric ratio . . . 33

7 Symmetric neighborhoods and boundary comparison **35**
7.1 Symmetric Schwarz as an exact neighborhood statement . . . 35
7.2 An explicit angular estimate . . . 38
7.3 Punctured disks and isolated punctures . . . 38
7.4 A positive global comparison on the sphere . . . 40
7.5 A local version for an arbitrary smooth background metric . . . 42
7.6 Explicit definite estimates and their scope . . . 43

8 Hyperbolic surfaces and the geometry of a cusp **46**
8.1 From a local metric to a complete surface . . . 46
8.2 Deck transformations and closed geodesics . . . 47
8.3 Injectivity radius and local geometric scale . . . 48
8.4 A matrix estimate for an embedded cusp . . . 49
8.5 Coordinates, completeness, and decay in a cusp . . . 50
8.6 Solved investigations: quotient geometry . . . 51

9 Closed geodesics, collars, and conformal modulus **52**
9.1 Why a short closed geodesic needs room . . . 52
9.2 Normal coordinates around an axis . . . 52
9.3 Distance between two disjoint axes . . . 53
9.4 Separation of the lifts of a simple closed geodesic . . . 53
9.5 The embedded collar theorem . . . 55
9.6 Conformal modulus of the collar and cyclic cover . . . 55
9.7 Uniform expansion for a symmetric annular double covering . . . 58
9.8 Intersecting curves and simultaneous collars . . . 59
9.9 A strip geodesic and its Euclidean comparison . . . 59
9.10 Solved investigations: length and conformal width . . . 60

10 Area, thin regions, and quantitative geometric control **62**
10.1 Area is governed by topology . . . 62
10.2 Boundary length, doubling, and the limits of an area argument . . . 63
10.3 A precise classical input for the thin part . . . 64
10.4 Exact displacement tubes . . . 64
10.5 Finite area makes the thick part compact . . . 65
10.6 What a pinching collar becomes . . . 66
10.7 How far a closed geodesic can enter a cusp . . . 67
10.8 Moving endpoints and buffering a rectangle . . . 69
10.9 Solved investigations: area, degeneration, and buffers . . . 70

11 Normal families and pointed conformal domains **72**
11.1 The topology of local uniform convergence . . . 72
11.2 Zeros, injectivity and limits . . . 74
11.3 The basepoint and the stable component . . . 75
11.4 Conformal radius as a normalization parameter . . . 76
11.5 Worked investigations in analytic compactness . . . 77

12 The planar uniformization theorem with its existence proof **78**
12.1 From local primitives to a global logarithm . . . 78
12.2 Producing a bounded injective map . . . 79
12.3 The extremal derivative and surjectivity . . . 80
12.4 Conformal radius as an extremal quantity . . . 81
12.5 Solved investigations in existence and normalization . . . 82

13 Kernel convergence and maps on changing domains **83**
13.1 The normalized kernel theorem . . . 83
13.2 A quantitative inverse-map step . . . 84
13.3 A topology and a compactness theorem . . . 85
13.4 Distances and geodesics under kernel convergence . . . 86
13.5 An alternative proof for increasing domains . . . 88

13.6 Holomorphic functions on varying domains . . . 88
13.7 Worked investigations in changing geometry . . . 89

14 Holomorphic iteration and boundary attraction 91
14.1 An interior fixed point . . . 91
14.2 Linearizing an attracting fixed point . . . 92
14.3 Horodisks and a sequential Julia lemma . . . 95
14.4 From an invariant horodisk to an attracting point . . . 96
14.5 What boundary attraction does and does not say . . . 98
14.6 Stability of an attracting interior point . . . 98
14.7 Worked investigations in disk dynamics . . . 99

15 Harmonic analysis and the hyperbolic metric 102
15.1 Harmonic conjugates and the Poisson formula . . . 102
15.2 Harnack inequalities as metric statements . . . 103
15.3 Compactness of normalized positive harmonic fields . . . 104
15.4 Horocycles and a differential boundary coordinate . . . 105
15.5 The Bergman kernel from a complete orthogonal expansion . . . 106
15.6 Several values and an exact minimum-norm solution . . . 106
15.7 Conformal transfer and a kernel expression for the metric . . . 108
15.8 Dirichlet energy and conformal invariance in two dimensions . . . 109
15.9 Boundary flux and the cost of one Fourier mode . . . 110

16 Curvature equations, stability and geometric field models 112
16.1 The conformal connection and its curvature . . . 112
16.2 A strictly convex Dirichlet variational problem . . . 113
16.3 An explicit finite-boundary curvature problem . . . 114
16.4 A whole-domain error bound from a curvature residual . . . 115
16.5 An integral residual estimate . . . 116
16.6 Geodesic separation and the meaning of instability . . . 117
16.7 What the field interpretation contributes . . . 117

17 Finite interpolation and constructive function theory 118
17.1 The interpolation problem and its conventions . . . 118
17.2 Two elementary matrix facts . . . 119
17.3 Reduction of a function and reduction of its data . . . 119
17.4 The finite theorem, including singular cases . . . 120
17.5 What singular data force on the whole function . . . 121
17.6 One value and a prescribed derivative . . . 122
17.7 Worked three-point investigations . . . 123
17.8 Arithmetic work, conditioning and data perturbations . . . 124
17.9 The exact range of an unprescribed value . . . 124
17.10Solved exercises . . . 125

18 Passive systems, matching and analytic sensitivity 127
18.1 Positive real part and the reflection coordinate . . . 127
18.2 Hyperbolic sensitivity on the half-plane . . . 128
18.3 Finite matching constraints in impedance coordinates . . . 129
18.4 Propagation along an ideal lossless line . . . 129
18.5 Interior matching zeros and exact analytic constraints . . . 130
18.6 The ideal series-RC family . . . 131
18.7 The RC law in time and its stored energy . . . 132
18.8 Series and parallel composition within the analytic class . . . 133
18.9 An exact tolerance calculation . . . 133
18.10Solved exercises . . . 134

19 Controlled iteration and computational accuracy 136
19.1 An interior image produces a uniform factor . . . 136
19.2 Existence, uniqueness and usable stopping rules . . . 137

19.3 Inexact updates and stability of the answer . . . 137
19.4 Turning Euclidean evaluation bounds into admissible updates . . . 138
19.5 Analytic dependence of a uniformly attracting fixed point . . . 139
19.6 A nonaffine worked iteration . . . 140
19.7 An affine family and the cost of exact output . . . 140
19.8 Solved exercises . . . 141

20 Lorentz geometry, velocity space and rapidity 143
20.1 The hyperboloid and its disk coordinate . . . 143
20.2 Distance as relative rapidity . . . 144
20.3 The geodesic between two four-velocities . . . 145
20.4 Boosts and their fractional action . . . 146
20.5 Why noncollinear boosts are different . . . 147
20.6 Acceleration, energy and the boundary . . . 147
20.7 Solved exercises . . . 148

21 Source correspondence and mathematical perspective 150
21.1 The seven source subsections . . . 150
21.2 Which conclusions are classical inputs . . . 151
21.3 The applications and their boundaries . . . 151

References 152

Chapter 1

Analytic foundations for the disk arguments

This chapter proves the local complex-analysis tools used in the disk arguments from complex differentiability. The starting real-analysis facts are completeness of $\mathbb{R}$, compactness of closed bounded subsets of $\mathbb{R}^2$, the fundamental theorem of calculus for continuously differentiable paths, uniform convergence and its elementary integration rule, and the real second derivative test. These are the real-analysis prerequisites for the present development. Contours are piecewise continuously differentiable; their integrals are defined by parametrization. In particular $|\int_\gamma h(\zeta)\,d\zeta| \le \sup_\gamma |h|\ \mathrm{length}(\gamma)$.

1.1 Triangle integrals from differentiability

Theorem 1.1 (Goursat's triangle theorem)**.** *If f is complex differentiable on an open neighborhood of a closed triangle T, then its integral around the positively oriented boundary of T vanishes.*

Proof. Join the midpoints of the sides to form four similar subtriangles. Their oriented boundary integrals sum to the integral I of the original triangle because each interior edge occurs twice with opposite orientations. One subtriangle has integral of modulus at least $|I|/4$. Repeat the selection to obtain nested closed triangles T_n whose integrals I_n satisfy $|I| \le 4^n|I_n|$. Their diameters and perimeters are respectively $2^{-n}d$ and $2^{-n}p$, where d, p are those of T. Compactness and the diameters tending to zero give a common point z_*.

Differentiability at that point gives

$$f(\zeta) = f(z_*) + f'(z_*)(\zeta - z_*) + (\zeta - z_*)\epsilon(\zeta), \qquad \epsilon(\zeta) \longrightarrow 0.$$

The first two terms have polynomial primitives, so their integrals around every triangle vanish. Given $\eta > 0$, all sufficiently large T_n lie where $|\epsilon| \le \eta$. Consequently $|I_n| \le \eta dp\, 4^{-n}$ and $|I| \le \eta dp$. Let η tend to zero. No continuity of f' has been assumed. Degenerate triangles have zero integral by cancellation of collinear edges. □

Two features of the subdivision argument explain its effectiveness. First, no estimate on f' throughout a triangle is required. Only the single derivative at the common point is used. Second, the factor 4^n arising from selection is exactly cancelled by the product of the diameter and perimeter, each of which has acquired a factor 2^{-n}. Bounding the remainder by a fixed constant instead of by the arbitrarily small η would yield only a bound for the integral, not its vanishing. The proof also allows z_* to lie on an edge: differentiability is available on an open neighborhood of the entire original closed triangle.

Lemma 1.2 (One removable point for contour integrals)**.** *If h is continuous on an open set and complex differentiable there except possibly at one point w, its integral around every triangle contained with its interior in that open set is zero.*

Proof. When w is outside the triangle, apply Theorem 1.1. Otherwise cut the triangle by the four lines bounding a square of side 2ε centered at w. Triangulate the finitely many polygonal pieces outside the square. Their integrals vanish and cancel along shared edges. The original boundary integral is thus the sum of integrals along portions of the square boundary and, if necessary, the removed portions of the original triangle edges. The total length of these portions is at most $(8 + 6\sqrt{2})\varepsilon$. Continuity bounds $|h|$ on a fixed neighborhood of w, so the remaining integral tends to zero. This also covers a point on an edge or vertex. Every closed piece used outside the square is a positive distance from w. □

Proposition 1.3 (Primitives on convex sets)**.** *Suppose U is open and convex, and a continuous $h : U \to \mathbb{C}$ has zero integral around every triangle in U. Fix $b \in U$. Then $H(z) = \int_{[b,z]} h(\zeta)\, d\zeta$ is complex differentiable with $H'(z) = h(z)$. All integrals of h around closed contours in U vanish.*

Proof. The triangle with vertices $b, z, z + k$ gives $H(z + k) - H(z) = \int_{[z,z+k]} h$. For $k \neq 0$ the quotient equals $\int_0^1 h(z + tk)\, dt$ and tends to $h(z)$ by continuity. Along any piecewise smooth path the real chain rule applied to H gives the endpoint difference. For a closed path that difference is zero. □

The segment construction is useful because it exhibits the exact role of convexity. The three vertices $b, z, z + k$ and the triangle they span remain inside U, so the triangle identity compares the two prescribed segments with the short increment segment. For example, the estimate

$$\left| \frac{H(z + k) - H(z)}{k} - h(z) \right| \leq \sup_{0 \leq t \leq 1} |h(z + tk) - h(z)|$$

tends to zero independently of the direction of k. Thus the derivative is a complex derivative, not merely a derivative along a chosen real line. Once $H' = h$ is known and h is continuous, the usual real chain rule is applicable to $H \circ \gamma$ on each smooth piece of a contour. Summing the endpoint differences gives the assertion even for contours which self-intersect. A Jordan-curve theorem is unnecessary at this stage.

1.2 The integral formula and local series

Theorem 1.4 (Cauchy's formula on a disk)**.** *Let f be holomorphic on an open neighborhood of the closed disk $\overline{B(c, r)}$. For $|w - c| < r$,*

$$f(w) = \frac{1}{2\pi i} \int_{|\zeta - c| = r} \frac{f(\zeta)}{\zeta - w}\, d\zeta.$$

Proof. On a slightly larger disk put $h(\zeta) = (f(\zeta) - f(w))/(\zeta - w)$ for $\zeta \neq w$ and $h(w) = f'(w)$. Differentiability makes h continuous at w, and it is holomorphic away from w. Lemma 1.2 and Proposition 1.3 make its circle integral zero. The uniformly convergent geometric expansion

$$\frac{1}{\zeta - w} = \sum_{n=0}^{\infty} \frac{(w - c)^n}{(\zeta - c)^{n+1}} \quad (|\zeta - c| = r)$$

shows by direct parametrization that the integral of $1/(\zeta - w)$ is $2\pi i$: only the term $n = 0$ has nonzero integral. Rearrangement gives the asserted formula. □

Corollary 1.5 (Taylor series and coefficient bounds)**.** *Under these hypotheses, for $|z-c|<r$,*

$$f(z)=\sum_{n=0}^{\infty}a_n(z-c)^n,\qquad a_n=\frac{1}{2\pi i}\int_{|\zeta-c|=r}\frac{f(\zeta)}{(\zeta-c)^{n+1}}\,d\zeta,\qquad |a_n|\le\frac{M_r}{r^n},$$

where $M_r=\max_{|\zeta-c|=r}|f(\zeta)|$. The series and each of its termwise derivatives converge uniformly on every smaller closed disk; $a_n=f^{(n)}(c)/n!$.

Proof. Insert the same geometric expansion into Cauchy's formula. Uniform convergence permits integration term by term. The integral estimate gives the displayed bound. For $0<\rho<r$, each differentiated series is majorized by a constant times $\sum n^k(\rho/r)^n$, which converges by the ratio test. The elementary differentiation theorem for a series with uniformly convergent derivatives then gives the derivative identities. It follows in particular that a holomorphic function is real smooth. □

The real differentiation theorem just used follows by integrating the uniformly convergent derivatives along segments and applying the fundamental theorem of calculus. Thus the argument does not presuppose higher complex derivatives when establishing their existence.

Proposition 1.6 (Remainders and derivative estimates)**.** *With the notation of Corollary 1.5, let $0\le\rho<r$ and $N\ge 0$. Then*

$$\sup_{|z-c|\le\rho}\left|f(z)-\sum_{n=0}^{N}a_n(z-c)^n\right|\le\frac{M_r(\rho/r)^{N+1}}{1-\rho/r}.$$

If f is holomorphic near $\overline{B(a,s)}$ and has modulus at most M on its boundary, then $|f^{(k)}(a)|\le k!M/s^k$ for every integer $k\ge 0$.

Proof. The coefficient bound controls the modulus of the tail by $M_r\sum_{n=N+1}^{\infty}(\rho/r)^n$. Its geometric sum is the first display. Apply the same coefficient formula at center a and radius s and multiply the degree-k bound by $k!$ to obtain the second. In particular, derivative estimates at a point require a whole analytic disk around that point; a bound only along one line through it is insufficient. □

For a uniform bound on an interior disk, the available radius must be tracked. If $|f|\le M$ throughout $B(c,r)$, then for $|a-c|\le\rho<r$ and $s<r-\rho$ the disk $\overline{B(a,s)}$ lies in $B(c,r)$. The proposition applies there, and letting $s\uparrow r-\rho$ gives

$$\sup_{|a-c|\le\rho}|f^{(k)}(a)|\le\frac{k!M}{(r-\rho)^k}.$$

This estimate does not require boundary values at $|z-c|=r$. The smaller circles are sufficient. It makes the loss near the analytic boundary explicit and will later turn locally uniform function bounds into bounds for derivatives. For a difference $f-g$, exactly the same inequality turns an interior bound on the difference into one on its derivatives.

Investigation 1.7 (A certified Taylor truncation)**.** Suppose f is holomorphic near the closed unit disk and $|f|\le 1$ on its boundary. How large a degree suffices to approximate it to absolute error at most $1/1024$ throughout $|z|\le 1/2$ by its Taylor polynomial at zero?

Solution. The remainder bound is $2(1/2)^{N+1}=2^{-N}$. Thus $N=10$ suffices. This is a uniform analytic estimate, not a conclusion from agreement at a finite collection of test points. It is a sufficient degree for every function in the stated class; a particular function, such as a polynomial of degree two, may need far less. Without the boundary modulus hypothesis, the coefficient bound and therefore the stated uniform error need not hold.

Proposition 1.8 (Isolated zeros and identity theorem)**.** *If a holomorphic function on a connected open set has zeros accumulating at an interior point, it vanishes throughout that set.*

Proof. At any zero c, if some Taylor coefficient is nonzero, let m be the first such index. Then $f(z) = (z-c)^m u(z)$ with $u(c) \neq 0$; continuity makes u nonzero near c, so that zero is isolated. An accumulation point therefore has all coefficients zero and a neighborhood on which $f = 0$. The set of points with such a neighborhood is open. It is also relatively closed: a limit point either already belongs to the set, or has distinct zeros approaching it, and the same Taylor argument applies. Connectedness finishes the proof. Applying this to a difference proves the usual identity theorem for two functions. □

1.3 Maximum modulus, including its equality case

Theorem 1.9. *A nonconstant holomorphic function on a connected open set cannot have a local maximum of its modulus. If it is holomorphic near a closed disk, its maximum modulus on that disk is attained on the boundary.*

Proof. Suppose $|f(z)| \leq |f(c)| = M$ in a neighborhood of c. If $M = 0$ then $f = 0$ there, and the identity theorem applies. Otherwise multiply f by a constant of modulus one to arrange $f(c) = M > 0$. Cauchy's formula at the center says that $f(c)$ is the average of $f(c + re^{it})$ for each sufficiently small r. The continuous function $M - \operatorname{Re} f$ on that circle is nonnegative and has integral zero; hence it vanishes at every point. The constraints $\operatorname{Re} f = M$ and $|f| \leq M$ force $f = M$ on that circle. Varying r gives constancy on a neighborhood and then on the connected domain. Finally a continuous modulus has a maximum on the closed disk; an interior maximum forces constancy, so the boundary always supplies a maximizing point. □

The quotient argument for Schwarz's lemma now has all its local analytic ingredients: Taylor expansion removes its zero, this theorem supplies the bound and rigidity, and the radius limit is an ordinary real limit.

1.4 Logarithms without assuming the desired contraction

Proposition 1.10 (Logarithms on a convex zero-free domain)**.** *Let U be open and convex and let g be holomorphic and nowhere zero on U. There is a holomorphic L on U with $e^L = g$ and $L' = g'/g$.*

Proof. The quotient g'/g is holomorphic by the local power-series development. Take its primitive H from Proposition 1.3. Differentiating ge^{-H} gives zero. Its derivative along each segment is zero, so it is a nonzero constant A. Choose one complex number ℓ with $e^\ell = A$ (write A in polar form). Then $L = H + \ell$ has both desired properties. □

In particular the zero-free function used in the omitted-point estimate has a logarithm on $\mathbb{D}$, without appealing to general covering theory. For an arbitrary simply connected domain the corresponding construction requires path independence; that topological extension is not silently included in the convex-domain proposition.

The freedom in choosing a logarithm is also precise. If L_1, L_2 are holomorphic logarithms of the same nonzero function on a connected domain, then $e^{L_1 - L_2} = 1$. Hence $L_1 - L_2$ takes its values in $2\pi i\mathbb{Z}$. Continuity and connectedness force a single constant value. Thus specifying one logarithm value at one point fixes the branch everywhere. This observation justifies the

normalized square roots used in the area-theorem argument: $\exp(L/2)$ is a holomorphic square root, and fixing its value at zero fixes its sign.

Convexity cannot simply be dropped without replacement. On the punctured disk the function $g(z) = z$ is holomorphic and nonzero. If it had a single-valued holomorphic logarithm, the derivative of that logarithm would be $1/z$, whose integral around a centered circle is $2\pi i$. A derivative of a single-valued primitive has zero integral around every closed contour. The contradiction concerns global branch consistency; there is still a logarithm on each sufficiently small disk avoiding zero.

1.5 Removable singularities and counting zeros

Two further local tools will be used for puncture coordinates and limits of conformal maps. Both follow from the preceding series and contour arguments.

Proposition 1.11 (Bounded removable singularity)**.** *If h is holomorphic and bounded on a punctured disk $0 < |z - a| < R$, it extends holomorphically across a.*

Proof. Put $F(z) = (z - a)^2 h(z)$ away from a and $F(a) = 0$. Boundedness makes F continuous at a and gives $F'(a) = \lim_{z \to a}(z - a)h(z) = 0$. Hence F is holomorphic on the whole disk. Its Taylor coefficients of degrees zero and one vanish, so $F(z) = (z - a)^2 H(z)$ with H holomorphic. The function H agrees with h off a and supplies the required extension. □

Proposition 1.12 (Argument principle on a disk)**.** *Let f be holomorphic near the closed disk $\overline{B(a, R)}$, with no zeros on its boundary. If its zeros inside the disk, counted with multiplicity, number N, then*

$$\frac{1}{2\pi i} \int_{|z-a|=R} \frac{f'(z)}{f(z)}\, dz = N.$$

Proof. There are finitely many zeros inside: otherwise compactness would give an accumulation point in the closed disk, contradicted either by the nonzero boundary or by the identity theorem. Factor them out as $f(z) = \prod_j (z - z_j)^{m_j} g(z)$. The quotient extends at every z_j by its local Taylor factorization. It is nonzero on the closed disk and, by compactness and continuity, on a slightly larger disk. Proposition 1.10 gives a logarithm of g there, so g'/g has zero circle integral. Logarithmic differentiation leaves $\sum_j m_j/(z - z_j)$. Each term integrates to $2\pi i m_j$, by the geometric-series calculation used in Cauchy's formula. Their sum is N. □

Theorem 1.13 (Rouché on a disk)**.** *Let f, g be holomorphic near a closed disk and suppose $|g - f| < |f|$ on its boundary. They have the same number of interior zeros, counted with multiplicity.*

Proof. For $0 \le t \le 1$ put $f_t = f + t(g - f)$. On the boundary, $|f_t| \ge |f| - |g - f|$, whose minimum is positive by compactness. Thus every f_t has no boundary zero. The contour integral of f_t'/f_t, divided by $2\pi i$, depends continuously on t because its denominator is uniformly bounded away from zero there. By Proposition 1.12, its value is always an integer, the number of zeros of f_t. A continuous integer-valued function on an interval is constant. The endpoints $t = 0, 1$ give the assertion. □

For completeness, the continuity in t used here is uniform along the circle. Let $\delta = \min_{\partial B}(|f| - |g - f|) > 0$ and $A = \max_{\partial B}(|f'| + |g' - f'|)$. Every denominator has modulus at least δ, and subtraction of the two fractions gives a bound by a constant times $|t - s|$, with a constant depending only on δ, A and the boundary bounds for f, g. Integration preserves this continuity. The strict inequality is needed on the boundary alone; zeros of either function inside the disk are

precisely what is being counted. A multiple zero may split into several simple zeros under the perturbation, so multiplicity cannot be omitted from the statement.

1.6 Local inversion with a complete zero-count argument

Proposition 1.14 (A local holomorphic inverse)**.** *If f is holomorphic near a and $c = f'(a) \neq 0$, there are neighborhoods U of a and W of $f(a)$ such that $f : U \to W$ is bijective and its inverse g is holomorphic, with $g'(w) = 1/f'(g(w))$.*

Proof. Continuity of f' permits a radius $r > 0$ such that f is holomorphic near $\overline{B(a,r)}$ and $|f'(z) - c| < |c|/2$ throughout that disk. For z_1, z_2 in it, integrate $f' - c$ along the straight segment:

$$|f(z_1) - f(z_2) - c(z_1 - z_2)| \leq \tfrac{1}{2}|c|\,|z_1 - z_2|.$$

Consequently $|f(z_1) - f(z_2)| \geq |c||z_1 - z_2|/2$, so f is injective there. With $z_2 = a$ the same estimate shows that on $|z - a| = r$ the nonlinear remainder has modulus at most $|c|r/2$. For $|w - f(a)| < |c|r/4$, we therefore have

$$|f(z) - w - c(z - a)| < \tfrac{3}{4}|c|r < |c(z - a)|.$$

Rouché's theorem gives exactly one zero of $f - w$ inside the disk, counted with multiplicity, because $c(z - a)$ has one. Thus every w in $W = B(f(a), |c|r/4)$ has a unique preimage there. Take $U = B(a, r) \cap f^{-1}(W)$.

The lower difference estimate gives $|g(w_1) - g(w_2)| \leq 2|w_1 - w_2|/|c|$, so the inverse is continuous. For $w \neq w_0$ in W its difference quotient is

$$\frac{g(w) - g(w_0)}{w - w_0} = \left(\frac{f(g(w)) - f(g(w_0))}{g(w) - g(w_0)}\right)^{-1}.$$

As $w \to w_0$, continuity of g and differentiability of f give the reciprocal derivative in the statement. That derivative is nonzero because $|f' - c| < |c|/2$ on the chosen disk. This proves holomorphic inversion without assuming an inverse-function theorem as an additional complex-analytic input. □

Proposition 1.15 (Local degree and openness)**.** *If f is nonconstant on a connected domain and $f(a) = b$, let $m \geq 1$ be the order of the zero of $f - b$ at a. There is a holomorphic local coordinate v, with $v(a) = 0$ and $v'(a) \neq 0$, in which $f(z) = b + v(z)^m$. In particular f is open, and injectivity near a forces $m = 1$.*

Proof. The Taylor factorization gives $f(z) - b = (z - a)^m u(z)$, where u is holomorphic and nonzero on a sufficiently small disk. Choose a logarithm of u there and put $v(z) = (z - a)\exp(\log u(z)/m)$. Its derivative at a is nonzero, so Proposition 1.14 makes v a local coordinate. Its image contains a disk centered at zero. The mth-power map sends that disk onto another disk centered at zero, proving that the image of every sufficiently small neighborhood of a contains a neighborhood of b. Applying the argument at each point proves openness. If $m \geq 2$, a small nonzero number has m distinct mth roots in the coordinate disk, giving m distinct preimages. This contradicts local injectivity. □

A nonzero derivative gives a local inverse, not global injectivity: e^z is periodic. For an injective map, the local inverse branches agree on overlaps and define a holomorphic global inverse.

Rouché's strict boundary inequality gives uniform separation. Applied to compact target sets separated from an image circle, it preserves their inclusion in perturbed images. This is the analytic step connecting map convergence to kernel convergence, without requiring smooth limiting boundaries.

Chapter 2

Disk geometry and the Schwarz principle

The disk metric is adapted to the automorphism group rather than to a chosen Euclidean center. Its apparently singular density makes the boundary infinitely far away and allows every interior point to play the role of the origin. We derive the metric identities explicitly before using their geometric interpretations.

2.1 Automorphisms and an exact deficit

For $a \in \mathbb{D}$ put $m_a(z) = (z-a)/(1-\bar{a}z)$. The denominator is nonzero on $\mathbb{D}$, since $|\bar{a}z| < 1$. Expanding squared moduli gives

$$|1-\bar{a}z|^2 - |z-a|^2 = (1-|a|^2)(1-|z|^2). \tag{2.1}$$

Consequently $|m_a(z)| < 1$. Direct substitution shows that its inverse is $m_a^{-1}(w) = (w+a)/(1+\bar{a}w)$. Differentiation gives $m_a'(z) = (1-|a|^2)/(1-\bar{a}z)^2$. Combining this derivative with (2.1) proves

$$\frac{2|m_a'(z)|}{1-|m_a(z)|^2} = \frac{2}{1-|z|^2}.$$

Thus disk automorphisms preserve the metric pointwise and hence preserve path lengths and distances. Multiplication by a complex number of modulus one also preserves them.

Here the algebra can be checked without an appeal to a general theorem about fractional linear transformations. The two squared moduli in (2.1) are

$$|1-\bar{a}z|^2 = 1 - 2\,\mathrm{Re}(\bar{a}z) + |a|^2|z|^2,$$
$$|z-a|^2 = |z|^2 - 2\,\mathrm{Re}(\bar{a}z) + |a|^2.$$

Their identical mixed terms cancel. The remaining product is positive precisely because both points lie in the open disk. Dividing by $|1-\bar{a}z|^2$ gives the useful identity

$$1-|m_a(z)|^2 = \frac{(1-|a|^2)(1-|z|^2)}{|1-\bar{a}z|^2}.$$

The derivative has the same denominator in modulus, so it cancels in the metric ratio. To pass from a pointwise identity to distance, apply the chain rule to every piecewise smooth path. The inverse map gives a bijection between paths with the corresponding endpoints; taking infima therefore yields equality of distances, not merely one inequality.

2.2 Schwarz lemma with equality

Theorem 2.1 (Disk Schwarz lemma)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic with $f(0) = 0$. Then $|f(z)| \le |z|$ and $|f'(0)| \le 1$. Equality at a nonzero point in the first inequality, or equality in the derivative inequality, implies $f(z) = cz$ for a constant $|c| = 1$.*

Proof. The quotient $g(z) = f(z)/z$ extends holomorphically at zero with value $f'(0)$, by the power series of f. On $|z| = r < 1$, $|g(z)| \le 1/r$. The maximum modulus principle gives this bound inside the circle. For each fixed z, let $r \uparrow 1$ to obtain $|g(z)| \le 1$. If g attains modulus one at an interior point, including zero, the equality case of that principle makes g constant. The converse follows by direct substitution. □

Theorem 2.2 (Schwarz–Pick)**.** *For holomorphic $f : \mathbb{D} \to \mathbb{D}$ and $z, w \in \mathbb{D}$,*

$$|m_{f(w)}(f(z))| \le |m_w(z)|, \qquad \frac{(1-|z|^2)|f'(z)|}{1-|f(z)|^2} \le 1.$$

Equality for distinct points, or equality in the differential inequality at one point, makes f a disk automorphism.

Proof. For fixed w, the holomorphic map $g = m_{f(w)} \circ f \circ m_w^{-1}$ maps $\mathbb{D}$ into itself and fixes zero. Put $\zeta = m_w(z)$. The inequality $|g(\zeta)| \le |\zeta|$ is exactly the first assertion after substitution. For the derivative, the three factors in the chain rule give

$$g'(0) = m'_{f(w)}(f(w))\, f'(w)\, (m_w^{-1})'(0) = \frac{(1-|w|^2)f'(w)}{1-|f(w)|^2}.$$

Schwarz's derivative estimate now gives the second assertion at w; the point w was arbitrary. The denominator is strictly positive because the target is the open disk. At a critical point the numerator is zero, so no separate division by f' is involved.

If equality holds for $z \ne w$, then $\zeta \ne 0$ because m_w is injective. Schwarz's equality case gives $g(\zeta) = e^{i\alpha}\zeta$ for every ζ. Equality in the derivative bound has the same consequence. Solving the conjugation relation gives

$$f = m_{f(w)}^{-1} \circ (\zeta \mapsto e^{i\alpha}\zeta) \circ m_w,$$

a composition of explicitly invertible disk maps. Conversely each factor preserves the density and distance, so their composition gives equality. □

The same Schwarz lemma is a shared dependency of the two displayed inequalities; they are not counted as independent proof routes.

These arguments also classify all holomorphic automorphisms of the disk. If A is such an automorphism, then $h = m_{A(0)} \circ A$ fixes zero, as does its inverse. Schwarz's lemma applied to both gives $|h(z)| \le |z| = |h^{-1}(h(z))| \le |h(z)|$. Equality at any nonzero point makes h a rotation. Thus every automorphism has the fractional linear form already used. This classification is a consequence of the lemma; it has not been assumed to prove its equality case.

2.3 Distance and completeness

For a path from 0 to z, $|(|\gamma|)'| \le |\gamma'|$ almost everywhere. Integrating $2\,dr/(1-r^2)$ therefore gives the lower bound $2\operatorname{artanh}|z|$ for its length. The radial segment attains it. Automorphism invariance gives the exact formula

$$d_{\mathbb{D}}(z,w) = 2\operatorname{artanh}|m_w(z)| = \log\frac{1+|m_w(z)|}{1-|m_w(z)|}. \tag{2.2}$$

This derivation uses the chain rule for absolutely continuous radial modulus; piecewise smooth paths suffice for the distance definition. It also shows that the radial distance diverges at the boundary.

More explicitly, write $\Phi(r) = \log((1+r)/(1-r))$ and let a path $\gamma : [0,1] \to \mathbb{D}$ join zero to z. Its compact image lies in some disk of radius $r_0 < 1$, where Φ' is bounded. The function $\rho(t) = |\gamma(t)|$ is absolutely continuous, and

$$\begin{aligned} L_{\mathbb{D}}(\gamma) &= \int_0^1 \frac{2|\gamma'(t)|}{1-\rho(t)^2}\,dt \\ &\geq \int_0^1 \frac{2|\rho'(t)|}{1-\rho(t)^2}\,dt \geq \left|\int_0^1 \frac{2\rho'(t)}{1-\rho(t)^2}\,dt\right| = \Phi(|z|). \end{aligned}$$

The possible nondifferentiability of the modulus when $\gamma = 0$ is irrelevant to these almost-everywhere statements. For $\gamma(t) = tz$ both inequalities are equalities. Moving w to zero by m_w then proves (2.2) for all pairs, including the coincident case when both sides are zero.

A hyperbolic Cauchy sequence is Euclidean Cauchy, since $\lambda_{\mathbb{D}} \geq 2$. Its distance from zero is bounded. Formula (2.2) confines it to a smaller closed Euclidean disk, so its Euclidean limit is interior. On that smaller disk the density is bounded above, proving convergence in the hyperbolic metric as well. Thus $(\mathbb{D}, d_{\mathbb{D}})$ is complete. Integrating Theorem 2.2 along paths shows that every holomorphic self-map is nonexpanding for this distance.

The comparison of the two topologies in this proof is quantitative. For z, w in the closed Euclidean disk of radius $r_0 < 1$, their straight segment stays in that disk, and

$$2|z-w| \leq d_{\mathbb{D}}(z,w) \leq \frac{2}{1-r_0^2}|z-w|.$$

The lower inequality holds on the entire disk; the upper one requires the fixed interior radius. Given a hyperbolic Cauchy sequence (z_n), choose M with $d_{\mathbb{D}}(0, z_n) \leq M$ for every n. Such an M exists by the triangle inequality and the Cauchy property. The distance formula gives $|z_n| \leq \tanh(M/2) < 1$, supplying exactly the radius needed for the upper comparison. Mere boundedness in Euclidean distance would not exclude a limit on the unit circle.

2.4 Geodesics and half-plane coordinates

A shortest path from zero to a nonzero point must attain equality in the radial lower bound. Apart from a change of parametrization, it is the radial segment: equality forces the tangent to have the outward radial direction almost everywhere, so neither angular motion nor radial reversal is possible. Applying a disk automorphism proves uniqueness of the unparametrized shortest segment between any two distinct points. Complete geodesics are diameters and circular arcs meeting the unit circle orthogonally. Indeed fractional linear maps preserve circles and angles, carry diameters to such circles, and carry any point and tangent direction to a diameter through zero. This last description is Euclidean; orthogonality at the ideal endpoints does not mean that those endpoints belong to the metric space.

The equality conditions in the length calculation explain uniqueness. Away from $\gamma = 0$, decompose a tangent into its radial and angular components. Equality in $|\rho'| \leq |\gamma'|$ eliminates its angular component. Equality in the subsequent integral triangle inequality makes $\rho' \geq 0$ almost everywhere, since the total radial displacement is positive. Thus the direction cannot change while the radius is positive, and the radius cannot return to zero after becoming positive. Stationary pieces are harmless, but no different geometric trace can minimize. This reasoning proves uniqueness as an unparametrized segment; it does not require choosing constant speed at the outset.

There is also an explicit constant-speed parametrization. For distinct z, w put $\eta = m_w(z)$, $D = 2\operatorname{artanh}|\eta|$, and $\omega = \eta/|\eta|$. Then

$$\gamma(s) = m_w^{-1}(\omega \tanh(s/2)), \qquad 0 \le s \le D,$$

joins w to z. The inner radial path has hyperbolic speed one because $d(\tanh(s/2))/ds = (1 - \tanh^2(s/2))/2$; invariance transfers that speed to γ. Extending s through all real values gives a complete geodesic whose two limiting boundary points are approached only as $s \to \pm\infty$.

The Cayley map $C(z) = i(1+z)/(1-z)$ sends $\mathbb{D}$ to $\mathbb{H}$. Its derivative and imaginary part satisfy

$$|C'(z)| = \frac{2}{|1-z|^2}, \qquad \operatorname{Im} C(z) = \frac{1-|z|^2}{|1-z|^2}.$$

Consequently the pullback of $|dw|/\operatorname{Im} w$ is exactly the disk metric. Half-plane geodesics are vertical lines and semicircles centered on the real axis. For $u, v \in \mathbb{H}$, the distance formula becomes

$$\cosh d_{\mathbb{H}}(u, v) = 1 + \frac{|u-v|^2}{2 \operatorname{Im} u \operatorname{Im} v}. \tag{2.3}$$

To verify it, move v to i by a positive scaling and a real translation, apply the inverse Cayley map, and use $\cosh(2\operatorname{artanh} r) = (1+r^2)/(1-r^2)$. Real translations and positive scalings preserve the half-plane density, so the calculation gives the formula for all pairs.

For the intermediate algebra, write the normalized point as $U = X + iY$, $Y > 0$. Its inverse Cayley coordinate is $(U-i)/(U+i)$, so its squared modulus is

$$r^2 = \frac{X^2 + (Y-1)^2}{X^2 + (Y+1)^2}.$$

Substituting in $(1+r^2)/(1-r^2)$ gives $(X^2+Y^2+1)/(2Y) = 1 + |U-i|^2/(2Y)$. Taking $U = (u - \operatorname{Re} v)/\operatorname{Im} v$ recovers (2.3). In particular the formula is invariant under simultaneous real translation and positive scaling of both points, as the half-plane metric requires.

For points iy_1, iy_2 on a vertical geodesic, $d_{\mathbb{H}}(iy_1, iy_2) = |\log(y_2/y_1)|$. Additive hyperbolic distance is therefore multiplicative change of Euclidean height. On the disk's real diameter, the analogous additive coordinate is $2\operatorname{artanh} x$. These coordinates make exact examples far easier to compare than direct integration along a general circular arc.

2.5 Metric balls, area and exponential growth

A hyperbolic ball of radius R about zero is the Euclidean disk of radius $r = \tanh(R/2)$. A ball of radius R about a is the set $|m_a(z)| < r$. Clearing the positive denominator and completing the square shows that it is a Euclidean disk with center and radius

$$c = \frac{(1-r^2)a}{1-r^2|a|^2}, \qquad \rho = \frac{r(1-|a|^2)}{1-r^2|a|^2}. \tag{2.4}$$

Indeed the inequality $|z-a|^2 < r^2|1-\bar{a}z|^2$ is

$$(1-r^2|a|^2)|z|^2 - 2(1-r^2)\operatorname{Re}(\bar{a}z) + |a|^2 - r^2 < 0.$$

The coefficient $A = 1 - r^2|a|^2$ is positive. Dividing by A and completing the square gives the displayed center; the squared radius is

$$\frac{(1-r^2)^2|a|^2 - (|a|^2 - r^2)(1-r^2|a|^2)}{A^2} = \frac{r^2(1-|a|^2)^2}{A^2}.$$

This derivation also verifies the limiting cases $a = 0$ and $r = 0$. Its Euclidean center is generally different from its hyperbolic center. For $a \neq 0$ the Euclidean center lies farther from zero than a only if the formula says so; in fact $(1 - r^2)/(1 - r^2|a|^2) < 1$, so it lies closer to zero. The disk extends farther toward the Euclidean interior than toward the nearby ideal boundary, compensating for the larger metric density near that boundary.

The area element is $dA_h = 4\,dx\,dy/(1 - |z|^2)^2$. Integrating in polar coordinates gives

$$\operatorname{area}_h B_h(0, R) = \int_0^{\tanh(R/2)} \frac{8\pi t}{(1 - t^2)^2}\,dt = 4\pi \sinh^2(R/2) = 2\pi(\cosh R - 1).$$

The circumference is $2\pi r \lambda_{\mathbb{D}}(r) = 2\pi \sinh R$. Differentiating the area with respect to R gives the same circumference. For small R these are $\pi R^2 + O(R^4)$ and $2\pi R + O(R^3)$, agreeing with Euclidean geometry to first order. For large R both grow exponentially. This growth will later control how many disjoint metric balls can fit into a finite-area surface.

In the radial distance coordinate $R = 2\operatorname{artanh} r$, the disk line element is

$$ds^2 = dR^2 + \sinh^2 R\, d\theta^2.$$

To derive it, substitute $dr = (1-r^2)dR/2$ into $4(dr^2+r^2d\theta^2)/(1-r^2)^2$ and use $2r/(1-r^2) = \sinh R$. The formula explains why two rays with fixed angle separate much faster than in the Euclidean plane.

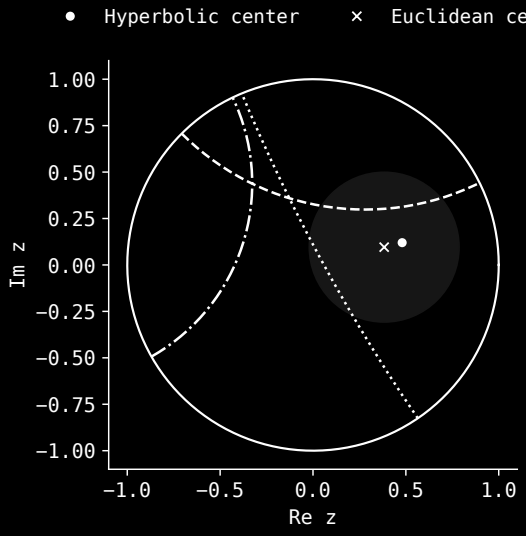


Figure 2.1: Disk geodesics are diameters or circular arcs orthogonal to the ideal boundary. The shaded metric ball has a hyperbolic center different from its Euclidean center, as in (2.4).

2.6 Equality, strictness and the lack of a universal deficit

If a disk self-map is not an automorphism, Schwarz–Pick says that its differential norm is strictly less than one at every point. On a fixed compact subset $K \subset \mathbb{D}$, continuity then gives $\sup_K \|Df\| < 1$. Both qualifications matter: the compact set is fixed, and the bound depends on the map. For $f_a(z) = az$, $0 < a < 1$, the norm at zero equals a and tends to one as $a \uparrow 1$. There can be no common deficit for all nonautomorphisms, even at a single fixed point.

The exact norm for this family is

$$\|Df_a(z)\| = \frac{a(1 - |z|^2)}{1 - a^2|z|^2} \leq a.$$

Subtracting its numerator from a times its denominator reduces the bound to $a(1-a^2)|z|^2 \geq 0$. Thus each map does have a global contraction constant; the failure concerns uniformity as the parameter varies. The boundary chapter will replace the qualitative exclusion of automorphisms by a controlled omitted point, producing an explicit parameter-dependent gap.

Investigation 2.3 (A fixed map with no global strict constant)**.** Show that $f(z) = z^2$ is strictly contracting in differential norm at every point of $\mathbb{D}$, but that the supremum of its norm on the whole disk is one. Find its optimal norm bound on $|z| \leq r < 1$.

Solution. At a point of modulus t, direct substitution gives

$$\|Df(z)\| = \frac{2t(1-t^2)}{1-t^4} = \frac{2t}{1+t^2}.$$

The difference from one is $(1-t)^2/(1+t^2) > 0$ for $t < 1$, but tends to zero as $t \uparrow 1$. Differentiation gives $2(1-t^2)/(1+t^2)^2 > 0$ on $[0,1)$, so the optimal bound on the closed radius-r disk is $2r/(1+r^2)$. The critical point at zero has norm zero and causes no difficulty. This example separates two failures of uniformity: a family of dilations loses its common gap as the map varies, whereas this single fixed map loses its gap as the observation set approaches the boundary.

An equality of endpoint distances is more rigid than an arbitrary small deficit. If $z \neq w$ and $d(f(z), f(w)) = d(z, w)$, then the increasing function in (2.2) and the two-point Schwarz–Pick inequality give equality in its pseudohyperbolic form. The equality theorem therefore makes f an automorphism. A map that preserves only the distance of a point to itself provides no such information; that equality is always zero.

2.7 Higher-order zeros and quantitative rigidity

Proposition 2.4 (Higher-order Schwarz lemma)**.** *If $f : \mathbb{D} \to \mathbb{D}$ is holomorphic and vanishes to order at least $m \geq 1$ at zero, then $|f(z)| \leq |z|^m$. Equality at a nonzero point holds only for $f(z) = cz^m$ with $|c| = 1$.*

Proof. The quotient $g = f/z^m$ extends holomorphically at zero. On the circle of radius r, $|g| \leq r^{-m}$. The maximum principle gives this bound inside, and $r \uparrow 1$ gives $|g| \leq 1$. An interior equality invokes the maximum principle again and makes g constant of modulus one. □

The condition "order at least m" means $f(0) = f'(0) = \cdots = f^{(m-1)}(0) = 0$. Its Taylor expansion then has $g(0) = f^{(m)}(0)/m!$, so the same proof gives the derivative estimate $|f^{(m)}(0)| \leq m!$. Equality in this derivative estimate also forces $f(z) = cz^m$ with $|c| = 1$, by the maximum principle applied to g at zero. If the zero has order strictly greater than m, then $g(0) = 0$; the equality possibility is therefore excluded unless the function is of the displayed exact order. No division at a zero is taken literally: throughout, the quotient denotes its holomorphic Taylor extension.

The bound applies just as well to a zero at a point $a \neq 0$ after moving that point to the origin. If $f : \mathbb{D} \to \mathbb{D}$ vanishes to order at least m at a, then

$$|f(z)| \leq |m_a(z)|^m.$$

Indeed $f \circ m_a^{-1}$ has a zero of the same order at zero because the derivative of the coordinate change is nonzero. Applying the higher-order lemma and substituting back proves the assertion. Equality at a point different from a forces $f(z) = c\, m_a(z)^m$. This is a model for how multiplicity and intrinsic distance interact: the coordinate m_a reflects the geometry, while the exponent reflects local analytic branching.

Investigation 2.5 (A double zero with prescribed location)**.** Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic with $f(1/2) = f'(1/2) = 0$. Bound $|f(0)|$ and identify all maps attaining the bound.

Solution. The shifted higher-order estimate with $m = 2$ gives $|f(0)| \le |m_{1/2}(0)|^2 = 1/4$. Since zero differs from the prescribed zero $1/2$, equality invokes the equality statement:

$$f(z) = c\left(\frac{z - 1/2}{1 - z/2}\right)^2, \qquad |c| = 1.$$

Conversely these maps are disk valued, have the required double zero, and have modulus $1/4$ at zero. A higher actual multiplicity would force a strictly smaller value there. The derivative condition and the value condition are both needed; a single zero alone gives only the larger bound $1/2$.

Suppose $f(0) = 0$ and write $f(z) = az + bz^2 + \cdots$. The holomorphic function $h = f/z$ maps into the closed disk. If $|a| = 1$, the equality case already makes $f(z) = az$. Otherwise h maps into the open disk, and Schwarz–Pick at zero gives

$$|b| = |h'(0)| \le 1 - |a|^2. \tag{2.5}$$

Unlike the coefficient bound for normalized univalent maps, this estimate uses a disk-valued range and does not assume injectivity. It is sharp: take h to be a disk automorphism with $h(0) = a$ and set $f = zh$.

Proposition 2.6 (Closeness to the first derivative)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic with $f(0) = 0$ and $f'(0) = a$, $|a| < 1$. For $|z| \le r < 1$,*

$$|f(z) - az| \le \frac{r^2(1 - |a|^2)}{1 - r|a|}.$$

For real $a \in [0, 1)$ this bound is attained at $z = r$ by an appropriate disk self-map. If $a \neq 0$, putting $u = a/|a|$ also gives

$$|f(z) - uz| \le (1 - |a|)r + \frac{r^2(1 - |a|^2)}{1 - r|a|}.$$

Proof. For $h = f/z$, Schwarz–Pick gives $|h(z) - a| \le |z|\,|1 - \bar{a}h(z)|$. The right side is at most $|z|(1 - |a|^2) + |z||a||h(z) - a|$. Since $1 - |z||a| > 0$,

$$|h(z) - a| \le \frac{|z|(1 - |a|^2)}{1 - |z||a|}.$$

Multiply by $|z|$ and use monotonicity in $|z|$. If a is real, $h(z) = (a - z)/(1 - az)$ gives equality in modulus at the positive real point r. The triangle inequality with $|a - u| = 1 - |a|$ proves the last bound. □

For each fixed $r < 1$, a derivative of modulus near one therefore forces uniform closeness to a rotation on $|z| \le r$. The bound deteriorates as $r \uparrow 1$, precisely where compact interior control no longer provides boundary control. This is a quantitative form of the equality case, with a specified norm and domain of comparison. It is stronger than the qualitative statement that every subsequential limit is a rotation, because it supplies an error from a measured derivative deficit.

Investigation 2.7 (A derivative deficit of one percent)**.** Assume $f(0) = 0$ and $f'(0) = 99/100$. Bound the distance from $f(z)$ to z throughout $|z| \le 1/2$, and compare this with the endpoint case $f'(0) = 1$.

Solution. The bound is

$$\frac{1}{200}+\frac{(1/4)(199/10000)}{1-99/200}=\frac{1}{200}+\frac{199}{20200}=\frac{3}{202}<\frac{3}{200}.$$

It is a uniform absolute error on the whole half disk, not a sampled error. At derivative one, Schwarz rigidity gives $f(z)=z$ exactly. The estimate connects these cases continuously on every fixed interior disk, but it does not claim that an arbitrary one-percent derivative error controls the entire closed unit disk.

Investigation 2.8 (An exact midpoint)**.** Find the hyperbolic midpoint between 0 and $x\in(0,1)$ on the real diameter and compare it with the Euclidean midpoint.

Solution. If m is the midpoint, then $2\operatorname{artanh} m=\operatorname{artanh} x$. The half-angle identity gives $m=x/(1+\sqrt{1-x^2})$. Since its denominator is less than two, $m>x/2$. More hyperbolic length is accumulated near x than near zero, so the midpoint lies farther along the Euclidean radius. For $x=3/5$, $m=1/3$; the two hyperbolic distances are both $\log 2$. Substitution in the distance formula verifies the result without approximation.

2.8 Lifting to hyperbolic surfaces

Suppose the uniformization and lifting theorems provide disk universal coverings $\pi:\mathbb{D}\to S$ and $\pi':\mathbb{D}\to S'$ with the quotient metrics. A holomorphic map $f:S\to S'$ has, after a basepoint choice, a disk lift $\widetilde{f}$ satisfying $\pi'\widetilde{f}=f\pi$. Applying Theorem 2.2 to the lift proves $\|Df\|\le 1$. If equality occurs, the lift is a disk automorphism. Its induced conjugation includes the source deck group in the target deck group, and the quotient map is a covering and a local isometry. Otherwise the differential inequality is strict everywhere.

To see each part of this passage, the topological lifting theorem applies to $f\circ\pi$ because its source disk is simply connected. The lift is holomorphic: locally it is the composition of $f\circ\pi$ with an inverse branch of the unramified covering π'. The quotient densities satisfy

$$\lambda_S(\pi(z))|D\pi(z)|=\lambda_{\mathbb{D}}(z),$$

where $D\pi(z)$ is the derivative in local coordinates. Together with the corresponding target identity and the differentiated lifting relation, this identifies $\|Df\|$ with $\|D\widetilde{f}\|$ at each lift. The statement is coordinate independent because the derivative factor in a conformal change is absorbed by the density.

In the equality case write $A=\widetilde{f}$ and let Γ,Γ' be the two deck groups. For $\gamma\in\Gamma$, the maps $A\gamma$ and A have the same projection to S'. The uniqueness of lifts gives $A\gamma=\gamma' A$ for some $\gamma'\in\Gamma'$. Since A is an automorphism, this is the inclusion $A\Gamma A^{-1}\subset\Gamma'$. The quotient of the disk by the smaller group maps surjectively to its quotient by the larger group. Evenly covered neighborhoods for the larger quotient show that this map is an unramified covering. Its local metric factor is one, so it is a local isometry. It need not be globally injective; equality for maps between surfaces has a covering conclusion rather than a universal biholomorphism conclusion.

For a proper connected inclusion $U\subsetneq V$ of hyperbolic surfaces, equality would make the inclusion a surjective covering; hence $\lambda_U>\lambda_V$ on U. If $p:U\to\Sigma$ is an unramified covering, $\|Dp\|_{V,\Sigma}=\lambda_U/\lambda_V>1$. This is the source's Corollary 7.1 with its topology and normalization stated. The surface statement uses the classical uniformization and covering-lift inputs specified in the preface.

Chapter 3

Univalent maps: area, distortion and conformal radius

Schwarz's lemma controls a map with a specified range. A univalent map may have an unbounded range, so an estimate depending only on the disk cannot hold until its value and derivative have been normalized. The family

$$\mathcal{S} = \{f : \mathbb{D} \to \mathbb{C} : \ f \text{ univalent}, \ f(0) = 0, \ f'(0) = 1\}$$

is the basic normalization. This chapter supplies the quantitative estimates for $\mathcal{S}$ used later in boundary geometry and convergence. The route is classical: an area calculation controls the second coefficient; a change of normalization converts that coefficient estimate into pointwise distortion. The geometric ideas are part of the standard univalent-function development; see the author lecture notes in [3] for further context.

3.1 Why an injective holomorphic map has no critical point

Let f be nonconstant and suppose $f(z) - f(a) = (z-a)^m u(z)$ near a, where $u(a) \neq 0$. On a small disk, u has a holomorphic logarithm. Hence $v(z) = (z-a)\exp(\log u(z)/m)$ is holomorphic with $v'(a) \neq 0$, and $f(z) - f(a) = v(z)^m$. Proposition 1.14 makes v a local coordinate. If $m \geq 2$, any sufficiently small nonzero value has m different preimages in that coordinate. This contradicts injectivity. Thus $m = 1$ for a univalent map, so its derivative never vanishes and its inverse on its image is holomorphic. This also proves that a nonconstant holomorphic map is open: the power map in the local coordinate sends neighborhoods to neighborhoods. These conclusions justify the inverse paths used below.

There is no circular use of the desired coefficient or distortion bounds here. The local inverse was constructed from a small derivative estimate, Rouché's theorem and the inverse difference quotient. The order m exists because a nonconstant holomorphic function cannot have every Taylor coefficient of $f - f(a)$ vanish. For a univalent function, the local inverses agree on overlaps by injectivity. They therefore define one holomorphic inverse on the entire image, even if that image is unbounded or has an irregular boundary. All paths used below remain strictly inside the original analytic domain, so no extension across the unit circle is assumed.

3.2 The exterior area theorem

Consider an injective holomorphic function on $\{\zeta : |\zeta| > 1\}$, including a simple pole at infinity, with expansion

$$F(\zeta) = \zeta + b_0 + \sum_{n=1}^{\infty} b_n \zeta^{-n}.$$

The leading coefficient one fixes the scale. For each $r > 1$, the curve $F(re^{it})$ is a smooth Jordan curve: injectivity gives simplicity, and the preceding local argument gives a nonzero tangent. The unbounded component is $F(\{|\zeta| > r\})$. Its bounded complement therefore has nonnegative area, regardless of how irregular the omitted set may become as r decreases to one.

Here the geometric inputs are the Jordan-curve theorem and Green's area formula for a smooth simple closed curve. The identification of the unbounded component can be justified as follows. Near infinity, $F(\zeta)/\zeta \to 1$, so the exterior image meets the unbounded component. It is connected and cannot cross $F(\{|\zeta| = r\})$ by injectivity. If a sequence of finite image points converges away from that curve, its preimages cannot escape to infinity, because $F(\zeta) \to \infty$ there, or approach the circle $|\zeta| = r$. A subsequence therefore converges in the exterior and its limit is a preimage of the limiting point. The image is both open and relatively closed in the unbounded component, and hence equals it. Orientation is positive around the bounded complement: it is positive near infinity, and the simple curves vary continuously with r without reversing their orientation.

Theorem 3.1 (Exterior area theorem)**.** *For every such F,*

$$\sum_{n=1}^{\infty} n|b_n|^2 \le 1.$$

Proof. Green's area formula for a positively oriented smooth Jordan curve is $\operatorname{area} E = (2i)^{-1} \int_{\partial E} \overline{w}\, dw$. On $|\zeta| = r > 1$ the Laurent series and its derivative converge uniformly. Insert $w = F(re^{it})$ in this formula. Different Fourier frequencies integrate to zero, the constant term contributes zero, and the frequency 1 contributes πr^2. The frequency $-n$ contributes $-\pi n|b_n|^2 r^{-2n}$. Thus

$$0 \le \operatorname{area} E_r = \pi \left(r^2 - \sum_{n=1}^{\infty} n|b_n|^2 r^{-2n} \right).$$

For each finite N, discard all terms with $n > N$ and let $r \downarrow 1$. This gives $\sum_{n=1}^{N} n|b_n|^2 \le 1$. Increasing N proves the assertion. The argument does not assume a rectifiable limiting boundary at $r = 1$, or an area formula on that limiting boundary. □

The frequency calculation is short enough to record explicitly. Write $w(t) = \sum_k c_k e^{ikt}$ on the circle, where $c_1 = r$, $c_0 = b_0$, and $c_{-n} = b_n r^{-n}$. For finite sums,

$$\frac{1}{2i} \int_0^{2\pi} \overline{w(t)} w'(t)\, dt = \frac{1}{2i} \sum_{j,k} \bar{c}_j (ik) c_k \int_0^{2\pi} e^{i(k-j)t}\, dt = \pi \sum_k k|c_k|^2.$$

The integral is zero when $k \ne j$ and 2π when $k = j$. Uniform convergence of both Laurent series on the radius-r circle allows passage to the limit in the product and its integral. Each negative frequency consequently has a negative signed contribution. The nonnegative series is handled at $r \downarrow 1$ through its finite partial sums; there is no unsupported interchange with a possibly divergent boundary series.

The sign of the negative frequencies is essential. On an exterior domain the principal term expands area, while the inverse powers subtract from the enclosed area. A bound on each coefficient alone would be weaker than the weighted sum. For example the theorem implies $|b_n| \le n^{-1/2}$, but these individual estimates cannot be added to recover the theorem.

Investigation 3.2 (The slit example)**.** For $F(\zeta) = \zeta + \zeta^{-1}$, determine its omitted set and check the equality case in the area formula for $r > 1$.

Solution. The identity $F(\zeta) - F(\eta) = (\zeta - \eta)(1 - (\zeta\eta)^{-1})$ proves injectivity when both moduli exceed one. The equation $F(\zeta) = w$ has roots of product one; off the real segment $[-2, 2]$ exactly one root has modulus greater than one. On that segment both roots are on the unit circle. Thus the omitted set is $[-2, 2]$. For $\zeta = re^{it}$, the image has coordinates $((r + r^{-1}) \cos t, (r - r^{-1}) \sin t)$, an ellipse of area $\pi(r^2 - r^{-2})$. This agrees with the theorem's intermediate formula, and tends to zero as $r \downarrow 1$. A limiting set of zero area need not be a single point.

3.3 The second coefficient and the quarter theorem

Theorem 3.3. *If $f(z) = z + a_2 z^2 + \cdots$ belongs to $\mathcal{S}$, then $|a_2| \leq 2$.*

Proof. Since f is injective and fixes zero, $f(z)/z$ is holomorphic and nonzero on the disk, with value one at zero. Choose its logarithm with value zero there, and define

$$g(z) = z \exp\left(\tfrac{1}{2} \log \frac{f(z^2)}{z^2}\right) = z + \frac{a_2}{2} z^3 + \cdots .$$

Then g is odd and $g(z)^2 = f(z^2)$. If $g(z) = g(w)$, injectivity of f implies $z^2 = w^2$. The possibility $w = -z$ forces $g(z) = 0$, hence $z = w = 0$. Thus g is univalent. The exterior function

$$F(\zeta) = \frac{1}{g(1/\zeta)} = \zeta - \frac{a_2}{2} \zeta^{-1} + \cdots$$

is univalent for $|\zeta| > 1$. The area theorem gives $|a_2/2|^2 \leq 1$. □

The branch selection in this construction is essential. The holomorphic function $f(\xi)/\xi$ has a removable value one at zero and no zero elsewhere, since a second zero of f would violate injectivity. Its logarithm $L(\xi)$ is chosen with $L(0) = 0$. The expression actually used is $g(z) = z \exp(L(z^2)/2)$, so the multiplier is an even function and g is odd. Squaring yields $g(z)^2 = f(z^2)$ on the whole disk, including zero by continuity. The two possible preimages after squaring are controlled by oddness; choosing pointwise square roots without this holomorphic branch would not establish an analytic univalent map.

The area argument also determines the equality case in this particular coefficient estimate. If $|a_2| = 2$, its exterior transform has $|b_1| = 1$. Every other term of the nonnegative area sum must then vanish. The transform is odd, so its constant coefficient vanishes too, and $F(\zeta) = \zeta + b_1/\zeta$. Taking reciprocals gives $g(z) = z/(1 + b_1 z^2)$ and hence

$$f(\xi) = \frac{\xi}{(1 + b_1 \xi)^2}, \qquad |b_1| = 1.$$

Every $\xi \in \mathbb{D}$ has a square root in $\mathbb{D}$, so the last identity holds everywhere, not only on a chosen ray. Thus equality yields the normalized rotations of the Koebe function. This conclusion uses only the degree-two coefficient estimate and its area proof; it does not invoke a general coefficient theorem.

Only the second coefficient is needed here. The estimate is not being extended without proof to the full coefficient theorem for $\mathcal{S}$. It is powerful because a normalization at another point produces a new second coefficient containing the logarithmic derivative of f.

Theorem 3.4 (Koebe quarter theorem)**.** *Every $f \in \mathcal{S}$ satisfies $B(0, 1/4) \subset f(\mathbb{D})$. More generally, if $f : \mathbb{D} \to U$ is conformal, $f(0) = a$, then*

$$B(a, |f'(0)|/4) \subset U.$$

The constant $1/4$ is best possible.

Proof. Suppose $w \notin f(\mathbb{D})$. Since $f(0) = 0$, $w \neq 0$. The map

$$h(z) = \frac{f(z)}{1 - f(z)/w} = z + (a_2 + 1/w)z^2 + \cdots$$

is a well-defined normalized univalent function: its final fractional linear map is injective off its pole. The preceding theorem applies to f and h, so $|1/w| \leq |a_2 + 1/w| + |a_2| \leq 4$. No point of modulus less than $1/4$ is omitted. Translation and scaling prove the second statement.

For sharpness, take $k(z) = z/(1 - z)^2$. If $k(z) = k(w)$, clearing denominators gives $(z - w)(1 - zw) = 0$, and $|zw| < 1$ forces $z = w$. Thus $k \in \mathcal{S}$. The value $-1/4$ is omitted, since the equation $4z = -(1 - z)^2$ reduces to $(1 + z)^2 = 0$, whose only root is on the unit circle. A larger universal centered disk is impossible. □

The proof concerns an *open* quarter disk. It does not assert that every point of modulus $1/4$ belongs to the image; the Koebe example shows exactly why such a boundary assertion would fail. The sign in the omitted-point transform can also be checked directly: expanding $(1 - f/w)^{-1} = 1 + f/w + \cdots$ gives the new coefficient $a_2 + 1/w$. Subtracting the two coefficient vectors gives $1/w$, and their two radius-two bounds give the radius-four bound. Using only the bound for the transformed map would leave the unknown a_2 uncontrolled.

The full image of the sharpness map has a transparent description. Put $q = (1 + z)/(1 - z)$, which maps $\mathbb{D}$ to the right half-plane. Then

$$k(z) = \frac{q^2 - 1}{4}.$$

Squaring maps the right half-plane bijectively onto $\mathbb{C} \setminus (-\infty, 0]$: every point in this slit plane has a unique square root with positive real part. Therefore $k(\mathbb{D}) = \mathbb{C} \setminus (-\infty, -1/4]$. The nearest omitted point is indeed $-1/4$, and the missing set extends along a full ray. Sharpness is therefore a geometric boundary phenomenon, not an accidental zero of a polynomial equation.

For a proper simply connected plane domain U and $a \in U$, choose its Riemann map $f : \mathbb{D} \to U$ with $f(0) = a$ and $f'(0) > 0$. Define the *conformal radius* $r_U(a) = f'(0)$. The only remaining freedom in a Riemann map fixing zero is a rotation, so the absolute derivative is intrinsic to this pointed domain. The real number $r_U(a)$ has Euclidean length units. It is not the hyperbolic distance to the boundary, which is infinite for the complete hyperbolic metric.

Corollary 3.5. *For $d = \operatorname{dist}(a, \partial U)$,*

$$d \leq r_U(a) \leq 4d, \qquad \lambda_U(a) = \frac{2}{r_U(a)}, \qquad \frac{1}{2d} \leq \lambda_U(a) \leq \frac{2}{d}.$$

Proof. The inverse Riemann map on $B(a, d)$, composed with $z \mapsto a + dz$, fixes zero and maps the disk to itself. Schwarz's derivative bound gives $d/r_U(a) \leq 1$. The quarter theorem gives $r_U(a)/4 \leq d$. The density formula is the pullback identity at zero. Substitution proves the last two inequalities, including their directions. □

3.4 Changing the normalization point

Fix $a \in \mathbb{D}$ and put $\varphi_a(z) = (a+z)/(1+\bar{a}z)$. For $f \in \mathcal{S}$ consider the Koebe transform

$$F_a(z) = \frac{f(\varphi_a(z)) - f(a)}{(1-|a|^2)f'(a)}.$$

It is normalized and univalent. Since $\varphi_a'(0) = 1-|a|^2$ and $\varphi_a''(0) = -2\bar{a}(1-|a|^2)$, its second coefficient is

$$\frac{(1-|a|^2)f''(a)}{2f'(a)} - \bar{a}.$$

The second-coefficient theorem therefore yields the disk-valued estimate

$$\left|(1-|a|^2)\frac{f''(a)}{f'(a)} - 2\bar{a}\right| \le 4. \tag{3.1}$$

The derivative $f'(a)$ is nonzero, so this expression is defined on all of $\mathbb{D}$. Notice that simply bounding $|f''/f'|$ by $4/(1-|a|^2)$ would lose the center term $2\bar{a}$ and would give the wrong sharp distortion constants.

To expose the normalization, differentiate the numerator twice at zero:

$$(f\circ\varphi_a)''(0) = f''(a)(1-|a|^2)^2 - 2\bar{a}f'(a)(1-|a|^2).$$

Division by the denominator of F_a and then by 2! gives precisely its second Taylor coefficient. The subtraction of $f(a)$ has no effect on derivatives, while the denominator makes $F_a'(0) = 1$. Multiplication of the bound $|[z^2]F_a| \le 2$ by two gives (3.1). Every normalization factor is retained; dropping $1-|a|^2$ at either derivative would change the radial rates.

Theorem 3.6 (Distortion and growth)**.** *For $f \in \mathcal{S}$ and $|z| = r < 1$,*

$$\frac{1-r}{(1+r)^3} \le |f'(z)| \le \frac{1+r}{(1-r)^3}, \qquad \frac{r}{(1+r)^2} \le |f(z)| \le \frac{r}{(1-r)^2}.$$

Each of the four bounds is sharp.

Proof. For $z = re^{i\theta}$ multiply (3.1) by $e^{i\theta}$ and take real parts. Since $\partial_r \log|f'(re^{i\theta})| = \mathrm{Re}(e^{i\theta}f''/f')$, we obtain

$$\frac{2r-4}{1-r^2} \le \partial_r \log|f'(re^{i\theta})| \le \frac{2r+4}{1-r^2}.$$

Integrate from zero to r, using $f'(0) = 1$. The left integral is $\log((1-r)/(1+r)^3)$ and the right integral is $\log((1+r)/(1-r)^3)$. Exponentiation proves distortion.

For the upper growth bound integrate along the radial segment:

$$|f(re^{i\theta})| \le \int_0^r \frac{1+t}{(1-t)^3}\,dt = \frac{r}{(1-r)^2}.$$

For the lower bound one cannot replace the modulus of an integral by the integral of the modulus. Instead choose a point w_0 of least modulus on $f(\{|z| = r\})$. The open segment from zero to w_0 lies in $f(B(0,r))$: otherwise it would first meet that domain's boundary at a smaller modulus. Its inverse under f is a smooth path γ from zero to some point of modulus r, including its endpoint since $r < 1$. As $f\circ\gamma$ traces a straight segment without reversals,

$$|w_0| = \int_\gamma |f'(z)|\,|dz| \ge \int_\gamma \frac{1-|z|}{(1+|z|)^3}|dz| \ge \int_0^r \frac{1-t}{(1+t)^3}\,dt = \frac{r}{(1+r)^2}.$$

The penultimate inequality follows by composing the increasing primitive of the radial weight with $|\gamma|$ and using $|(|\gamma|)'| \le |\gamma'|$. Every point on the image circle has modulus at least $|w_0|$. Finally $k(z) = z/(1-z)^2$ attains the upper bounds on the positive real radius and the lower bounds on the negative radius. Rotations give sharpness in any prescribed direction. □

Several details prevent misapplication of the growth argument. First, f is holomorphic and injective on a neighborhood of the closed radius-r disk because $r < 1$. Its image of that closed disk is compact, and its image of the circle is the boundary of $f(B(0, r))$. To verify the boundary statement, an image of an interior point is interior by openness, whereas a boundary point has nearby images from both sides of the source circle under a local inverse. Second, the least modulus $|w_0|$ is positive, since zero has the unique preimage zero. A radial segment of smaller modulus cannot leave the image without first meeting its boundary; such a first boundary point would contradict the choice of w_0. Third, the inverse curve is smooth up to its endpoint because the inverse derivative exists in a neighborhood of that point.

For the weighted path estimate, let $A(s) = \int_0^s (1 - t)/(1 + t)^3 \, dt = s/(1 + s)^2$. It is increasing on $[0, r]$. The chain rule and $|(|\gamma|)'| \le |\gamma'|$ give

$$\int_\gamma A'(|z|)|dz| \ge \left| \int_\gamma d(A(|z|)) \right| = A(r) - A(0).$$

There is no assumption that the inverse curve is radially monotone. Any radial backtracking only increases the first integral. Finally, $k'(z) = (1 + z)/(1 - z)^3$ verifies both derivative equalities explicitly, and the modulus of $k(\pm r)$ verifies both growth equalities.

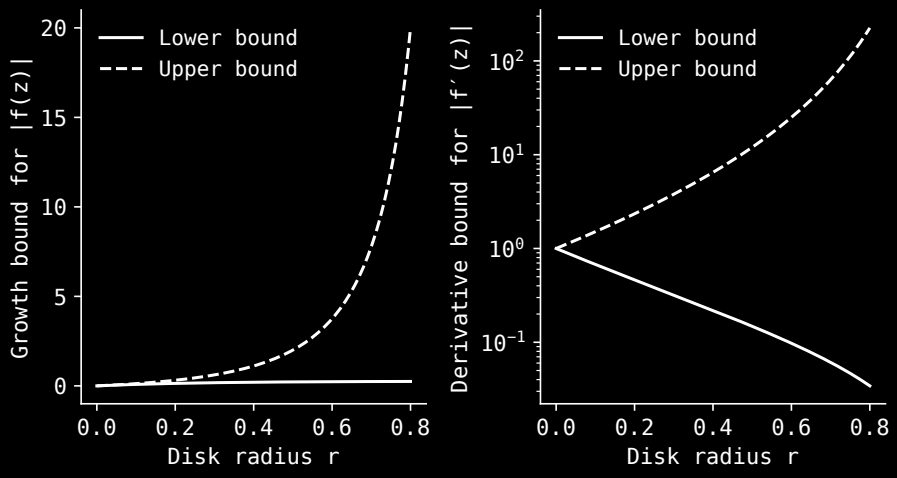


Figure 3.1: Sharp growth and derivative bounds for normalized univalent maps. The logarithmic vertical axis in the derivative panel shows the different boundary rates; all curves are formulas proved in the text.

The inverse-path argument illustrates a general principle: a lower bound for local length distortion becomes a lower bound for an endpoint only when the chosen image path has known length. Integrating along an arbitrary source radius would control the length of its image, which may be much larger than the distance between its endpoints.

3.5 Compact disks, inverse estimates and normalization errors

On $|z| \le r < 1$, the growth theorem bounds every normalized univalent function by $r/(1 - r)^2$. Cauchy estimates on an intermediate radius $r < R < 1$ therefore bound every derivative uniformly on the smaller disk. This is the quantitative input for compactness of $\mathcal{S}$ in the topology of locally uniform convergence. In the convergence chapter we will also prove that a nonconstant locally uniform limit of univalent maps is univalent, and will use the normalization $f'(0) = 1$ to exclude the constant alternative.

For a general univalent f on $B(a,R)$ define $F(\zeta) = (f(a+R\zeta) - f(a))/(Rf'(a))$. Applying the growth theorem gives, for $|z-a| = rR$,

$$\frac{rR|f'(a)|}{(1+r)^2} \le |f(z) - f(a)| \le \frac{rR|f'(a)|}{(1-r)^2}.$$

Thus the useful small parameter is the fraction of the available analytic radius, not the Euclidean displacement by itself. A fixed displacement approaching the boundary of the analytic domain cannot have a uniform distortion constant.

If x, y lie in the closed disk of radius rR about a, comparison of the derivative bounds gives

$$\frac{|f'(x)|}{|f'(y)|} \le \left(\frac{1+r}{1-r}\right)^4.$$

This estimate is convenient when the derivative is known at some point but the location of that point is not controlled. A smaller bound may be available when one point is the center, or when both points lie on a specified radius; the displayed statement deliberately covers any pair.

Proposition 3.7 (An interior inverse estimate)**.** *If $f \in \mathcal{S}$ and $0 < r < 1$, set $\alpha = r/(1+r)^2$. The image $f(B(0,r))$ contains $B(0,\alpha)$. On this latter disk the inverse $g = f^{-1}$ satisfies*

$$|g'(w)| \le \frac{(1+r)^3}{1-r}, \qquad |g(w_1) - g(w_2)| \le \frac{(1+r)^3}{1-r}|w_1 - w_2|.$$

Proof. The lower growth estimate on the radius-r circle places every point of its image boundary at distance at least α from zero. The same first-boundary-point argument used above puts the open centered disk inside $f(B(0,r))$. Hence $|g(w)| < r$ on that disk. Inverse differentiation and the lower distortion estimate give

$$|g'(w)| = \frac{1}{|f'(g(w))|} \le \frac{(1+|g(w)|)^3}{1-|g(w)|} \le \frac{(1+r)^3}{1-r}.$$

The last scalar function is increasing on $[0,1)$. The target disk is convex, so its straight segment from w_1 to w_2 remains within the region of the derivative bound. Integrating g' along that segment gives the difference estimate. □

The convex target disk matters: a derivative bound on an arbitrary nonconvex domain bounds change along paths by their lengths, and those lengths may be much larger than the endpoint distance. For $r = 1/2$ the proposition supplies the explicit target disk $|w| < 2/9$ and the inverse Lipschitz constant $27/4$. These constants are uniform over $\mathcal{S}$, not assertions of optimality for every individual map. They provide a controlled local inverse region even when the full image boundary is unknown.

Investigation 3.8 (A buffered interval)**.** Let L be a real interval of length $\ell > 0$, and suppose f is univalent on the disk centered at its midpoint with radius $(\delta + 1/2)\ell$, where $\delta > 0$. Prove that the derivative distortion on L is at most $((1+\delta)/\delta)^4$.

Solution. Every point of L is at relative radius at most $r = (1+2\delta)^{-1}$. Then $(1+r)/(1-r) = (1+\delta)/\delta$, and the preceding estimate applies. The fact that δ is strictly positive is essential: an interval with no analytic buffer can approach a singularity of the inverse or an endpoint of a slit. The estimate controls the oscillation of a derivative on an interval, not a universal distance-contraction deficit for every map of a slit domain.

Investigation 3.9 (Inner radius versus conformal radius)**.** Find domains showing that the two ends of $d \le r_U(a) \le 4d$ cannot be improved uniformly.

Solution. For $U = B(a, d)$ the map $z \mapsto a + dz$ gives $r_U(a) = d$. For the Koebe image $U = k(\mathbb{D})$ at $a = 0$, the quarter theorem gives $d \geq 1/4$ while the omitted point $-1/4$ gives $d \leq 1/4$. Thus $r_U(0) = 1 = 4d$. The domains giving equality are geometrically different: one is a disk, the other a slit plane. A statement that replaces conformal radius by Euclidean inner radius without a factor must therefore impose extra geometry.

3.6 A useful link with extremal problems

The conformal radius also solves an extremal derivative problem. Among holomorphic maps $h : U \to \mathbb{D}$ with $h(a) = 0$, the largest possible value of $|h'(a)|$ is $1/r_U(a)$. To see this, compose with a normalized Riemann map $f : \mathbb{D} \to U$: Schwarz's lemma gives $|h'(a)|r_U(a) \leq 1$, and the inverse Riemann map attains equality. The equality maps differ only by a unit complex factor.

Conversely, among univalent maps $g : \mathbb{D} \to U$ fixing a, the largest possible $|g'(0)|$ is $r_U(a)$, attained precisely by the conformal maps onto U. Indeed $f^{-1} \circ g$ fixes zero and its derivative has modulus at most one. Equality invokes the equality case of Schwarz's lemma. This extremal characterization explains monotonicity: if $U \subset V$ are proper simply connected domains containing a, every admissible g for U is admissible for V, so $r_U(a) \leq r_V(a)$. For a proper inclusion equality is impossible, since it would force the smaller map to be onto the larger domain. The corresponding hyperbolic densities satisfy the reversed inequality, with strictness at every point.

These extremal descriptions are useful even when no explicit Riemann map is available. The boundary-distance bounds supply rough scale information, the distortion theorem gives compactness, and a limiting extremal argument can then identify the map. The later kernel theorem makes the limiting step precise, including the basepoint normalization that prevents escape.

Chapter 4

A second contraction proof from curvature

The disk proof by conjugating to a zero and dividing by z is one route. This chapter gives a different route for the differential inequality: a real second derivative test applied to the ratio of conformal densities. The classical curvature method is due to Ahlfors; see [2]. The argument below fixes curvature -1, treats critical points explicitly and states exactly which equality conclusion is obtained by which route.

4.1 Conventions and the differential calculation

A positive twice continuously differentiable density σ on an open set $V \subset \mathbb{C}$ defines the conformal metric $\sigma(w)^2(du^2 + dv^2)$. With the Euclidean Laplacian $\Delta = \partial_u^2 + \partial_v^2$, its Gaussian curvature is

$$K_\sigma = -\frac{\Delta \log \sigma}{\sigma^2}.$$

This is the conformal-coordinate curvature formula, derivable by inserting the diagonal metric coefficients into the Levi–Civita connection. It can also serve here as the specified differential curvature condition; the comparison proof below uses only the displayed formula, not an unproved global curvature comparison theorem.

The coordinate calculation can be made explicit. Put $u = \log \sigma$, so the metric tensor is $g_{ij} = e^{2u}\delta_{ij}$ and its inverse is $g^{ij} = e^{-2u}\delta_{ij}$. Substitution in

$$\Gamma_{ij}^k = \tfrac{1}{2} \sum_\ell g^{k\ell}(\partial_i g_{j\ell} + \partial_j g_{i\ell} - \partial_\ell g_{ij})$$

gives

$$\begin{array}{lll} \Gamma_{11}^1 = u_x, & \Gamma_{12}^1 = u_y, & \Gamma_{22}^1 = -u_x, \\ \Gamma_{11}^2 = -u_y, & \Gamma_{12}^2 = u_x, & \Gamma_{22}^2 = u_y. \end{array}$$

With the convention $R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z$ for commuting coordinate fields, the first component of $R(\partial_x, \partial_y)\partial_y$ is $-u_{xx} - u_{yy}$; the products of first derivatives cancel. Its second component is zero. Taking its inner product with ∂_x gives $-e^{2u}\Delta u$, and dividing by the squared area $g_{11}g_{22} - g_{12}^2 = e^{4u}$ gives $K = -e^{-2u}\Delta u$. This specifies both the curvature sign and the scale used in the theorem.

Multiplication of a density by a positive constant c multiplies all lengths by c and changes curvature to K/c^2. In particular, $1/(1-|z|^2)$ has curvature -4, whereas the density $2/(1-|z|^2)$ used here has curvature -1. The factor of two cannot be moved between formulas without changing the associated curvature.

For $R > 0$, set $\lambda_R(z) = 2R/(R^2 - |z|^2)$ on $B(0, R)$. Writing $s = x^2 + y^2$, direct differentiation gives

$$\partial_x \log \lambda_R = \frac{2x}{R^2 - s}, \qquad \partial_{xx} \log \lambda_R = \frac{2}{R^2 - s} + \frac{4x^2}{(R^2 - s)^2}.$$

Adding the analogous y derivative yields

$$\Delta \log \lambda_R = \frac{4R^2}{(R^2 - s)^2} = \lambda_R^2. \tag{4.1}$$

Thus this metric has curvature -1 at every point of the open disk. Neither the boundary singularity nor a nonpositive R is in its domain.

Lemma 4.1 (Conformal Laplacian chain rule)**.** *If f is holomorphic and H is real C^2 near $f(z)$, then $\Delta(H \circ f)(z) = (\Delta H)(f(z))|f'(z)|^2$. Where $f' \neq 0$, $\Delta \log |f'| = 0$.*

Proof. Write $f = u + iv$. The Cauchy–Riemann equations give $u_x = v_y$, $u_y = -v_x$. Their derivatives give $\Delta u = \Delta v = 0$. In the ordinary second derivative chain rule the terms containing $H_u \Delta u + H_v \Delta v$ vanish. The mixed term has coefficient $2(u_x v_x + u_y v_y) = 0$, and the two remaining coefficients are equal to $u_x^2 + u_y^2 = |f'|^2$. This proves the first assertion, also at a critical point. Near a point with $f' \neq 0$, Proposition 1.10 supplies a local holomorphic logarithm of f'. Its real part is $\log |f'|$ and is harmonic by the same Cauchy–Riemann calculation. The local logarithm is all that is needed; a global logarithm of f' is not asserted. □

For reference, the full second-derivative expansion in the first part is

$$\begin{aligned}\Delta(H(u, v)) = {} & H_{uu}(u_x^2 + u_y^2) + 2H_{uv}(u_x v_x + u_y v_y) \\ & + H_{vv}(v_x^2 + v_y^2) + H_u \Delta u + H_v \Delta v.\end{aligned}$$

The Cauchy–Riemann equations make the two squared-gradient coefficients equal, kill the mixed coefficient, and make both component Laplacians zero. This explains why a factor $|f'|^2$ appears rather than $|f'|$ or its reciprocal. At a critical point all first derivatives of u, v vanish and the formula still reads $0 = 0$; only the separate logarithm of f' needs the noncritical hypothesis.

4.2 Comparison including critical points

Theorem 4.2 (Curvature contraction)**.** *Let $V \subset \mathbb{C}$ be open with positive C^2 density σ satisfying $K_\sigma \leq -\kappa$ everywhere, where $\kappa > 0$ is constant. For every holomorphic $f : \mathbb{D} \to V$ and every $z \in \mathbb{D}$,*

$$\sigma(f(z))|f'(z)| \leq \frac{2}{\sqrt{\kappa}(1 - |z|^2)}.$$

Proof. Fix $0 < R < 1$, and put $\mu = \sigma(f)|f'|$ and $v_R = \lambda_R/\sqrt{\kappa}$. The continuous function μ is bounded on $\overline{B(0, R)}$: the image of that compact set is a compact subset of V, and f' and σ are continuous. The quotient $q = \mu/v_R$ on the open disk therefore extends continuously by zero to its boundary. Suppose q exceeds one somewhere. Its positive maximum is attained at an interior point z_0 with $f'(z_0) \neq 0$. Near z_0 the function $u = \log q$ is C^2 and has a local maximum. Hence $\Delta u(z_0) \leq 0$.

On this neighborhood the chain rule and the harmonic logarithm give

$$\Delta \log \mu = (\Delta \log \sigma)(f)|f'|^2 \geq \kappa \mu^2, \qquad \Delta \log v_R = \lambda_R^2 = \kappa v_R^2.$$

It follows that $\Delta u(z_0) \geq \kappa(\mu(z_0)^2 - v_R(z_0)^2) > 0$, a contradiction. Thus $\mu \leq v_R$ on $B(0, R)$. For fixed z, let R increase to one. Equation (4.1) gives exactly the asserted bound. Critical points

cause no logarithmic singularity in this argument because the hypothetical positive maximum is necessarily away from them. If f is constant, the left side is identically zero and the assertion is immediate. □

The smaller radius is indispensable to this maximum argument. On the unit disk, the function $\sigma(f)|f'|$ need not have bounded boundary values, and a supremum of its ratio with the disk density need not be attained. On $\overline{B(0,R)}$, by contrast, the map is defined on a neighborhood of the compact set and its image stays inside V. The comparison density tends to infinity at $|z| = R$, so the quotient has boundary value zero without any assumption on the behavior of f at the ideal unit circle. If its maximum were $M > 1$, the factor

$$\kappa(\mu^2 - v_R^2) = \kappa v_R^2(M^2 - 1)$$

would be strictly positive at that maximum. The contradiction is therefore strict even when the target curvature inequality is an equality. Finally, for a fixed z choose $R > |z|$ before passing to the limit; there is no comparison with a radius-R density outside its domain.

Taking $V = \mathbb{D}$, $\sigma = \lambda_1$ and $\kappa = 1$ proves the differential Schwarz–Pick inequality. Integration along piecewise smooth paths gives $L_{\mathbb{D}}(f \circ \gamma) \le L_{\mathbb{D}}(\gamma)$, and then $d_{\mathbb{D}}(f(z), f(w)) \le d_{\mathbb{D}}(z, w)$. Using the explicit disk distance and strict monotonicity of artanh yields the two-point pseudohyperbolic inequality as well.

This route to the inequality does not use Schwarz's lemma or its quotient proof. It shares the local regularity and logarithm consequences of complex differentiability developed in Chapter 1. A non-strict maximum test alone does not settle the equality case: at a maximum of value one, it yields $0 \ge 0$ rather than a contradiction. The next argument supplies the stronger local statement that is actually needed.

4.3 Contact rigidity by a circular-mean argument

Lemma 4.3 (A local strong minimum principle)**.** *Let w be a nonnegative real C^2 function near a point p in the plane. Suppose $\Delta w \le Cw$ there for a constant $C \ge 0$. If $w(p) = 0$, then w vanishes on a neighborhood of p.*

Proof. Translate p to zero and choose a closed disk in the stated neighborhood. Let $M(r) = (2\pi)^{-1} \int_0^{2\pi} w(re^{i\theta})\, d\theta$ be its circular mean. We have $M(0) = 0$ and $M'(0) = 0$; the latter follows by averaging the directional derivative of w at zero. For $r > 0$, the polar Laplacian and integration of the angular second derivative give

$$M''(r) + \frac{1}{r} M'(r) = \frac{1}{2\pi} \int_0^{2\pi} \Delta w(re^{i\theta})\, d\theta \le CM(r).$$

Multiply by r and integrate from zero to s. The term $rM'(r)$ tends to zero at zero, so $sM'(s) \le C \int_0^s tM(t)\, dt$. Integrating once more gives

$$M(s) \le C \int_0^s \frac{1}{u} \int_0^u tM(t)\, dt\, du.$$

Choose $r > 0$ so small that $Cr^2/4 < 1$, and write $M_* = \max_{0 \le t \le r} M(t)$. Nonnegativity of w gives $M_* \ge 0$. For every $s \le r$ the last display implies $M(s) \le Cs^2 M_*/4 \le Cr^2 M_*/4$. Taking the maximum yields $M_* \le (Cr^2/4) M_*$, which forces $M_* = 0$. A nonnegative continuous function with zero integral on a circle is zero at every point of that circle. Consequently w vanishes throughout the disk of radius r. If $C = 0$, the same argument works for any sufficiently small radius. □

The lemma uses an upper bound on Δw, together with nonnegativity and a zero value. No assertion that w itself is harmonic has been made. The coefficient can also be a locally bounded nonnegative function: replace it on a smaller disk by a finite upper bound before applying the lemma.

Proposition 4.4 (Equality of densities propagates)**.** *Under the hypotheses of Theorem 4.2, put* $\mu(z) = \sigma(f(z))|f'(z)|$ *and* $v(z) = 2/(\sqrt{\kappa}(1-|z|^2))$*. If* $\mu(p) = v(p)$ *at one point, then* $\mu = v$ *throughout* $\mathbb{D}$*. In this case* f' *has no zeros and* $K_\sigma(f(z)) = -\kappa$ *for every* $z \in \mathbb{D}$*.*

Proof. The contraction theorem gives $0 \le \mu \le v$. Let $E = \{z : \mu(z) = v(z)\}$. This set is nonempty by hypothesis and closed relative to the disk by continuity. At any point of E, $\mu = v > 0$, so $f' \neq 0$ on a neighborhood of that point. On this neighborhood define $w = \log(v/\mu)$. It is C^2, is nonnegative, and vanishes at the chosen point. The curvature calculation gives

$$\Delta w \le \kappa v^2 - \kappa \mu^2 = \kappa v^2 (1 - e^{-2w}) \le 2\kappa v^2 w.$$

The final inequality follows from $e^{-2w} \ge 1 - 2w$ for $w \ge 0$. On a smaller closed disk $2\kappa v^2$ is bounded. Lemma 4.3 therefore forces $w = 0$ near the point. Thus E is open as well as closed. Connectedness of $\mathbb{D}$ implies $E = \mathbb{D}$. The equality $\mu = v > 0$ excludes every critical point. Finally

$$\Delta \log \mu = -K_\sigma(f)\mu^2, \qquad \Delta \log v = \kappa v^2.$$

Since $\mu = v$, division by their positive square gives $K_\sigma(f) = -\kappa$ along the image, as asserted. □

For a disk target with its standard density, the proposition proves propagation of metric equality by curvature alone. The classification of the resulting map as a disk automorphism can then be read from the already proved equality case of Schwarz–Pick. These are two distinct steps: local density equality is the conclusion of the circular-mean argument, while the global automorphism identification uses the earlier disk classification. For a more general target, equality need not imply injectivity; a covering map may be a local isometry everywhere.

4.4 A finite-radius error estimate

The limiting step can be quantified without sampling a grid. For $|z| \le r < R < 1$,

$$\frac{\lambda_R(z)}{\lambda_1(z)} - 1 = \frac{(1-R)(R+|z|^2)}{R^2 - |z|^2} \le \frac{(1-R)(R+r^2)}{R^2 - r^2}.$$

Both denominators are positive, and the rational function on the left increases with $|z|^2$ because its derivative is $(1-R)(R+R^2)/(R^2-|z|^2)^2 \ge 0$. Thus the comparison by the radius-R metric converges uniformly on each fixed smaller closed disk, with the displayed deterministic upper error. For example $r = 1/2$, $R = 9/10$ gives an excess at most $23/112$. This is a whole-disk bound. A plotted or sampled value is unnecessary for the proof and cannot replace the condition $r < R$.

For a desired uniform excess $\varepsilon > 0$ on $|z| \le r < 1$, one can choose a radius without solving the exact rational inequality. If $R \ge (1+r)/2$, then $R - r \ge (1-r)/2$ and $R + r \ge 1/2$, so $R^2 - r^2 \ge (1-r)/4$. Also $R + r^2 \le 2$. The displayed bound is therefore at most $8(1-R)/(1-r)$. Choosing

$$1 - R \le \min\{(1-r)/2,\ \varepsilon(1-r)/8\}$$

is sufficient. The constants are deliberately elementary rather than optimal, but every denominator and parameter range is explicit. This turns the radius limit into a finite analytic error estimate whenever one is useful for a later approximation argument.

4.5 A model of negative-curvature target geometry

For the right half-plane $\{w : \operatorname{Re} w > 0\}$ the density $\sigma(w) = 1/\operatorname{Re} w$ has $\Delta \log \sigma = 1/(\operatorname{Re} w)^2 = \sigma^2$. The theorem gives, for every holomorphic $h : \mathbb{D} \to \{\operatorname{Re} > 0\}$,

$$|h'(z)| \le \frac{2 \operatorname{Re} h(z)}{1 - |z|^2}.$$

In particular at zero this is the derivative inequality used for the logarithm of an omitted-point map. It supplies a second route to that inequality, while its sharpness example is still checked by explicit substitution. The target need not have been globally uniformized for this local-density theorem. General hyperbolic surface metrics, on the other hand, still need their existence and coordinate compatibility established.

4.6 Rescaling and an entire-map obstruction

If $f : B(a, R) \to V$ is holomorphic and the target satisfies the same curvature bound, apply Theorem 4.2 to $F(\zeta) = f(a + R\zeta)$. The derivative is $F' = Rf'(a + R\zeta)$, and substitution gives

$$\sigma(f(z))|f'(z)| \le \frac{2R}{\sqrt{\kappa}\,(R^2 - |z - a|^2)}.$$

At the center this is $2/(\sqrt{\kappa}R)$. The radius is an available analytic scale; replacing it by an arbitrary large number is justified only when the map is defined on every such disk.

Corollary 4.5 (No nonconstant entire map into this target)**.** *Under the target hypotheses of Theorem 4.2, every holomorphic map $f : \mathbb{C} \to V$ is constant.*

Proof. Fix $a \in \mathbb{C}$. The rescaled estimate holds on $B(a, R)$ for every $R > 0$, so $\sigma(f(a))|f'(a)| \le 2/(\sqrt{\kappa}R)$. Letting $R \to \infty$ and using $\sigma(f(a)) > 0$ gives $f'(a) = 0$. Since a was arbitrary, integration along any segment shows that f has the same value everywhere. □

No completeness hypothesis on the target density was needed for this obstruction. Positivity, C^2 regularity and one fixed negative curvature upper bound were sufficient. The fixed value $\kappa > 0$ matters: a bound that deteriorated with the chosen source radius could make the final right side fail to tend to zero. For the disk target the conclusion is also a special case of the elementary bounded-entire argument; here its mechanism is exhibited through curvature comparison.

Chapter 5

Symmetric maps: strictness and uniformity

5.1 An explicit slit-plane coordinate

Let $\Omega = \mathbb{C} \setminus ((-\infty, -1] \cup [1, \infty))$ and define

$$J(w) = \frac{2w}{1+w^2}, \qquad h(z) = \frac{z}{1+\sqrt{1-z^2}}.$$

The square root is the branch with positive real part. It is defined on Ω: if $1-z^2$ were nonpositive real, writing $z = x+iy$ would force $xy = 0$; the only possible case has $y = 0$ and $|x| \geq 1$, excluded from Ω. Write $s = \sqrt{1-z^2}$. Then $|h(z)|^2 = |1-s|/|1+s| < 1$, since $\operatorname{Re} s > 0$. Algebra gives $J(h(z)) = z$.

For $|u|, |v| < 1$,

$$J(u) - J(v) = \frac{2(u-v)(1-uv)}{(1+u^2)(1+v^2)},$$

so J is injective. Furthermore

$$\operatorname{Im} J(w) = \frac{2 \operatorname{Im} w(1-|w|^2)}{|1+w^2|^2}.$$

Nonreal points in the disk map to nonreal points, and the real diameter maps increasingly onto $(-1, 1)$. Thus $J(\mathbb{D}) \subset \Omega$ and the inverse calculation proves that J and h are mutually inverse conformal maps.

At the origin the expression for h has no singularity: its denominator is two and $h(0) = 0$. Away from zero, rationalizing gives $h(z) = (1-\sqrt{1-z^2})/z$. These are two expressions for the same holomorphic branch, not a choice of sign made separately at each point. The identity $z^2 = (1-s)(1+s)$ gives

$$|h(z)|^2 = \frac{|z|^2}{|1+s|^2} = \frac{|1-s|}{|1+s|}.$$

Since $|1+s|^2 - |1-s|^2 = 4 \operatorname{Re} s > 0$, this is less than one. At zero the conclusion follows directly and is consistent with the same identity. For the other direction, the denominators of J do not vanish in $\mathbb{D}$, and $1-uv$ in its difference formula cannot vanish when $u, v \in \mathbb{D}$. These checks establish both the analytic domain and the injectivity before any metric calculation.

5.2 Counterexample to a map-independent contraction gap

Proposition 5.1 (A symmetric family)**.** *For $0 < a < 1$, put*

$$F_a(z) = J(ah(z)), \qquad r_a = \frac{2a}{1+a^2}, \qquad I'_a = (-r_a, r_a).$$

Then F_a is a real-symmetric univalent map $\Omega \to \mathbb{C} \setminus ((-\infty, -r_a] \cup [r_a, \infty))$, maps $(-1, 1)$ diffeomorphically onto I'_a, is not Möbius, and its differential norm at zero between the round disks with these diameters is

$$s(a) = \frac{a}{r_a} = \frac{1+a^2}{2}.$$

Proof. Univalence follows by composing h, a nonzero dilation and J. The imaginary-part identity shows that a nonreal input cannot map to the removed real rays. On the real interval, h and J are increasing, and the endpoint limits are $\pm r_a$. Their derivatives are positive there. The power series at zero is

$$F_a(z) = az + \frac{a(1-a^2)}{4} z^3 + O(z^5).$$

An odd Möbius map holomorphic at zero and fixing zero must be linear: write it $Az/(Cz+D)$ and compare its values at z and $-z$ to obtain $C = 0$. The displayed cubic coefficient is nonzero, so F_a is not Möbius. The densities at the centers of the two round disks are 2 and $2/r_a$, and $F'_a(0) = a$, giving the stated norm. □

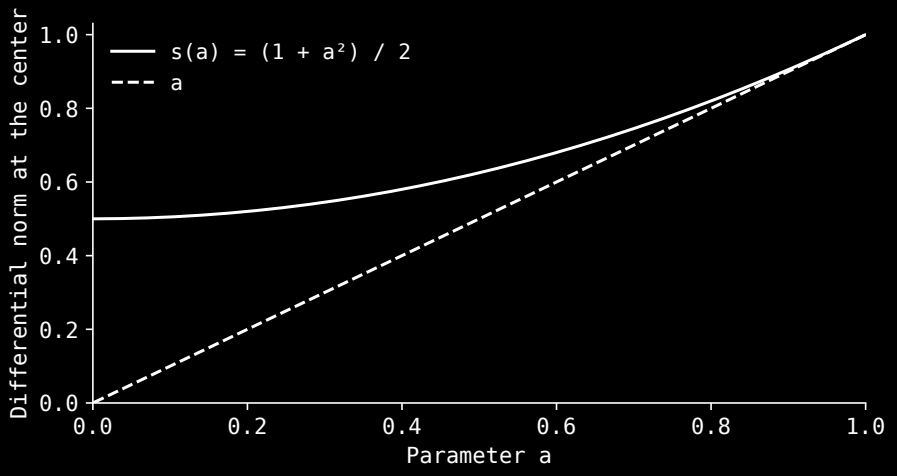


Figure 5.1: The exact scalar $s(a) = (1+a^2)/2$ approaches one within the non-Möbius family. The dashed line is a; their difference is $(1-a)^2/2$. The plot illustrates the written argument and does not replace its quantifiers.

The map takes its values in the asserted slit domain, but no surjectivity onto that slit domain is claimed. Its actual image is $J(a\mathbb{D})$, which is a proper subdomain when $a < 1$. To check the interval statement, h sends $(-1, 1)$ increasingly onto $(-1, 1)$ and dilation sends it onto $(-a, a)$. On the real diameter, $J'(t) = 2(1-t^2)/(1+t^2)^2 > 0$, so the final image is $(-2a/(1+a^2), 2a/(1+a^2))$. In the cubic calculation,

$$h(z) = \tfrac{1}{2}z + \tfrac{1}{8}z^3 + O(z^5), \qquad J(w) = 2w - 2w^3 + O(w^5),$$

which gives the displayed coefficient after composition. Its nonvanishing for $0 < a < 1$ is the reason every family member used in the counterexample is excluded from the Möbius class.

For any $0 < q < 1$ choose $a = q$. Then

$$s(a) - q = \frac{(1-q)^2}{2} > 0, \qquad s(a) < 1.$$

Fix any compact subinterval containing zero, for instance $[-1/2, 1/2]$. Its separation from the endpoints of $(-1, 1)$ does not vary with a. Thus a bound $\rho(\delta) < 1$ depending only on that fixed interior separation cannot bound the entire family. This refutes the map-independent uniform-gap assertion in the literal second part of source Corollary 7.2. It does not refute pointwise strictness for each fixed non-Möbius map.

A valid compact-set conclusion for a fixed map follows from continuity: if its norm is everywhere less than one, its maximum on a compact set is less than one. The constant then depends on the map. A family-uniform version needs a quantitative exclusion of the automorphism limit, such as an omitted target region controlled in the target hyperbolic metric.

5.3 A selected exact example

Take $a = 1/2$. Then $r_a = 4/5$, $s(a) = 5/8$ and the cubic coefficient is $3/32$. The norm exceeds a by $1/8$. At $a = 99/100$ it is $19801/20000$, already larger than $99/100$.

Chapter 6

Boundary geometry and definite contraction

6.1 Normalization and the Koebe estimate

Proposition 6.1 (Corrected conformal-disk bounds)**.** *For a proper simply connected plane domain U, let $d = \operatorname{dist}(z, \partial U)$. In curvature -1 normalization,*

$$\frac{1}{2d} \le \lambda_U(z) \le \frac{2}{d}.$$

Proof. The Euclidean disk $B(z, d)$ lies in U. Metric monotonicity and scaling give $\lambda_U(z) \le 2/d$. If $\psi : \mathbb{D} \to U$ is conformal with $\psi(0) = z$, then $\lambda_U(z) = 2/|\psi'(0)|$. Theorem 3.4 says $B(z, |\psi'(0)|/4) \subset U$, hence $d \ge |\psi'(0)|/4$. Substitution gives the lower bound. □

The direction of the two comparisons has a useful geometric check. Enlarging a hyperbolic domain gives holomorphic paths more room and decreases its density. The inscribed Euclidean disk therefore gives an upper bound for λ_U. Conversely, a Riemann map with a very large derivative at its center would, by the quarter theorem, place a large Euclidean disk in U. The actual boundary distance restricts that derivative from above and hence restricts the density from below. For $U = \mathbb{D}$ at zero, $d = 1$ and $\lambda_U(0) = 2$, attaining the upper endpoint. For the Koebe slit plane at zero, $d = 1/4$ and $\lambda_U(0) = 2$, attaining the lower endpoint $1/(2d)$. Both constants are sharp under precisely the stated simply connected plane-domain hypotheses. The source defines curvature -1 and disk density $2/(1 - |z|^2)$ on page 65, but its Lemma 7.7 uses density one at zero. Its displayed upper bound fails for $U = \mathbb{D}, z = 0$ in the earlier convention. The printed proof also reverses the Koebe inequality. Rescaling the metric can explain a factor of two; it cannot justify reversing the direction of set inclusion.

6.2 A boundary asymptotic is not a global estimate

The literal global lower bound in source Exercise 7.6 also needs a range restriction. Take $S = \mathbb{C}$, $K = \{0, 1\}$ and the Euclidean reference density, and let $z = it$ with $0 < t < 1$. Its Euclidean distance from K is t. The hyperbolic density of $\mathbb{C} \setminus \{0, 1\}$ is smooth and finite near i, whereas

$$\frac{\kappa}{-t \log t} \longrightarrow +\infty \quad (t \uparrow 1)$$

for every fixed $\kappa > 0$. Thus the proposed lower bound cannot hold at all these points, even with distance strictly below one. This argument uses the classical hyperbolicity of the twice-punctured plane as a declared uniformization dependency. The next chapter proves a small-distance replacement near compact removed sets and a globally positive comparison on the sphere.

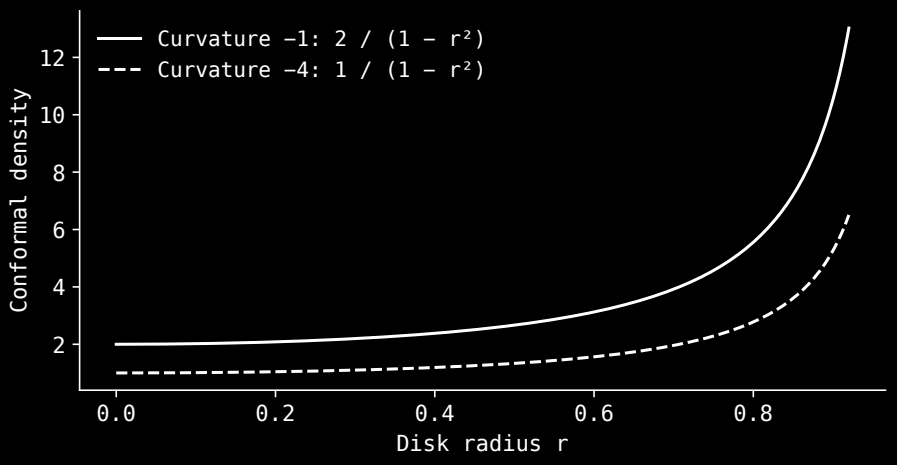


Figure 6.1: The two standard disk densities differ by a factor of two. This volume uses curvature -1 throughout. Constants copied between the conventions require corresponding changes.

6.3 Omitting one point gives an explicit sharp bound

Theorem 6.2. *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic, $f(0) = 0$, and suppose $b \notin f(\mathbb{D})$, where $0 < |b| < 1$. Then*

$$|f'(0)| \le B(|b|) := \frac{2|b| \log(1/|b|)}{1 - |b|^2} < 1.$$

For each such b, equality is attained by a holomorphic map with these properties. The function $B(r)$ is increasing on $(0, 1)$.

Proof. Set $g = m_b \circ f$. This map has no zero, maps to $\mathbb{D}$, and satisfies $g(0) = -b$. Proposition 1.10 supplies a holomorphic logarithm on the convex disk. Put $u = -\log g$, and write $u(0) = c + id$, where $c = \log(1/|b|) > 0$. Its real part is positive because $\operatorname{Re} u = -\log|g|$ and $0 < |g| < 1$. Normalize the right half-plane by $v = (u - id)/c$, so $v(0) = 1$, and put $H = (v - 1)/(v + 1)$. The denominator is nonzero and $|v + 1|^2 - |v - 1|^2 = 4 \operatorname{Re} v > 0$, so $H : \mathbb{D} \to \mathbb{D}$ is holomorphic with $H(0) = 0$. Differentiation at zero gives $H'(0) = u'(0)/(2c)$. Schwarz's derivative bound therefore yields $|u'(0)| \le 2c = 2\log(1/|b|)$.

The chain rule now has two complementary forms:

$$|g'(0)| = |g(0)||u'(0)| = |b||u'(0)|, \qquad |g'(0)| = |m_b'(0)||f'(0)| = (1 - |b|^2)|f'(0)|.$$

Combining them and dividing by the positive $1 - |b|^2$ gives the bound.

To see sharpness, write $c = \log(1/|b|) > 0$ and choose a phase θ with $e^{i\theta} = -b/|b|$. Set $g(z) = e^{i\theta} \exp[-c(1 + z)/(1 - z)]$ and $f = m_b^{-1} \circ g$. The Cayley transform has positive real part, so $0 < |g(z)| < 1$; the image omits zero and $g(0) = -b$. Direct differentiation gives equality.

Indeed the derivative of $(1 + z)/(1 - z)$ at zero is two, so $|g'(0)| = 2c|b|$. The value of g cannot equal zero because it is an exponential, and it cannot reach the unit circle because the real part of the exponent is strictly negative. Thus f is disk valued, $f(0) = m_b^{-1}(-b) = 0$, and f omits b since $m_b(b) = 0$. This checks every hypothesis for the equality example, not only its derivative.

Write $r = e^{-t}$, $t > 0$. Then $B(r) = t/\sinh t$. Since $\sinh t > t$, it is less than one. The derivative in t has numerator $\sinh t - t\cosh t$, whose derivative is $-t\sinh t < 0$ and whose value at zero is zero. Therefore B increases with r. □

The half-plane derivative inequality follows from Theorem 2.2 by a Cayley transform. This is a complete short argument relative to the stated logarithm and Schwarz dependencies. No new

priority is claimed. It supplies an explicit function for source Lemma 7.8, rather than deriving a false uniform gap from nonautomorphism alone.

The endpoint behavior is also informative. As $r \downarrow 0$, $B(r) \sim 2r\log(1/r) \to 0$: omitting a point close to the normalized center strongly restricts the derivative. As $r \uparrow 1$, set $t = \log(1/r) \downarrow 0$ to obtain $B(r) = 1 - t^2/6 + O(t^4)$. Omission far toward the ideal boundary can therefore give an arbitrarily weak strict gap. An exact lower bound for the gap follows from $\sinh t \geq t + t^3/6$:

$$1 - B(r) \geq \frac{t^2}{6 + t^2}, \qquad t = \log(1/r) > 0.$$

The power-series terms of $\sinh t$ are nonnegative for $t > 0$, so this estimate holds on the whole stated range, not merely asymptotically.

Corollary 6.3 (Omission at an arbitrary basepoint)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic, let $a \in \mathbb{D}$, and suppose $b \in \mathbb{D} \setminus f(\mathbb{D})$. Put $\rho = |m_{f(a)}(b)|$. Then $0 < \rho < 1$ and*

$$\|Df(a)\| \leq B(\rho).$$

In particular, if $d_{\mathbb{D}}(f(a), b) \leq R$ for some $R > 0$, then $\|Df(a)\| \leq B(\tanh(R/2)) < 1$.

Proof. The normalized map $F = m_{f(a)} \circ f \circ m_a^{-1}$ fixes zero and omits $\beta = m_{f(a)}(b)$. The omitted point differs from zero because $b \neq f(a)$, and the automorphism sends it into the disk. The chain-rule computation in Schwarz–Pick gives $|F'(0)| = \|Df(a)\|$. Apply Theorem 6.2 to F and β. The distance formula gives $\rho = \tanh(d_{\mathbb{D}}(f(a), b)/2)$; monotonicity of both scalar functions proves the final assertion. □

For a fixed omitted point and a fixed compact set of source points, the continuous distance $d_{\mathbb{D}}(f(a), b)$ has a finite maximum. The corollary therefore supplies an explicit strict bound throughout that compact set. For a family of maps it supplies a common bound only if a common finite distance upper bound has actually been established. Omission is a sufficient geometric condition, not a consequence of nonautomorphism: the map $z \mapsto z^2$ is onto the disk and is not an automorphism.

Investigation 6.4 (A specified omitted radius)**.** Suppose $f(0) = 0$ and f omits a point of modulus $1/2$. Give a strict derivative bound and exhibit a map attaining it.

Solution. The bound is $B(1/2) = 4\log 2/3 < 1$. Its strictness follows either from the proved $t/\sinh t < 1$, or from $\sinh(\log 2) = (2 - 1/2)/2 = 3/4 > \log 2$. After a rotation it suffices to omit $b = 1/2$. Set $g(z) = -\exp[-(\log 2)(1 + z)/(1 - z)]$ and $f(z) = (g(z) + 1/2)/(1 + g(z)/2)$. The preceding construction verifies that it fixes zero, omits $1/2$, and has derivative modulus $4\log 2/3$. The extremal map need not be univalent; injectivity was not among this theorem's hypotheses.

6.4 A distance buffer controls the metric ratio

Proposition 6.5. *Let $U \subset V$ be hyperbolic Riemann surfaces with their complete curvature -1 metrics. If $L = d_V(x, V \setminus U) > 0$, then*

$$1 \leq \frac{\lambda_U(x)}{\lambda_V(x)} \leq \coth(L/2).$$

If $L \geq \log 2$, the ratio is at most $1 + 4e^{-L}$.

Proof. Let $p : \mathbb{D} \to V$ be a universal covering with $p(0) = x$, and let W be the component of $p^{-1}(U)$ containing zero. The restriction $p : W \to U$ is a covering onto the component of U containing x: lift paths from x to any other point of that component. Their lifts stay in $p^{-1}(U)$ and hence in W. By the quotient-metric rule this restricted covering is a local isometry for the intrinsic metric of U.

For $0 < \ell < L$, every point z with $d_{\mathbb{D}}(0, z) < \ell$ is joined to zero by its radial hyperbolic geodesic of length less than ℓ. The projection of every initial segment has that same length. If a projected point lay in $V \setminus U$, it would be at V-distance less than L from x, contrary to the definition of L. Thus the whole lifted path remains over U, and the radius-ℓ hyperbolic disk is contained in W. By (2.2) it is the Euclidean disk of radius $r_\ell = \tanh(\ell/2)$.

The inclusions $B(0, r_\ell) \subset W \subset \mathbb{D}$ and metric monotonicity give

$$2 = \lambda_{\mathbb{D}}(0) \leq \lambda_W(0) \leq \lambda_{B(0,r_\ell)}(0) = 2/r_\ell.$$

In local coordinates at x, the two covering identities are $\lambda_W(0) = \lambda_U(x)|p'(0)|$ and $2 = \lambda_V(x)|p'(0)|$. Their quotient cancels the common, nonzero covering derivative. Letting $\ell \uparrow L$ proves the ratio bound, with no requirement that a point of the complement realize the infimum distance. Finally $\coth(L/2) - 1 = 2/(e^L - 1) \leq 4e^{-L}$ when $e^L \geq 2$. □

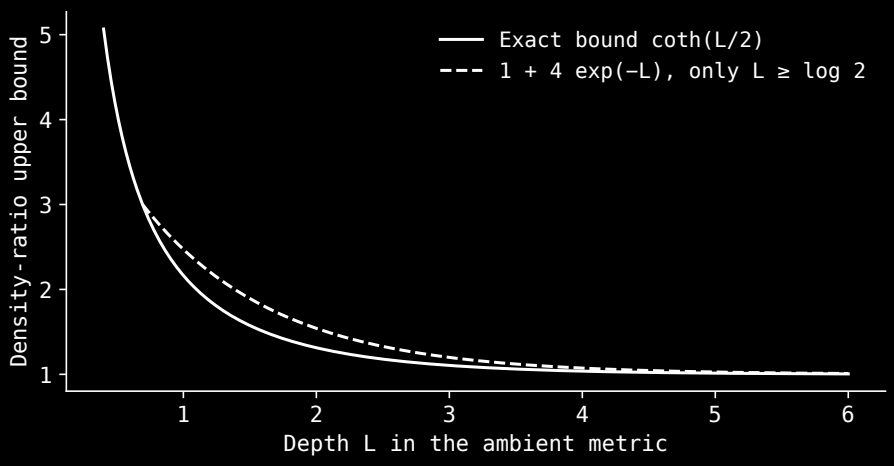


Figure 6.2: The exact comparison upper bound $\coth(L/2)$ and its large-distance majorant $1 + 4e^{-L}$. The latter is drawn only for $L \geq \log 2$, its stated validity range.

Inversion gives $\lambda_V/\lambda_U \geq \tanh(L/2)$, the quantitative form of source Exercise 7.13; the large-depth estimate makes Lemma 7.21's constant explicit. If $U = V$, the metrics coincide. The disks used lie in the universal cover, so no ambient injectivity-radius restriction is needed: projected path lengths exclude the complement. A coordinate change cancels in the density ratio.

Sharpness holds for $V = \mathbb{D}$, $U = B(0, r)$, $x = 0$: here $L = 2 \operatorname{artanh} r$ and $\lambda_U(0)/\lambda_V(0) = 1/r = \coth(L/2)$. The exact excess $2/(e^L - 1)$ is at most $\varepsilon > 0$ whenever $L \geq \log(1 + 2/\varepsilon)$; thus $L \geq \log 201$ guarantees a ratio at most $101/100$. Depth is measured in V, not in the generally larger intrinsic metric of U. Proposition 7.15 develops the complementary sharp bound from this same geometric datum.

Chapter 7

Symmetric neighborhoods and boundary comparison

This chapter follows the symmetric and boundary principles in the source, stating the precise ranges of the estimates. Disk uniformization and covering lifts are used for surface statements; the univalence estimates were proved in Chapter 3. The isolated-cusp coordinate input is stated where needed and developed geometrically in the quotient-surface chapters.

7.1 Symmetric Schwarz as an exact neighborhood statement

Source correspondence. Lyubich, the named Symmetric Schwarz principle, p. 134; Corollary 7.2..

For a finite real interval $I = (a, b)$, write $\ell = b - a$ and

$$\Omega_I = \mathbb{C} \setminus ((-\infty, a] \cup [b, \infty)), \quad Q_I(z) = \frac{z-a}{b-z}, \quad T_I(z) = \sqrt{Q_I(z)}.$$

The square root is the principal branch. The Möbius map Q_I sends Ω_I onto the plane slit along the nonpositive real axis; its square root sends that slit plane onto the right half-plane. Thus T_I is a conformal bijection and sends I onto the positive real axis. For $0 < \theta < \pi$ define the circular neighborhood by

$$D_\theta(I) = \{z \in \Omega_I : |\arg Q_I(z)| < \pi - \theta\}.$$

Its two boundary arcs are circles through a, b, one in each half-plane. At $\theta = \pi/2$ it is the round disk with diameter I.

7.1.1 Coordinates, branches and the actual density

The branch choice can be checked without an appeal to a picture. Solving $q = Q_I(z)$ for z gives

$$z = \frac{a + bq}{1 + q}.$$

The value $q = -1$ is not attained at a finite point. The two removed real rays correspond exactly to the nonpositive real axis, including its endpoints on the sphere. On its complement there is a single logarithm with imaginary part in $(-\pi, \pi)$, so $T_I = \exp(\frac{1}{2} \log Q_I)$ has positive real part. Conversely, if $\operatorname{Re} w > 0$, then w^2 does not belong to the nonpositive axis, and

$$T_I^{-1}(w) = \frac{a + bw^2}{1 + w^2}.$$

The denominator cannot vanish in the open right half-plane. These two explicit inverses establish both bijectivity and holomorphicity. They also specify what happens near the endpoints: a

corresponds to zero, b to infinity, and neither endpoint is an interior point of either coordinate domain.

Put $\phi = \arg Q_I(z) \in (-\pi, \pi)$. Since $Q_I'(z) = \ell/(b-z)^2$, the covering-normalized density is

$$\lambda_{\Omega_I}(z) = \frac{|T_I'(z)|}{\operatorname{Re} T_I(z)} = \frac{\ell}{2|z-a||b-z|\cos(\phi/2)}. \tag{7.1}$$

To obtain the middle expression, use the right-half-plane metric $|dw|/\operatorname{Re} w$. To obtain the last, write $|T_I'| = |Q_I'|/(2|Q_I|^{1/2})$ and $\operatorname{Re} T_I = |Q_I|^{1/2}\cos(\phi/2)$. Every factor in the denominator is positive on Ω_I; in particular this formula does not require a convention for an angle on a removed ray. On I, the cosine equals one, and the density reduces to $\ell/(2(x-a)(b-x))$. Thus the metric calculation itself explains the factor of two relative to the diameter disk.

For an explicit description of the circular arcs, write $m = (a+b)/2$ and $h = \ell/2$. In the upper half-plane their equation is

$$(x-m)^2 + (y - h\cot\theta)^2 = h^2\csc^2\theta, \qquad y > 0.$$

Indeed, rescaling z to $(z-m)/h$ changes the interval to $(-1,1)$. There Q_I has argument equal to the argument of $1-x^2-y^2+2iy$. Requiring this argument to be $\pi-\theta$ gives $x^2+y^2-2y\cot\theta = 1$; completing the square gives the displayed equation. The neighborhood is the upper circular cap and its reflection, together with the interval. When $\theta < \pi/2$ the upper center is above the real axis, so the cap exceeds a semicircle. This is why an argument that assumes every cap has height at most its radius would lose the small-angle cases.

Lemma 7.1 (The circular neighborhood is a hyperbolic tube)**.** *In the intrinsic metric of Ω_I,*

$$D_\theta(I) = \{z : d_{\Omega_I}(z, I) < R_\theta\}, \qquad R_\theta = \operatorname{arsinh}\big(\cot(\theta/2)\big).$$

Proof. The right half-plane has density $1/\operatorname{Re} w$. The disk distance formula, transported by a Cayley map, gives, for $w = u+iv$, $u>0$, and $s>0$,

$$\cosh d(w,s) = 1 + \frac{|w-s|^2}{2us} = \frac{u^2+v^2+s^2}{2us}.$$

As a function of s, the right side is minimized at $s = |w|$, by $s + |w|^2/s \geq 2|w|$. Therefore

$$d(w, \mathbb{R}_{>0}) = \operatorname{arsinh}(|v|/u) = \operatorname{arsinh}(\tan|\arg w|).$$

Since $\arg T_I = \frac{1}{2}\arg Q_I$, the claimed strict sublevel set and its radius follow. This also proves the asserted geometric meaning without relying on a diagram. □

The minimization also identifies the closest point on the interval, not just its distance. The strictly convex expression in s has the unique minimum $s = |T_I(z)|$. Pulling this point back gives

$$\operatorname{proj}_I(z) = \frac{a + b|Q_I(z)|}{1 + |Q_I(z)|}. \tag{7.2}$$

This is an interior point of I for every $z \in \Omega_I$. The right-half-plane geodesic from w to $|w|$ is the arc of the circle centered at zero with radius $|w|$. It meets the positive real axis orthogonally; conformal transfer gives the corresponding perpendicular in Ω_I. The Euclidean closest point and this intrinsic closest point need not agree. For example, above a point beyond the right endpoint, the Euclidean infimum is approached at that endpoint, whereas (7.2) always lies strictly between a and b.

There are useful endpoint checks on the tube radius. As $\theta \downarrow 0$, $\cot(\theta/2)$ tends to infinity, so the tubes exhaust Ω_I. As $\theta \uparrow \pi$, the radius decreases to zero and the neighborhoods shrink to the interval. The monotonic direction agrees with the defining angular inequality: a smaller angle parameter permits a larger set. Neither endpoint is included in the definition of a nondegenerate circular neighborhood.

Investigation 7.2 (An off-axis point in slit coordinates)**.** For $I = (-1, 1)$ and $z = iy$ with $y > 0$, find the nearest point of I in the intrinsic metric and determine exactly when $iy \in D_\theta(I)$.

Solution. The endpoint distances $|iy + 1|$ and $|1 - iy|$ coincide, hence $|Q_I(iy)| = 1$ and (7.2) gives the nearest point 0. Since $Q_I(iy) = (1 + iy)/(1 - iy) = \exp(2i \arctan y)$, its square root has argument $\arctan y$. The tube distance is consequently $\operatorname{arsinh} y$. Membership is equivalent to $y < \cot(\theta/2)$. The same upper height follows from the circle equation: its top point has height $\cot\theta + \csc\theta = \cot(\theta/2)$. The agreement checks both the branch and the choice of cap.

Proposition 7.3 (Symmetric inclusion with explicit sufficient hypotheses)**.** *If $f : \Omega_I \to \Omega_{I'}$ is holomorphic and $f(I) \subset I'$, then $f(D_\theta(I)) \subset D_\theta(I')$ for every $0 < \theta < \pi$.*

Proof. Distance contraction gives

$$d_{\Omega_{I'}}(f(z), I') \le \inf_{x \in I} d_{\Omega_{I'}}(f(z), f(x)) \le \inf_{x \in I} d_{\Omega_I}(z, x).$$

Apply Lemma 7.1. The source's real-symmetric univalent maps carrying I diffeomorphically onto I' satisfy these hypotheses. The weaker sufficient hypotheses in this proposition are an explicit extension, not a silent change to the definition of the source's map class. □

On the real interval, the diameter-disk density is $\ell/((x - a)(b - x))$. The intrinsic slit-plane density is half this: indeed $T_I'/T_I = \ell/(2(x - a)(b - x))$. The interval density printed on source p. 134 is twice the diameter-disk density. Multiplying both source and target interval metrics by the same constant cancels in the derivative ratio; it does not repair a false uniformity assertion.

Proposition 7.4 (Strictness and a valid distortion constant)**.** *Let $f : \Omega_I \to \Omega_{I'}$ be univalent, real-symmetric, and map I diffeomorphically onto I'. If f is not Möbius, then its differential norm between the diameter disks is strictly less than one at every point of I. On every compact subinterval it has a maximum strictly below one, depending on f and that subinterval.*

Separately, let $L = (c, d) \subset I$ have positive length $s = d - c$ and suppose both endpoint gaps are at least δs, where $\delta > 0$. For every $x, y \in L$,

$$\frac{|f'(x)|}{|f'(y)|} \le C(\delta), \qquad C(\delta) = \left(\frac{1 + \delta}{\delta}\right)^4.$$

Proof. Proposition 7.3 for $\theta = \pi/2$ permits Schwarz–Pick on the diameter disks. Equality at one point would make that restriction a Möbius map. The identity theorem would then make f Möbius on Ω_I. Strictness and continuity give the fixed-map compact-set conclusion.

For distortion, let $m = (c + d)/2$ and $R = (\delta + 1/2)s$. The disk $B(m, R)$ lies in Ω_I, because both removed real rays are at least distance R from m. Normalize the restriction of f to a univalent function g on the unit disk with $g(0) = 0$, $g'(0) = 1$. For points corresponding to L, $|\zeta| \le q = (1 + 2\delta)^{-1}$. Theorem 3.6 gives

$$\frac{1 - q}{(1 + q)^3} \le |g'(\zeta)| \le \frac{1 + q}{(1 - q)^3}.$$

Dividing the upper by the lower bound yields $((1 + q)/(1 - q))^4 = ((1 + \delta)/\delta)^4$. Univalence ensures the derivatives being divided by are nonzero. □

The family F_a already constructed in this manuscript has norm $(1 + a^2)/2 \to 1$ at the fixed center while remaining non-Möbius. It refutes the map-independent contraction deficit in Corollary 7.2. The distortion estimate above is a different, valid clause and does not use that deficit.

7.2 An explicit angular estimate

Source correspondence. Lyubich, Proposition 7.3, p. 135..

For nonreal z put

$$\alpha(z, I) = \min\{|\operatorname{Arg}(z-b)|, |\operatorname{Arg}(a-z)|\},$$

where Arg is the principal argument in $(-\pi, \pi]$. This gives a reflection-invariant angle and avoids the source's impossible global argument convention in $[0, \pi]$.

Proposition 7.5 (A quantified angular bound)**.** *Let f be as in Proposition 7.4, and let $0 < \theta < \pi$. If $z \in \Omega_I$ satisfies $D = \operatorname{dist}(z, I) \geq \ell$ and $\alpha(z, I) > \theta$, then*

$$\frac{\operatorname{dist}(f(z), I')}{|I'|} \leq \frac{4}{\sin\theta}\frac{D}{\ell}.$$

Proof. The hypotheses imply that z is nonreal. Reflect if necessary and write $\operatorname{Im} z = y > 0$. Let R be the circumradius of the triangle with vertices a, b, z. Its area is $\ell y/2$, so

$$R = \frac{|z-a||z-b|}{2y}.$$

If $a \leq \operatorname{Re} z \leq b$, then $D = y \geq \ell$ and both endpoint distances are at most $\sqrt{D^2+\ell^2} \leq \sqrt{2}D$. Consequently $R \leq D \leq D/\sin\theta$. If $\operatorname{Re} z \geq b$, then $D = |z-b|$ and $|z-a| \leq D + \ell \leq 2D$. The angle $\operatorname{Arg}(z-b)$ lies in $(0, \pi/2]$ and exceeds θ, so $y \geq D\sin\theta$. Again $R \leq D/\sin\theta$. The left-hand case is identical after reflection about the midpoint of I.

There is a unique $\eta \in (0, \pi)$ for which z belongs to the upper boundary arc of $D_\eta(I)$, namely $\eta = \pi - \arg Q_I(z)$. That arc lies on the circle just used. Its center and the center of its reflected circle have distance $\sqrt{R^2 - (\ell/2)^2} \leq R$ from the midpoint of I. Each half of the circular neighborhood is the cap inside its corresponding circle. Thus the whole neighborhood lies in the disk of radius $2R$ about that midpoint, and

$$\operatorname{diam} D_\eta(I) \leq 4R \leq 4D/\sin\theta.$$

This remains true when the cap is larger than a semicircle; no assertion that the nearest point of I is an endpoint was used.

For $\eta' < \eta$, z lies in the open neighborhood $D_{\eta'}(I)$. Proposition 7.3, followed by $\eta' \uparrow \eta$, places $f(z)$ in $\overline{D_\eta(I')}$. Circular neighborhoods for the same angle are Euclidean similarities, so their diameters scale by $|I'|/\ell$. Their closures contain I'. Hence $\operatorname{dist}(f(z), I') \leq (|I'|/\ell)\operatorname{diam} D_\eta(I)$, as required. □

The constant here is proved as a sufficient bound, not asserted optimal. In particular this argument replaces the source proof's incorrect equality between distance to the interval and distance to a selected endpoint.

7.3 Punctured disks and isolated punctures

Source correspondence. Lyubich, Lemma 7.4 and its punctured-disk dependency, p. 136 and p. 73..

Proposition 7.6 (Exact punctured-disk density)**.** *For $R > 0$ and $0 < |z| < R$,*

$$\lambda_{\{0<|z|<R\}}(z) = \frac{1}{|z|\log(R/|z|)}.$$

Proof. The map $q : \mathbb{H} \to \{0 < |z| < R\}$, $q(\zeta) = Re^{i\zeta}$, is a holomorphic universal covering. Its deck translations are $2\pi\mathbb{Z}$. The pullback identity for a hyperbolic covering and $\lambda_{\mathbb{H}}(\zeta) = 1/\operatorname{Im}\zeta$ give

$$\lambda(q(\zeta))|q'(\zeta)| = 1/\operatorname{Im}\zeta.$$

Now $|q'| = |q|$ and $\operatorname{Im}\zeta = \log(R/|q|)$. Substitution proves the formula at every point, independent of the lift. □

Proposition 7.7 (Cusp asymptotic with its geometric dependency)**.** *Let p be a point of a Riemann surface S such that $U = S \setminus \{p\}$ is hyperbolic. In a local holomorphic coordinate z centered at p,*

$$\lim_{z\to 0} \lambda_U(z)|z|\log(1/|z|) = 1.$$

Here the cusp-coordinate theorem is an explicit prerequisite: an isolated puncture has a neighborhood with a coordinate w in which the restriction of the ambient metric is $|dw|/(|w|\log(1/|w|))$ on $0 < |w| < r < 1$.

Proof. The coordinate w tends to zero at the puncture and extends across it by the removable singularity theorem. The inverse coordinate extends as well; their compositions remain the identity. Consequently $w(z) = cz(1 + O(z))$ with $c \neq 0$, and $w'(z) = c + O(z)$. Pulling back the stated cusp density gives

$$\lambda_U(z)|z|\log(1/|z|) = \frac{|w'(z)||z|}{|w(z)|}\frac{\log(1/|z|)}{\log(1/|w(z)|)} \longrightarrow 1.$$

The first factor tends to one and the two logarithms differ by the bounded quantity $-\log|c| + o(1)$. □

For every $0 < \varepsilon < 1$ this limit gives a radius $r_\varepsilon > 0$ on which the density is between $(1-\varepsilon)$ and $(1+\varepsilon)$ times $1/(|z|\log(1/|z|))$. This proves the source's comparison once the cusp-coordinate theorem is available; it does not prove that theorem. In particular the metric of the ambient punctured unit disk restricted to $|w| < r$ is not the intrinsic metric of the smaller punctured disk: the latter contains $\log(r/|w|)$.

7.3.1 A quantitative form of the coordinate change

The limiting assertion conceals two separate errors: the derivative of the coordinate change and the additive constant inside its logarithm. They can be separated explicitly. This is useful whenever a specified relative error is required, since the logarithmic term often decays much more slowly than the Taylor remainder.

Proposition 7.8 (Coordinate error at a cusp)**.** *Under the cusp-coordinate hypothesis of Proposition 7.7, write $w(z) = cz(1 + u(z))$, where $c \neq 0$, $u(0) = 0$, and u is holomorphic on a neighborhood of $\{|z| \le r_0\}$. Suppose $|u'(z)| \le A$ there, with $A \ge 0$. At a point with $r = |z| > 0$, put*

$$L = \log(1/r), \qquad E = \frac{Ar}{1-Ar}, \qquad D = |\log|c|| - \log(1-Ar).$$

If $r \le r_0$, $Ar \le 1/3$ and $L > D$, then

$$\frac{1-E}{1+D/L} \le \lambda_U(z)rL \le \frac{1+E}{1-D/L}, \qquad |\lambda_U(z)rL - 1| \le \frac{EL+D}{L-D}.$$

In particular, as $r \downarrow 0$,

$$\lambda_U(z)r\log(1/r) = 1 + \frac{\log|c|}{\log(1/r)} + O\left(\frac{1}{\log^2(1/r)} + r\right), \tag{7.3}$$

uniformly in the argument of z. Constants depend on the chosen cusp coordinate and its stated neighborhood.

Proof. Integrate u' along the radial segment from 0 to z to get $|u(z)| \le Ar$. Thus $1+u(z)$ is nonzero and

$$\frac{|w'(z)|r}{|w(z)|} = \left|1 + \frac{zu'(z)}{1+u(z)}\right| \in [1-E, 1+E].$$

The interval is positive under $Ar \le 1/3$. Also $1 - Ar \le |1+u| \le 1 + Ar$, which implies $|\log|1+u|| \le -\log(1-Ar)$. Consequently

$$\log(1/|w(z)|) = L - \log|c| - \log|1+u(z)| \in [L-D, L+D].$$

The lower endpoint is positive by hypothesis. Substituting both intervals in the exact pullback formula proves the two-sided estimate. The upper relative deviation is at most $(EL+D)/(L-D)$; the lower deviation is at most $(EL+D)/(L+D)$ and is therefore bounded by the same number.

For the expansion, the first factor above is $1+O(r)$, uniformly in angle, and the logarithmic denominator is $L - \log|c| + O(r)$. Divide by L and use $(1-v)^{-1} = 1 + v + v^2/(1-v)$ for $|v| \le 1/2$. This gives $L/(L - \log|c| + O(r)) = 1 + \log|c|/L + O(L^{-2} + r/L)$. Multiplication by $1+O(r)$ proves (7.3). The restrictions used in this last step hold for all sufficiently small r and are uniform in angle. □

For example, if one requests relative error at most ε, it is sufficient to choose a radius satisfying $E \le \varepsilon/4$ and $D/L \le \varepsilon/(4+2\varepsilon)$, with $0 < \varepsilon \le 1$. Indeed the error bound is $(E + D/L)/(1 - D/L)$, which under these conditions is less than ε. This is a sufficient radius test once A, c, r_0 are known. It is not an effective construction of the cusp coordinate from an arbitrary surface description. That global existence input remains the one identified in Proposition 7.7.

Investigation 7.9 (Why the logarithmic correction cannot be dropped)**.** In an exact cusp coordinate, replace w by $z = w/c$ with $c > 0$. Compute the relative metric factor and its first correction. Explain why an $O(|z|)$ relative-error claim generally fails.

Solution. Here $w = cz$ and $u = 0$, so the exact factor is $\lambda_U(z)|z|L = L/(L - \log c)$. For $c \ne 1$ its difference from one is $\log c/(L - \log c)$, of order $1/L$. This cannot be bounded by a constant times $r = e^{-L}$ for all large L, since e^L/L is unbounded. The leading equivalence is invariant under a coordinate rescaling; the rate of approach and first correction are not.

7.4 A positive global comparison on the sphere

Source correspondence. Lyubich, Proposition 7.5, p. 136..

Use the round sphere of radius one, so its geodesic diameter is π, and write σ_{sph} for its density. For $0 < d \le \pi$, put $F(d) = d\log(e\pi/d) > 0$.

Proposition 7.10 (A corrected global spherical bound)**.** *For every connected hyperbolic domain $U \subset \widehat{\mathbb{C}}$ there is $c_U > 0$ such that, for every $z \in U$ and $d = d_{\mathrm{sph}}(z, \partial U)$,*

$$\frac{\lambda_U(z)}{\sigma_{\mathrm{sph}}(z)} \ge \frac{c_U}{d\log(e\pi/d)}.$$

The proof uses the cusp asymptotic for the three-punctured sphere.

Proof. Choose three distinct boundary points p_1, p_2, p_3, with minimum pairwise spherical distance $\delta > 0$. A nearest boundary point a to z exists by compactness. At most one p_i is within $\delta/3$ of a. Choose two of the remaining points, b, c. The ordered triple (a, b, c) is separated by $\delta/3$.

The unique Möbius map M sending this triple to $(0, 1, \infty)$ and its inverse are uniformly K-Lipschitz for spherical distance, for a constant $K \geq 1$ depending only on δ. To justify uniformity, the space of such separated ordered triples is compact; the normalizing Möbius map depends continuously on the triple, and its positive spherical differential norm and that of its inverse are continuous also in the point of the compact sphere. Their maxima are therefore finite.

On $V = \widehat{\mathbb{C}} \setminus \{0, 1, \infty\}$ the cusp asymptotic at each of its three ends, and a positive minimum on the complementary compact set, give $c_0 > 0$ with $\lambda_V(w)/\sigma_{\mathrm{sph}}(w) \geq c_0/F(s)$, where $s = d_{\mathrm{sph}}(w, \{0, 1, \infty\})$. At $w = M(z)$ the Lipschitz bounds give $d/K \leq s \leq Kd$. Since $\log(e\pi/d) \geq 1$,

$$F(s) \leq Kd\big(\log(e\pi/d) + \log K\big) \leq K(1 + \log K)F(d).$$

The inclusion $U \subset \widehat{\mathbb{C}} \setminus \{a, b, c\}$ and Schwarz monotonicity now yield the assertion with $c_U = c_0/(K^2(1 + \log K))$. □

This formulation is defined and positive at $d = 1$ and for $d > 1$. The source's literal expression $-1/(d \log d)$ is undefined at one and cannot serve as this global positive comparison function.

7.4.1 Where the global comparison constant comes from

There are three compactness steps in Proposition 7.10, and keeping them separate prevents a hidden uniformity assumption. First, the three initially selected boundary points have positive separation; this fixes a scale δ. Second, the possible normalized triples (a, b, c) lie in the compact set of triples with pairwise separation at least $\delta/3$. This controls the spherical derivatives of their normalizing Möbius maps. Third, the fixed three-punctured sphere is split into its three cusp neighborhoods and a compact remainder. The positive continuous density ratio has a positive minimum only on that remainder; near a puncture its lower estimate comes from the cusp asymptotic, not from compactness of the punctured surface.

For completeness, if s is spherical distance to a puncture and w is a regular local coordinate at it, then $s = c|w|(1 + O(|w|))$ with $c > 0$, and the spherical density is $c(1 + O(|w|))$. The coordinate cusp expression therefore gives

$$\frac{\lambda_V(w)}{\sigma_{\mathrm{sph}}(w)}\, s \log(e\pi/s) \longrightarrow 1.$$

Choose disjoint cusp neighborhoods on which this product is at least $1/2$. On the remaining compact set the product is continuous and strictly positive. Taking the smaller of its minimum and $1/2$ produces the constant c_0 used in the proof. This also shows why the harmless positive normalization $e\pi$ inside the logarithm does not alter the local asymptotic.

The distance comparison used there deserves an explicit check. Each of a, b, c lies in ∂U, so its distance from z is at least d. The inverse Lipschitz estimate yields $s \geq d/K$. Since a was chosen nearest to z, the forward estimate gives $s \leq Kd$. The pullback density is $\lambda_{\widehat{\mathbb{C}}\setminus\{a,b,c\}}(z) = \lambda_V(Mz)|M'(z)|$. Dividing by the spherical density multiplies the target ratio by the spherical differential norm of M, at least $1/K$. The other factor $K(1 + \log K)$ comes from comparing $F(s)$ with $F(d)$. Thus the two factors of K in c_U have distinct origins.

Corollary 7.11 (Uniformity for separated boundary triples)**.** *Fix $\delta > 0$. Among connected hyperbolic spherical domains whose boundaries each contain three points with pairwise spherical distances at least δ, the constant in Proposition 7.10 can be chosen to depend only on δ.*

Proof. Choose such a triple separately for each domain. Every step of the proof then uses the same compact space of $\delta/3$-separated triples and the same fixed three-punctured sphere. Hence the same K and c_0 work, and $c_\delta = c_0/(K^2(1 + \log K))$ is admissible. No bound on the number of the domain's other boundary components enters. □

This statement permits highly irregular boundaries. It does not assert a constant uniform under collisions of all possible boundary triples. Such collisions remove the compactness hypothesis responsible for the normalizing-map bound.

7.5 A local version for an arbitrary smooth background metric

Source correspondence. Lyubich, Exercise 7.6, p. 136..

The all-point claim in the source fails already for $S = \mathbb{C}$, $K = \{0, 1\}$, and the Euclidean background metric: at $z = it$ as $t \uparrow 1$, distance to K equals t, while $\lambda_{\mathbb{C}\setminus\{0,1\}}(it)$ has a finite positive limit and $1/(-t \log t)$ diverges. The next statement is a local replacement.

Proposition 7.12 (Boundary comparison near a compact removed set)**.** *Let S be a connected Riemann surface with a smooth positive conformal metric σ. Let $K \subset S$ be nonempty and compact, and suppose $U = S \setminus K$ is connected and hyperbolic. There are constants $\varepsilon \in (0, e^{-1})$ and $c > 0$, depending on S, K, σ, such that, whenever $0 < d = d_\sigma(z, K) < \varepsilon$,*

$$\frac{\lambda_U(z)}{\sigma(z)} \geq \frac{c}{d \log(1/d)}.$$

No completeness of σ and no regularity of ∂K is assumed. Dependencies are uniformization of S, covering lifts, cusp asymptotics for the three-punctured sphere, and local comparison of smooth metrics.

Proof. Uniformize S by $\pi : X \to S$, where X is the disk, the plane or the sphere. Every component W of $\pi^{-1}(U)$ covers U, so $\lambda_W = \pi^* \lambda_U$. Put $\widetilde{K} = \pi^{-1}(K)$.

We first make the local uniformity precise. Cover K by finitely many inner coordinate neighborhoods whose closures lie in larger relatively compact, evenly covered coordinate neighborhoods. Fix a lift of each larger neighborhood. The selected lifted closures are compact. On them the pulled-back smooth metric has positive lower and finite upper bounds relative to a Euclidean coordinate (or a sphere chart). The σ-distance from each inner closure to the outside of its larger neighborhood has a positive lower bound: any exit path must cross a fixed coordinate collar on which the metric has a positive lower bound. Take $\eta > 0$ smaller than all these finitely many exit distances, and require $d < \eta/3$. Choose an inner neighborhood containing a nearest $k \in K$. A path of length less than $2d$ from k to z cannot leave its larger neighborhood, since its length is less than η. Such a nearest k exists because K is compact. Lifting the path gives points a, w in a fixed compact chart with $a \in \widetilde{K}$ and

$$e = |w - a| \leq Cd, \qquad (\pi^*\sigma)(w) \leq M,$$

with constants uniform over the finite chart family. Shrinking the allowed d also keeps all these points in the stated compact charts.

If $X = \mathbb{D}$, then $W \subset \mathbb{D} \setminus \{a\}$. With $m_a(w) = (w - a)/(1 - \bar{a}w)$, Proposition 7.6 gives

$$\lambda_{\mathbb{D}\setminus\{a\}}(w) = \frac{|m_a'(w)|}{|m_a(w)| \log(1/|m_a(w)|)}.$$

For a, w in the fixed compact subsets of $\mathbb{D}$ and e sufficiently small, the numerator is bounded below, $|m_a(w)|$ is bounded above and below by positive multiples of e, and the logarithm differs from $\log(1/e)$ by a bounded quantity. Monotonicity therefore gives $\lambda_W(w) \geq c_1/(e \log(1/e))$ uniformly.

If $X = \mathbb{C}$, then $\widetilde{K}$ contains two distinct points b_0, c_0. Otherwise its complement is $\mathbb{C}$ or $\mathbb{C}^*$, which is parabolic and cannot cover the hyperbolic surface U. For each a choose $b \in \{b_0, c_0\}$ with

$|a-b| \geq |b_0-c_0|/2$. Since the possible a lie in a fixed compact set, $|a-b|$ also has a uniform upper bound. Now $W \subset \mathbb{C} \setminus \{a,b\}$. The affine map $(w-a)/(b-a)$ identifies this latter domain with $\widehat{\mathbb{C}} \setminus \{0,1,\infty\}$. The cusp asymptotic at zero, these two bounds on $|a-b|$, and monotonicity give the same lower estimate $c_1/(e\log(1/e))$.

If $X = \widehat{\mathbb{C}}$, the covering is trivial: every nonidentity Möbius transformation has a fixed point and cannot be a deck transformation. The compact sphere's smooth metric σ is uniformly comparable to its round metric. Proposition 7.10 and this metric comparison give the desired estimate for small d directly.

In the first two cases divide by the bound M for the background density. For small positive arguments the function $t\log(1/t)$ is increasing. Since $e \leq Cd$, after enlarging C to be at least one and making d smaller, $e\log(1/e) \leq Cd\log(1/(Cd)) \leq Cd\log(1/d)$. This proves the claimed inequality uniformly. The finite-chart argument is local near K; no properness or completeness assertion for the background metric was used. □

The constant includes the background metric and the estimate is restricted to a sufficiently small boundary neighborhood. These are substantive parts of the replacement contract, not suppressed qualifications of the false literal all-point expression.

7.6 Explicit definite estimates and their scope

Source correspondence. Lyubich, Lemma 7.7 and Lemma 7.8 through Exercise 7.13, pp. 136–138..

For a proper simply connected plane domain, the corrected conformal-disk bounds in this normalization are $1/(2d) \leq \lambda_U \leq 2/d$, where d is Euclidean distance to the boundary. The lower estimate uses Koebe's quarter theorem in the direction $d \geq |\psi'(0)|/4$; the upper estimate uses the included radius-d disk. This is the normalization-corrected contract for 7.7, not its literal printed constants.

For $0 < r < 1$ define

$$B(r) = \frac{2r\log(1/r)}{1-r^2}, \qquad Q(R) = B(\tanh(R/2)) \quad (R > 0).$$

Writing $r = e^{-t}$ gives $B(r) = t/\sinh t$. Thus $0 < B(r) < 1$, B is increasing, and $B(r) \to 1$ as $r \uparrow 1$. Indeed $\sinh t > t$ for $t > 0$, and $(t\cosh t - \sinh t)' = t\sinh t > 0$, which proves the required monotonicity.

The omitted-point argument already given in the manuscript shows that if $f : \mathbb{D} \to \mathbb{D}$, $f(0) = 0$, and f omits b with $0 < |b| \leq r < 1$, then $|f'(0)| \leq B(r)$. To recall its analytic content, the zero-free function $m_b \circ f$ has a disk logarithm. The positive-real-part function $-\log(m_b \circ f)$ obeys $|h'(0)| \leq 2\operatorname{Re} h(0)$. This gives $(1-|b|^2)|f'(0)|/|b| \leq 2\log(1/|b|)$ and the stated bound. The curvature route gives a separate derivation of the half-plane derivative inequality; neither route presupposes the desired contraction constant. The excluded case $b = 0$ is incompatible with $f(0) = 0$.

Proposition 7.13 (Definite contraction on surfaces)**.** *Let $f : S \to V$ be holomorphic between connected hyperbolic surfaces, $x \in S$, and $R > 0$. Give the empty set distance $+\infty$. If $d_V(f(x), V \setminus f(S)) \leq R$, then*

$$\|Df(x)\|_{S,V} \leq Q(R) < 1.$$

For a nonconstant map its image is open, and the complement distance is the distance to its relative boundary in V.

Proof. A constant map has zero derivative. Otherwise choose normalized universal disk coverings at $x, f(x)$ and the lift $\widetilde{f}$ fixing zero. For every $\eta > 0$, choose an omitted point b in V and a path

from $f(x)$ to b of length less than $R + 2\eta$. Lift that path to the target disk starting at zero. Its endpoint β is omitted by $\widetilde{f}$, since its projection b is omitted by f. Its disk distance from zero is at most the path length, whence $|\beta| \leq \tanh((R + 2\eta)/2)$. The disk omitted-point bound applies. Let η tend to zero and use continuity of B. This argument needs no assertion that the infimum distance is attained by a boundary point. □

Apply this to an inclusion $U \subsetneq V$ to obtain the corrected quantitative form of 7.10:

$$d_V(x, V \setminus U) \leq R \quad \Longrightarrow \quad \lambda_V(x)/\lambda_U(x) \leq Q(R).$$

If $f : U \to V$ is an unramified covering, local isometry gives $\lambda_V(f(x))|f'(x)| = \lambda_U(x)$. Its differential norm when both source and target are measured using V is therefore at least $1/Q(R) > 1$. This is 7.11 with its quantifiers and metric roles fixed.

Proposition 7.14 (An explicit interior-depth estimate)**.** *If $U \subset V$ are connected hyperbolic surfaces, $x \in U$, and $d_V(x, V \setminus U) \geq r > 0$, then*

$$\tanh(r/2) \leq \lambda_V(x)/\lambda_U(x) \leq 1.$$

The lower constant tends to one as r tends to infinity and is sharp.

Proof. Lift V to the disk with x under zero, and let W be the component of the preimage of U containing zero. Every radial segment of hyperbolic length less than r projects into U: otherwise its first exit would contradict the assumed depth. Thus $B(0, \tanh(r/2)) \subset W$. Monotonicity and covering isometry give $\lambda_U(x)/\lambda_V(x) = \lambda_W(0)/2 \leq 1/\tanh(r/2)$. The upper bound is ordinary monotonicity. For equality in the lower bound take $V = \mathbb{D}$, $U = B(0, \tanh(r/2))$, $x = 0$. If $U = V$, the depth is infinite and the ratio is one. □

7.6.1 Sharp two-sided information from one depth

The definite-contraction and interior-depth estimates use the same distance in opposite directions. Combining them records exactly what that single geometric datum tells us.

Proposition 7.15 (Two sharp depth bounds)**.** *Let $U \subsetneq V$ be connected hyperbolic surfaces and $x \in U$. If $L = d_V(x, V \setminus U)$ is finite and positive, then*

$$\tanh(L/2) \leq \frac{\lambda_V(x)}{\lambda_U(x)} \leq Q(L), \qquad Q(L) = B(\tanh(L/2)).$$

For every $L > 0$, each bound is attained by a suitable inclusion into the disk. Furthermore,

$$\frac{2}{3}e^{-2L} \leq 1 - Q(L) \leq e^{-2L}, \qquad 1 - Q(L) = \frac{2}{3}e^{-2L} + O(e^{-4L}) \quad (L \to \infty). \tag{7.4}$$

Proof. The lower bound is Proposition 7.14; the upper is Proposition 7.13 applied to the inclusion with $R = L$. To attain the lower bound, set $V = \mathbb{D}$, $x = 0$ and $U = \{|z| < r\}$, where $r = \tanh(L/2)$. The depth is L and the center densities are 2 and $2/r$.

For the upper bound, choose $a \in \mathbb{D}$ with $|a| = r$ and put $U = \mathbb{D} \setminus \{a\}$. Its complement in V is the singleton $\{a\}$, so the depth from 0 is again L. The automorphism m_a maps this punctured disk onto $\mathbb{D} \setminus \{0\}$. At zero, $|m_a(0)| = r$ and $|m_a'(0)| = 1 - r^2$. Pulling back the exact punctured-disk density gives

$$\lambda_U(0) = \frac{1 - r^2}{r \log(1/r)}, \qquad \frac{\lambda_{\mathbb{D}}(0)}{\lambda_U(0)} = \frac{2r \log(1/r)}{1 - r^2} = Q(L).$$

Thus sharpness is within the stated surface category; the upper example is multiply connected, whereas the lower one is simply connected.

For the quantitative gap set $s = e^{-L} \in (0,1)$. Then $r = (1-s)/(1+s)$ and $\log(1/r) = 2\,\operatorname{artanh} s$. Since $\sinh(2\,\operatorname{artanh} s) = 2s/(1-s^2)$,

$$Q(L) = \frac{1-s^2}{s}\operatorname{artanh} s = 1 - 2\sum_{n=1}^{\infty} \frac{s^{2n}}{(2n-1)(2n+1)}.$$

The power series follows by integrating $1/(1-s^2) = \sum_{n\geq 0} s^{2n}$ on compact subintervals of $[0,1)$. Subtraction of the two absolutely convergent series gives the displayed coefficients. Their sum is one, because $2/((2n-1)(2n+1)) = 1/(2n-1) - 1/(2n+1)$ telescopes. Keeping the first term gives the lower gap bound; using $s^{2n} \leq s^2$ in every term gives the upper. Finally the remainder after the first term is at most $s^4 \sum_{n\geq 2} 2/((2n-1)(2n+1)) = s^4/3$. This proves the asymptotic with an explicit remainder bound. □

The two sharp examples explain a feature that would be obscured by a single unspecified constant. A large included disk guarantees a metric ratio whose deficit is at most of order e^{-L}, since $1 - \tanh(L/2) = 2/(e^L+1)$. An omitted point at distance L forces a deficit at least of order e^{-2L}. Those statements have different extremizers. They cannot be identified merely because both ratios approach one when the boundary is far away.

Investigation 7.16 (An exact depth comparison)**.** Compare the center metric ratios for $\{|z| < 1/2\} \subset \mathbb{D}$ and $\mathbb{D} \setminus \{1/2\} \subset \mathbb{D}$. Check that both inclusions have the same target-metric depth and identify which endpoint of Proposition 7.15 each realizes.

Solution. In both cases the nearest complement has Euclidean radius $1/2$, hence $L = 2\,\operatorname{artanh}(1/2) = \log 3$. The first density ratio is $1/2$ and the second is $4\log 2/3$. These are respectively $\tanh(L/2)$ and $Q(L)$. For the latter, $2/27 \leq 1 - 4\log 2/3 \leq 1/9$ follows from (7.4). No decimal approximation is required either to distinguish the two geometries or to verify their relative order.

Finally, the annular inclusion clause in 7.12 is refuted by the round annuli $A_t = \{e^{-t} < |z| < e^t\}$: their modulus is t/π, their density on the unit circle is $\pi/(2t)$, and $\lambda_{A_{t+\varepsilon}}/\lambda_{A_t} = t/(t+\varepsilon) \to 1$. One valid replacement imposes a target-metric depth bound and applies Proposition 7.13. For round annuli another imposes $\varepsilon/t \geq \eta > 0$, giving the bound $1/(1+\eta)$. The additional double-cover clause has a genuine extra constraint. For a symmetric essential inclusion and a two-sheeted covering onto the larger annulus, modulus is halved. Theorem 9.4 proves a uniform expansion factor from that modulus relation and an exact cylinder-band estimate. Thus the false inclusion-only clause and the valid covering clause have different conclusions and different proofs.

Chapter 8

Hyperbolic surfaces and the geometry of a cusp

8.1 From a local metric to a complete surface

The disk and half-plane models describe more than simply connected domains. A hyperbolic Riemann surface has a universal covering by the disk; after a Cayley transformation it can be written

$$S = \mathbb{H}/\Gamma, \qquad \mathbb{H} = \{x + iy : y > 0\}, \qquad \mathrm{d}s^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2},$$

where Γ is a discrete torsion-free subgroup of $\mathrm{PSL}_2(\mathbb{R})$. We use the uniformization and covering-space theorems developed in the preceding volume to obtain this representation. The arguments in this chapter concern the geometry of the resulting quotient. In particular, they do not require an explicit formula for its covering map.

Throughout these chapters curvature is -1. The corresponding disk density is $2/(1-|z|^2)$, the half-plane density is $1/y$, and the area form is $\mathrm{d}x\,\mathrm{d}y/y^2$. A factor of two in the density changes lengths, areas, and the numerical constants in collar formulas. Fixing the normalization at the outset makes the geometric and conformal calculations compatible.

A matrix $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of determinant one acts by $Az = (az+b)/(cz+d)$. The two identities

$$\operatorname{Im}(Az) = \frac{y}{|cz+d|^2}, \qquad A'(z) = \frac{1}{(cz+d)^2}$$

show directly that this action preserves the metric. Matrices A and $-A$ give the same transformation. Since the group acts freely and properly discontinuously, sufficiently small disks project isometrically onto neighborhoods in S. The metric therefore descends without a singularity at an ordinary point of the quotient.

The half-plane is complete. One can see why the two possible ways of leaving it both cost infinite length: approaching $y = 0$ or $y = \infty$ requires unbounded variation of $\log y$, while a path with y bounded above and below has hyperbolic length comparable to its Euclidean length. More formally, these observations make every closed metric ball stay in a Euclidean compact subset of $\mathbb{H}$. Geodesics on the quotient lift to geodesics in $\mathbb{H}$ and extend for all time. The usual completeness equivalence for connected Riemannian manifolds then gives completeness of S. A puncture in the conformal compactification can thus be infinitely far away in the intrinsic metric.

The geodesics in $\mathbb{H}$ are vertical lines and semicircles orthogonal to the real axis. The imaginary axis minimizes length between its points because $\mathrm{d}s \geq |\,\mathrm{d}y|/y$; applying real Möbius isometries gives the general case. The disk formula gives

$$\cosh d_{\mathbb{H}}(z,w) = 1 + \frac{|z-w|^2}{2\operatorname{Im} z \operatorname{Im} w}. \tag{8.1}$$

This single expression will determine both cusp and collar displacement functions.

8.1.1 Completeness and quotient distance, with the quantifiers exposed

The compact-ball argument for $\mathbb{H}$ can be made quantitative. If a piecewise smooth path starts at $z_0 = x_0 + iy_0$ and has hyperbolic length at most R, then every initial portion has length at most R. Since $|\,\mathrm{d}\log y| \le \mathrm{d}s$, its height satisfies

$$y_0 e^{-R} \le y \le y_0 e^{R}.$$

Also $|\,\mathrm{d}x| \le y\,\mathrm{d}s$, so $|x - x_0| \le y_0 e^R R$ along the path. If $d(z_0, z) \le R$, use paths of length at most $R + 1$ to obtain these bounds with $R + 1$ in place of R. The closed metric ball is therefore contained in a Euclidean rectangle lying strictly above the real axis. On that rectangle the Euclidean and hyperbolic metrics are uniformly comparable. The ball is Euclidean closed there, because distance is continuous, and hence compact. A hyperbolic Cauchy sequence eventually belongs to such a ball; a convergent subsequence and the Cauchy condition then give convergence of the whole sequence.

There is a similarly useful formula for distances after taking the quotient. Write $\pi : \mathbb{H} \to S$ for the projection, choose lifts z, w of p, q, and set

$$d_S(p, q) = \min_{\gamma \in \Gamma} d_{\mathbb{H}}(z, \gamma w). \tag{8.2}$$

First prove equality with an infimum. A path from p to q lifts from z to some γw, with the same length, so its length is at least the infimum on the right. Conversely, the geodesic from z to each γw projects to an admissible path of exactly that length. These two inequalities prove the formula with inf.

To see attainment, choose any finite upper bound for the infimum, for example $d(z, w) + 1$. Only finitely many orbit points γw lie in the compact ball of that radius about z. This finiteness is a consequence of proper discontinuity: the intersection of a discrete deck orbit with a compact set is finite. At least one orbit point lies below the chosen upper bound, and taking the minimum among these finitely many points gives the infimum over the entire orbit. Projection of its joining geodesic is a minimizing path in S. There may be several minimizing lifts, so uniqueness of shortest paths in the plane does not imply uniqueness on a quotient.

This formula also gives a direct metric proof that S is complete. If (p_n) is Cauchy, pass to a subsequence with $d_S(p_{n_j}, p_{n_{j+1}}) < 2^{-j}$. Choose a lift of the first point and then, successively, a lift of the next point realizing the quotient distance. The lifted sequence is Cauchy, since the tail distances are bounded by a convergent geometric series. It converges in $\mathbb{H}$, and its projection supplies a limit of the original Cauchy sequence. This argument proves completeness without assuming that a single global inverse branch of π exists.

Finally, the half-plane formula itself follows from the disk formula without solving a differential equation. For $w = u + iv$, the map

$$\Phi_w(z) = \frac{z - w}{z - \overline{w}}$$

maps $\mathbb{H}$ to $\mathbb{D}$ and sends w to zero. Its pseudohyperbolic radius is $t = |z - w|/|z - \overline{w}|$. Since $|z - \overline{w}|^2 - |z - w|^2 = 4yv$ and $\cosh(2\,\operatorname{artanh} t) = (1 + t^2)/(1 - t^2)$, substitution gives (8.1). All denominators are positive because $y, v > 0$. The computation covers $z = w$ as well, when $t = 0$ and the distance vanishes.

8.2 Deck transformations and closed geodesics

A nonidentity real Möbius transformation is elliptic, parabolic, or hyperbolic according as the absolute value of its determinant-one trace is less than, equal to, or greater than two. The

fixed-point equation gives the classification: negative discriminant gives one fixed point in $\mathbb{H}$; a repeated real root gives one ideal fixed point; two distinct real roots give two ideal fixed points. The first case is conjugate to a rotation, the second to a nonzero translation, and the third to a dilation.

No elliptic transformation belongs to our group. A rotation through a rational multiple of a full turn has finite order, contradicting torsion-freeness. An irrational rotation has distinct powers converging to the identity, contradicting discreteness. These are different exclusions: merely ruling out finite-order elements would not by itself exclude an irrational rotation in an arbitrary nondiscrete group.

A hyperbolic element with translation length $\ell > 0$ can be normalized to

$$A_\ell z = e^\ell z, \qquad A_\ell = \begin{pmatrix} e^{\ell/2} & 0 \\ 0 & e^{-\ell/2} \end{pmatrix}.$$

Its invariant geodesic is the imaginary axis, and its displacement there is exactly ℓ. The distance formula shows that its displacement off the axis is larger. Consequently the axis projects to the unique closed geodesic in the associated free homotopy class. To justify the length assertion, lift a representative loop to a path from z to $A_\ell z$; its length is at least their distance, and hence at least ℓ. The projected axis attains this bound. Uniqueness follows from the uniqueness of the invariant geodesic with the two fixed ideal endpoints.

By contrast, the translation $Tz = z + 1$ has displacement tending to zero as $\operatorname{Im} z \to \infty$, but no interior point has zero displacement. Its free homotopy class has no closed geodesic of minimum length. A cusp is therefore not a closed geodesic whose length has literally become zero. It is a different quotient model, although families of short closed geodesics can degenerate toward cusps.

We call a closed geodesic *primitive* when it traverses its image once rather than repeating a shorter closed geodesic. Primitivity identifies the correct generator of its cyclic stabilizer. Simplicity is a separate condition: it says that the image has no self-intersection. The collar theorem later uses both conditions, and neither should disappear from its statement.

8.3 Injectivity radius and local geometric scale

Definition 8.1. For $p \in S$, the injectivity radius $\operatorname{inj}_S(p)$ is the supremum of radii for which the covering projection is injective on a disk in $\mathbb{H}$ centered at a lift of p. Such a disk projects by a Riemannian isometry onto its image; distances between its points measured through the whole surface may still admit shortcuts outside that image.

If z is a lift of p, then

$$\operatorname{inj}_S(p) = \frac{1}{2} \inf_{\gamma \in \Gamma \setminus \{1\}} d_{\mathbb{H}}(z, \gamma z). \tag{8.3}$$

For the trivial group the infimum is $+\infty$. Otherwise the value is positive at every point. Indeed, proper discontinuity prevents nontrivial translates of a sufficiently small disk from meeting it. For a direct verification of the factor $1/2$, disks of radius r centered at z and γz are disjoint whenever $2r < d(z, \gamma z)$. If the reverse strict inequality holds, their joining geodesic has a midpoint in both disks. Thus a projection ceases to be injective precisely at half the first displacement scale.

Proposition 8.2. *On a non-simply-connected complete hyperbolic surface, injectivity radius is a positive* 1*-Lipschitz function:*

$$|\operatorname{inj}_S(p) - \operatorname{inj}_S(q)| \le d_S(p, q).$$

Proof. Choose a path from p to q with length within an arbitrary positive amount of $d_S(p,q)$ and lift it from z to w. The triangle inequality and the isometry property give

$$d(z,\gamma z) \le 2d(z,w) + d(w,\gamma w)$$

for each deck transformation. Take the infimum, divide by two, and interchange z, w. Let the excess path length tend to zero. □

This estimate turns a pointwise lower bound into a neighborhood bound. If $\operatorname{inj}_S(p) \ge \varepsilon$, then every point within distance $\varepsilon/2$ has injectivity radius at least $\varepsilon/2$. It also shows that a compact subset cannot contain points with injectivity radius tending to zero.

For later area estimates, write the disk radius as $s = \tanh(r/2)$. Substitution into the disk metric gives polar coordinates

$$\mathrm{d}s^2 = \mathrm{d}r^2 + \sinh^2 r\ \mathrm{d}\theta^2, \quad \operatorname{area}(B(r)) = 2\pi(\cosh r - 1), \quad \operatorname{length}(\partial B(r)) = 2\pi \sinh r. \tag{8.4}$$

Here the repeated notation $\mathrm{d}s^2$ denotes the metric, while s in the substitution was the Euclidean radius. The area formula follows by integrating $\sinh r\ \mathrm{d}r\ \mathrm{d}\theta$. For small r, the area is $\pi r^2 + O(r^4)$, as expected from a smooth two-dimensional metric. On a quotient these formulas apply to a disk whenever its radius is below the injectivity radius of its center.

8.4 A matrix estimate for an embedded cusp

A parabolic fixed point can be moved to infinity. Its stabilizer then contains a translation. After scaling, a primitive positive translation is $Tz = z + 1$. The following estimate is often called the Shimizu lemma; the proof is short enough to include in full.

Lemma 8.3. *Suppose a discrete subgroup of* $\mathrm{PSL}_2(\mathbb{R})$ *contains* $Tz = z+1$. *If* $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ *belongs to the group, then either* $c = 0$ *or* $|c| \ge 1$.

Proof. Suppose $0 < |c_0| = q < 1$ for a representative A_0. Recursively put $A_{n+1} = A_n T A_n^{-1}$. Matrix multiplication gives

$$A_{n+1} = \begin{pmatrix} 1 - a_n c_n & a_n^2 \\ -c_n^2 & 1 + a_n c_n \end{pmatrix}.$$

In particular $|c_n| = q^{2^n}$, so these entries are nonzero, strictly decrease in absolute value, and tend to zero. Also

$$|a_{n+1}| \le 1 + q|a_n|.$$

Taking $M = \max\{|a_0|, (1-q)^{-1}\}$ proves inductively that $|a_n| \le M$. Hence $a_n c_n \to 0$, so $a_n \to 1$. The displayed matrix recurrence now gives $b_n \to 1$ and $d_n \to 1$. Thus $A_n \to T$, while the nonzero distinct values of $|c_n|$ show that these are distinct group elements. This contradicts discreteness at T, or equivalently discreteness at the identity after multiplication by T^{-1}. □

For completeness, the full stabilizer of infinity is a group of translations in this setting. If it contained $z \mapsto az + b$ with $a > 0$, $a \ne 1$, conjugating a nonzero translation by its positive or negative powers would produce translations tending to the identity. This is impossible. A discrete additive subgroup of the translation line is cyclic, so choosing its least positive translation and normalizing its period to one is legitimate.

Theorem 8.4. *A primitive parabolic end of a complete torsion-free hyperbolic quotient contains an embedded open cusp*

$$\{y > 1\}/\langle z \mapsto z + 1\rangle.$$

Its area is one. Every smaller cusp $\{y > Y\}/\langle T\rangle$, $Y > 1$, *has embedded boundary of length* $1/Y$, *and its area is also* $1/Y$.

Proof. Let $H_Y = \{\operatorname{Im} z > Y\}$. If A does not fix infinity, Lemma 8.3 gives

$$\operatorname{Im}(Az) = \frac{y}{|cz+d|^2} \le \frac{1}{c^2 y} \le \frac{1}{y}.$$

Consequently $A(H_1)$ lies below the line $y = 1$ and is disjoint from H_1. Elements fixing infinity are precisely powers of the primitive translation. Thus H_1 is precisely invariant under this cyclic subgroup, and its quotient embeds in S. At $Y > 1$ the same strict separation applies to the closure of H_Y, giving an embedded boundary. Finally,

$$\operatorname{area}(H_Y/\langle T\rangle) = \int_0^1 \int_Y^\infty \frac{\mathrm{d}y\,\mathrm{d}x}{y^2} = \frac{1}{Y}, \qquad \operatorname{length}\{y = Y\}/\langle T\rangle = \int_0^1 \frac{\mathrm{d}x}{Y} = \frac{1}{Y}.$$

□

The boundary at the limiting level $Y = 1$ can touch another translate; the theorem deliberately asserts embedding of the open cusp there. This distinction is harmless for its area and essential for its boundary geometry.

Unit open cusps belonging to different primitive parabolic classes can also be chosen disjoint. Normalize the first to H_1 with generator T. Write a primitive generator of the other stabilizer as $B = MTM^{-1}$, choosing its inverse if needed, where $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Distinct fixed points mean $c \neq 0$. The lower-left entry of B is $-c^2$, so Lemma 8.3 applied to B gives $c^2 \ge 1$. The horoball $M(H_1)$ has top height $1/c^2 \le 1$ and hence does not meet the first. Applying this argument after normalization to every pair gives pairwise disjoint open cusps. This is a statement about parabolic ends; the identification of the ends of a finite-type surface as cusps or funnels is part of the classical end classification.

8.5 Coordinates, completeness, and decay in a cusp

Put $y = Ye^r$ and $x = t \pmod 1$ in the cusp above height Y. Its metric becomes

$$\mathrm{d}s^2 = \mathrm{d}r^2 + Y^{-2}e^{-2r}\,\mathrm{d}t^2, \qquad r > 0, \quad t \in \mathbb{R}/\mathbb{Z}. \tag{8.5}$$

Thus r is the distance from the initial horocycle along a vertical geodesic, and the circumference at depth r is $Y^{-1}e^{-r}$. The remaining area beyond that depth equals the same number. The cusp has infinite length in its longitudinal direction and finite area in its transverse direction. There is no contradiction: its cross-sections shrink exponentially.

The coordinate $q = e^{2\pi i z}$ identifies this cusp conformally with the punctured disk $0 < |q| < e^{-2\pi Y}$. Since $|\,\mathrm{d}q| = 2\pi|q|\,|\,\mathrm{d}z|$ and $\log(1/|q|) = 2\pi y$, its metric is

$$\frac{|\,\mathrm{d}q|}{|q|\log(1/|q|)}.$$

This recovers the punctured-disk density by a geometric quotient calculation. It also explains why Euclidean closeness to a puncture is a poor guide to hyperbolic distance: an additional fixed hyperbolic depth multiplies y by a fixed factor and raises $|q|$ to a corresponding power.

The exact displacement of T follows from (8.1):

$$d_{\mathbb{H}}(x+iy, x+1+iy) = 2\operatorname{arsinh}\frac{1}{2y}. \tag{8.6}$$

For the complete parabolic cylinder $\mathbb{H}/\langle T\rangle$, the powers $T^{\pm 1}$ minimize nontrivial displacement, and hence

$$\operatorname{inj}(x+iy) = \operatorname{arsinh}\frac{1}{2y}.$$

In a larger surface a cusp is only a subdomain, and other deck transformations must initially be considered too. The thin-part argument below identifies a regime where the cyclic expression is the actual injectivity radius of the ambient surface. We do not infer that fact merely from the embedding of a cusp neighborhood.

8.6 Solved investigations: quotient geometry

Investigation 8.5 (Circumference and a based loop)**.** In the parabolic cylinder, compare the horocycle through iy with the shortest based loop in the same homotopy class. Explain why their lengths differ.

Solution. The horocycle has length $1/y$. The shortest based loop is the projection of the hyperbolic geodesic joining iy to $1+iy$, so its length is $2\operatorname{arsinh}(1/(2y))$. Since $\operatorname{arsinh} u < u$ for $u > 0$, the latter is strictly shorter. In the cover this geodesic rises above the horizontal line, where horizontal motion costs less. Its projection has a corner at the specified base point, so it is not a smooth closed geodesic. There is no conflict with the nonexistence of a closed geodesic in a parabolic free homotopy class. As $y \to \infty$, the ratio of the two lengths tends to one, but equality never occurs at a finite height.

Investigation 8.6 (Half the area lies one logarithmic step farther in)**.** Start with a cusp bounded by a horocycle of length h. Find the depth at which precisely half of its area remains. Determine whether the constant function is square-integrable on the cusp.

Solution. The area beyond depth R is he^{-R}, so the required depth is $R = \log 2$, independently of h. The constant function 1 has squared integral h and is therefore in L^2 on this finite-area end. Infinite metric extent alone does not imply infinite measure. For a radial function $u(r)$, the exact formulas are

$$\|u\|_{L^2}^2 = h\int_0^\infty |u(r)|^2 e^{-r}\,\mathrm{d}r, \qquad \int |\nabla u|^2\,\mathrm{d}A = h\int_0^\infty |u'(r)|^2 e^{-r}\,\mathrm{d}r.$$

These follow directly from (8.5), and show how cusp geometry enters energy estimates without replacing the hyperbolic measure by Euclidean area.

Investigation 8.7 (A covering preserves local scale but can enlarge injectivity radius)**.** Compare the cyclic surfaces $S_1 = \mathbb{H}/\langle z \mapsto e^{\ell} z\rangle$ and $S_k = \mathbb{H}/\langle z \mapsto e^{k\ell} z\rangle$, where k is a positive integer.

Solution. The inclusion of the second deck group in the first gives a degree-k covering $S_k \to S_1$. Both metrics pull back to the same metric on $\mathbb{H}$, so the covering is a local isometry. On the core axis, however, the shortest nontrivial deck displacement is $k\ell$ in S_k and ℓ in S_1. Their injectivity radii there are $k\ell/2$ and $\ell/2$. A local isometry does not preserve all global distances or shortest loops. More generally, passing to a subgroup removes competitors from the infimum in (8.3); it can increase injectivity radius, never decrease it at corresponding lifted points.

Chapter 9

Closed geodesics, collars, and conformal modulus

9.1 Why a short closed geodesic needs room

An annulus can be long in its conformal coordinate while having a very short central geodesic. The relation between these quantities is one of the most useful bridges between extremal length and hyperbolic geometry. To make the bridge precise we first prove that a primitive simple closed geodesic of length ℓ has an embedded collar of a specified width. The argument is elementary once the surface has been represented as a quotient: disjoint lifts, a trace computation, and explicit normal coordinates suffice.

The conclusion is classical. Our proof develops the equal-axis matrix argument used in Volume VI, now with the surrounding geometric details needed for independent reading. The general simultaneous collar theorem is attributed separately below [4]. In particular, a proof about distinct lifts of one geodesic must not silently become a proof about arbitrary geodesics of unequal lengths.

9.2 Normal coordinates around an axis

Around the imaginary axis introduce coordinates $(r, \tau) \in \mathbb{R}^2$ by

$$x = e^{\tau} \tanh r, \qquad y = e^{\tau} \operatorname{sech} r. \tag{9.1}$$

They cover the whole half-plane: the inverse formulas are

$$r = \operatorname{arsinh}(x/y), \qquad \tau = \tfrac{1}{2} \log(x^2 + y^2).$$

Differentiating (9.1) gives

$$\mathrm{d}s^2 = \mathrm{d}r^2 + \cosh^2 r \, \mathrm{d}\tau^2. \tag{9.2}$$

The curves of constant τ are the orthogonal geodesics to the axis. A path joining (r, τ) to the axis has length at least $|r|$, because the metric dominates $|\,\mathrm{d}r|$; the constant-τ segment attains it. Thus r is signed distance to the axis, and τ is arclength along it.

The dilation $z \mapsto e^{\ell} z$ adds ℓ to τ. Quotienting gives the complete hyperbolic cylinder

$$\mathrm{d}s^2 = \mathrm{d}r^2 + \ell^2 \cosh^2 r \, \mathrm{d}t^2, \qquad r \in \mathbb{R}, \quad t \in \mathbb{R}/\mathbb{Z}. \tag{9.3}$$

At distance r its circumference is $\ell \cosh r$. This increases away from the central geodesic in both directions. Unlike a cusp, the complete cylinder has two flaring ends and infinite area.

9.2.1 The coordinate calculation behind the cylinder

The two coordinate differentials in (9.1) are

$$\begin{aligned} \mathrm{d}x &= e^{\tau}(\operatorname{sech}^2 r\ \mathrm{d}r + \tanh r\ \mathrm{d}\tau), \\ \mathrm{d}y &= e^{\tau}(-\operatorname{sech} r \tanh r\ \mathrm{d}r + \operatorname{sech} r\ \mathrm{d}\tau). \end{aligned}$$

The coefficient of $\mathrm{d}r\ \mathrm{d}\tau$ in $\mathrm{d}x^2 + \mathrm{d}y^2$ is zero. The coefficient of $\mathrm{d}r^2$ is $e^{2\tau}\operatorname{sech}^2 r$ and that of $\mathrm{d}\tau^2$ is $e^{2\tau}$. Dividing by $y^2 = e^{2\tau}\operatorname{sech}^2 r$ gives (9.2). In particular, the absence of a mixed term proves orthogonality, rather than merely suggesting it from a picture of the coordinates. The area element of the quotient cylinder is $\ell \cosh r\ \mathrm{d}r\ \mathrm{d}t$. Integrating it from $-b$ to b gives $2\ell \sinh b$, whereas the two boundary lengths together are $2\ell \cosh b$. Area and boundary length therefore have different dependence on the half-width even in this particularly simple model.

9.3 Distance between two disjoint axes

Let one axis be the imaginary axis and let a second have endpoints $0 < u < v$. It is the semicircle with center $c = (u + v)/2$ and radius $R = (v - u)/2$. Put $s = \sqrt{uv}$, so $c^2 = R^2 + s^2$. The circle centered at zero with radius s is perpendicular to both axes and is their common perpendicular.

To calculate the distance D, write its intersection with the second semicircle as $se^{i\theta}$, with $0 < \theta < \pi/2$. Orthogonality gives $\sin\theta = R/c$. Along a circle centered at zero the half-plane length element is $\mathrm{d}\theta/\sin\theta$, so

$$D = \int_{\theta}^{\pi/2} \frac{\mathrm{d}v}{\sin v} = -\log\tan(\theta/2), \qquad \cosh D = \frac{1}{\sin\theta} = \frac{u+v}{v-u}. \tag{9.4}$$

For a direct check that this connector realizes the infimum, a point of the second axis has $x = c + R\cos\varphi$, $y = R\sin\varphi$. Its distance to the first axis is $\operatorname{arsinh}(x/y)$. The quotient x/y has its minimum at $\cos\varphi = -R/c$, where it equals s/R. Thus the minimum distance is $\operatorname{arsinh}(s/R)$, equal to the displayed value of D.

This calculation assumes the axes have four distinct ideal endpoints. That hypothesis will follow from discreteness in the next argument; it should not be inferred merely from disjointness in the interior.

9.4 Separation of the lifts of a simple closed geodesic

Lemma 9.1. *Let γ be a primitive simple closed geodesic of length ℓ on S. Any two distinct lifts of γ to $\mathbb{H}$ have distance D satisfying*

$$\sinh(D/2)\sinh(\ell/2) \geq 1.$$

Proof. Distinct lifts cannot cross. Otherwise their projections would give a self-intersection of the simple geodesic; local uniqueness of a geodesic with prescribed tangent deals with a putative tangency. They also cannot share just one ideal endpoint. To prove this assertion, move that endpoint to infinity and the other endpoint of the first axis to zero. The corresponding deck transformations have the form

$$A(z) = \lambda z, \qquad B(z) = \mu z + t, \qquad \lambda > 1, \quad \mu > 0.$$

They are hyperbolic, so $\mu \neq 1$. Different second endpoints imply $t \neq 0$. Their commutator is a translation by $(\lambda - 1)t$, hence a nonzero translation. Conjugating this translation by A^{-n}

produces nonidentity translations whose lengths tend to zero. This contradicts discreteness. If both endpoints were shared, the two axes would be the same, contrary to their choice.

Thus the axes have positive distance. Normalize the first to the imaginary axis. By interchanging its endpoints if necessary, the other has endpoints $0 < u < v$. Put $a = \ell/2$ and choose determinant-one representatives

$$A = \begin{pmatrix} e^{a} & 0 \\ 0 & e^{-a} \end{pmatrix}, \qquad P = \frac{1}{\sqrt{v-u}} \begin{pmatrix} v & u \\ 1 & 1 \end{pmatrix}, \qquad B = PAP^{-1}.$$

The translation attached to the second lift has the same length ℓ, because the lifts belong to the same closed geodesic. Replacing that deck transformation by its inverse, if necessary, gives precisely the displayed B in $\mathrm{PSL}_2(\mathbb{R})$. Matrix multiplication yields

$$\begin{aligned} \operatorname{tr}(AB^{-1}) &= 2\left(\cosh^2 a - \frac{u+v}{v-u}\sinh^2 a\right) \\ &= 2\left(1 - (\cosh D - 1)\sinh^2 a\right) < 2. \end{aligned}$$

The transformation AB^{-1} is not the identity: equality of A and B as transformations would force equality of their axes. A nonidentity element of a discrete torsion-free group cannot be elliptic, so its determinant-one trace cannot lie strictly between -2 and 2. The particular representative computed above therefore has trace at most -2. It follows that

$$(\cosh D - 1)\sinh^2 a \geq 2.$$

Using $\cosh D - 1 = 2\sinh^2(D/2)$ and taking positive square roots gives the result. □

Three points in this proof carry real mathematical content. Simplicity gives disjoint axes; discreteness rules out an asymptotic pair; equality of the translation lengths makes the short trace identity useful. Omitting any of them would leave an unsupported inference. Torsion-freeness and discreteness together are also what exclude the elliptic alternative, including rotations of infinite order.

9.4.1 Checking the trace computation

The matrix argument is short because it encodes the geometric separation in one scalar. To make that encoding reproducible by hand, write $h = e^{a}$, so that

$$P^{-1} = \frac{1}{\sqrt{v-u}} \begin{pmatrix} 1 & -u \\ -1 & v \end{pmatrix}, \qquad B = \frac{1}{v-u} \begin{pmatrix} vh - uh^{-1} & uv(h^{-1} - h) \\ h - h^{-1} & vh^{-1} - uh \end{pmatrix}.$$

The determinant is one, so the diagonal entries of B^{-1} are those of B in reverse order. Multiplication by $A = \operatorname{diag}(h, h^{-1})$ gives

$$\operatorname{tr}(AB^{-1}) = \frac{2v - u(h^2 + h^{-2})}{v-u} = 2\left(1 - \frac{2u}{v-u}\sinh^2 a\right).$$

Since $2u/(v-u) = \cosh D - 1$, this is exactly the identity used above. It is the trace of the indicated determinant-one representative, not an absolute trace chosen after the calculation. Its strict upper bound 2 comes from $D > 0$ and $a > 0$. Excluding the interval $(-2, 2)$ therefore leaves the alternative $\operatorname{tr}(AB^{-1}) \leq -2$, which is why the final inequality has the direction needed for a collar.

The equality case is also geometrically informative. If $\sinh(D/2)\sinh(\ell/2) = 1$, the computed trace is -2. The product is nonidentity, so it is parabolic. In particular, equality cannot

occur when the quotient is compact. Compactness rules out a parabolic element because the sequence of points high in its normalized cusp has injectivity radius tending to zero, whereas the positive continuous injectivity-radius function has a positive minimum on a compact surface. The standard collar width remains the same universal number; a particular compact surface may have extra room beyond it.

9.5 The embedded collar theorem

Theorem 9.2. *A primitive simple closed geodesic of length $\ell > 0$ has an embedded open collar of half-width*

$$w(\ell) = \operatorname{arsinh} \frac{1}{\sinh(\ell/2)} = \log \coth(\ell/4). \tag{9.5}$$

Its metric in coordinates $|r| < w(\ell)$ and $t \in \mathbb{R}/\mathbb{Z}$ is

$$\mathrm{d}r^2 + \ell^2 \cosh^2 r \; \mathrm{d}t^2.$$

Every closed collar with strictly smaller half-width is embedded as well.

Proof. Lemma 9.1 says that distinct lifts of the geodesic are separated by at least $2w(\ell)$. Their open distance-$w(\ell)$ neighborhoods cannot meet: a point in both would join the axes by a path of length less than $2w(\ell)$. The stabilizer of an axis consists of translations along it. An orientation-preserving isometry reversing its endpoints would be an elliptic half-turn, which is excluded. The translation lengths form a discrete additive subgroup of $\mathbb{R}$, hence are integer multiples of a least positive length. Primitivity says that this length is ℓ. Consequently the strip about one lift descends injectively after quotienting by precisely this cyclic stabilizer. Formula (9.3) gives the metric. The identity between the two expressions for w follows by writing $\operatorname{arsinh} u = \log(u + \sqrt{1+u^2})$ and using $\coth(x/2) = \operatorname{csch} x + \coth x$. □

At the limiting boundary, different closed strips can touch. We therefore retain the word “open” in the sharp statement. Most integral estimates are unchanged by taking limits from smaller widths; an assertion about embedded boundary circles is not.

The collar becomes wider when the central geodesic becomes shorter:

$$w(\ell) = \log(4/\ell) + O(\ell^2) \quad (\ell \to 0).$$

Its area and the length of either limiting boundary curve are

$$A_c(\ell) = 2\ell \sinh w(\ell) = \frac{2\ell}{\sinh(\ell/2)}, \qquad b_c(\ell) = \ell \cosh w(\ell) = \ell \coth(\ell/2). \tag{9.6}$$

Thus $A_c(\ell) \to 4$ and $b_c(\ell) \to 2$ as $\ell \to 0$. The geometric width diverges while the area stays bounded. This combination is a recurring feature of negatively curved ends and explains why area alone does not prevent long thin regions.

9.6 Conformal modulus of the collar and cyclic cover

A cylinder of Euclidean circumference one and height m has conformal modulus m. In (9.3) define

$$Y(r) = \frac{1}{\ell} \int_0^r \operatorname{sech} s \, \mathrm{d}s = \frac{1}{\ell} \arctan(\sinh r).$$

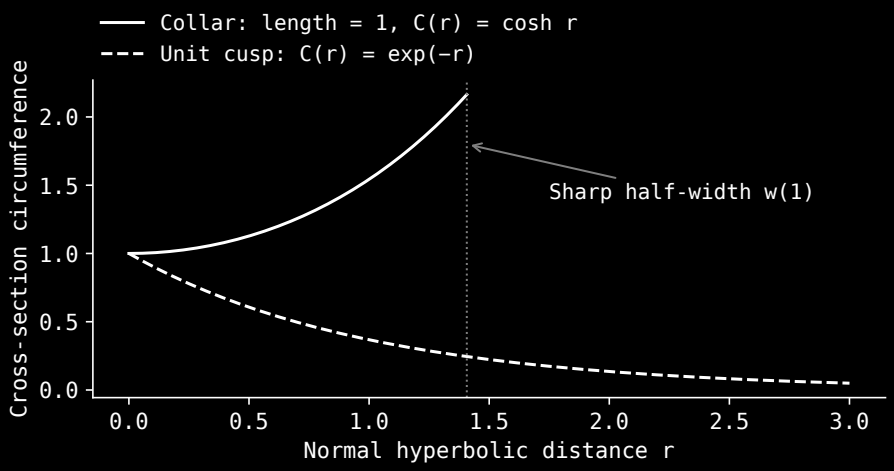


Figure 9.1: Exact circumference profiles in two different models. For the positive half of the embedded collar with $\ell = 1$, the circumference is $\cosh r$ and r is distance from the core. For the unit cusp it is e^{-r} and r is depth from its initial horocycle. The curves are analytic model formulas, not measurements.

Then $\mathrm{d}s^2 = \ell^2 \cosh^2 r\,(\mathrm{d}Y^2 + \mathrm{d}t^2)$, so (t, Y) are conformal cylinder coordinates. The complete cyclic cover has height π/ℓ; the embedded collar has modulus

$$m_c(\ell) = \frac{2}{\ell} \arctan \frac{1}{\sinh(\ell/2)}. \tag{9.7}$$

The difference between these two moduli has the particularly useful expression

$$\frac{\pi}{\ell} - m_c(\ell) = \frac{2}{\ell} \arctan(\sinh(\ell/2)) < 1. \tag{9.8}$$

Indeed, $f(x) = \arctan(\sinh(x/2))$ has $f(0) = 0$ and $f'(x) = \frac{1}{2}\operatorname{sech}(x/2) < 1/2$ for $x > 0$. Integrating the strict derivative inequality proves the bound. It also gives $m_c(\ell) = \pi/\ell - 1 + O(\ell^2)$ at a pinching geodesic.

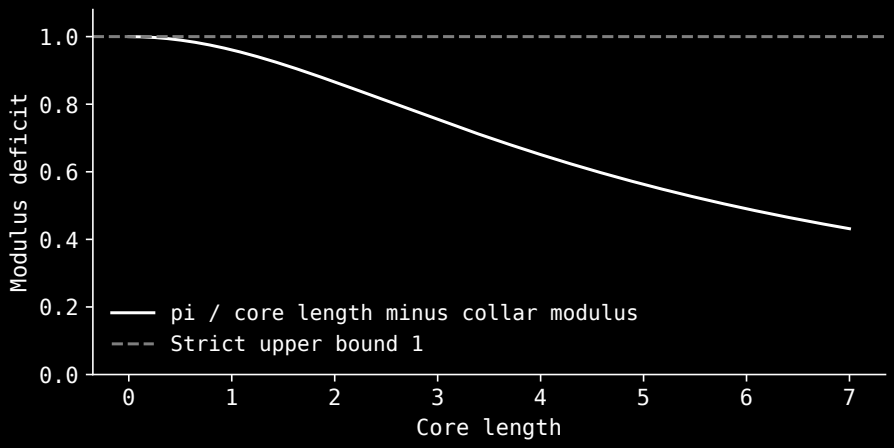


Figure 9.2: The exact deficit $\Delta(\ell) = \pi/\ell - m_c(\ell) = 2\arctan(\sinh(\ell/2))/\ell$. It lies strictly between zero and one for every $\ell > 0$, tends to one as $\ell \downarrow 0$, and tends to zero as $\ell \to \infty$. The plotted formula illustrates the proved bound.

9.6.1 Why essential annuli have monotone modulus

The comparison of an embedded annulus with its cyclic cover uses modulus monotonicity. We give the variational argument, including the direction of the inequality, because it is easy to confuse it with ordinary area monotonicity. Here only annuli of finite positive modulus are needed. The standard conformal classification of an annulus supplies a flat cylinder $\mathbb{R}/\mathbb{Z} \times (0, m)$ of circumference one and height m; its modulus is m.

For a family $\mathcal{C}$ of closed curves in a domain A, define its extremal length by

$$\mathrm{EL}_A(\mathcal{C}) = \sup_{\rho} \frac{\left(\inf_{\gamma \in \mathcal{C}} \int_\gamma \rho |\,\mathrm{d}z|\right)^2}{\int_A \rho^2 \,\mathrm{d}x\,\mathrm{d}y}.$$

The supremum ranges over nonnegative Borel densities of finite positive area. Curves for which the integral is infinite are allowed; they do not lower the infimum. A conformal change of coordinate multiplies the line density by the derivative modulus and the area density by its square, so the quotient and its supremum are conformally invariant. Let $\mathcal{C}_A$ consist of closed curves going once around the annulus in either orientation.

On the flat cylinder, the constant density $\rho = 1$ has area m. Every curve in $\mathcal{C}_A$ lifts to a path with net horizontal displacement one in absolute value. Its Euclidean length is therefore at least one; a horizontal circle attains one. This density gives $\mathrm{EL}_A(\mathcal{C}_A) \geq 1/m$. For the reverse bound, take any admissible ρ. Almost every horizontal circle has length $L(y) = \int_0^1 \rho(x, y)\,\mathrm{d}x$, and Cauchy–Schwarz gives

$$\int_0^m L(y)^2 \,\mathrm{d}y \leq \int_0^m \int_0^1 \rho(x, y)^2 \,\mathrm{d}x\,\mathrm{d}y.$$

Thus the infimum of the squared lengths of all essential curves is at most the mean squared length of those horizontal circles, which is at most $\mathrm{area}_\rho(A)/m$. Dividing by area and taking the supremum proves

$$\mathrm{EL}_A(\mathcal{C}_A) = \frac{1}{\operatorname{mod} A}. \tag{9.9}$$

No minimizing curve for a rough density is required: the infimum and the integral estimate suffice.

Now let A be an embedded essential subannulus of B. Its core goes once around B; for an embedded annulus this is the usual primitive essential inclusion, rather than a covering of higher degree. Given a density ρ on B, restrict it to A. The curves of $\mathcal{C}_A$ are among the curves of $\mathcal{C}_B$, so their infimum length can only increase. The area can only decrease. Whenever the restricted area is positive, this gives

$$\frac{\inf_{\gamma \in \mathcal{C}_B} L_\rho(\gamma)^2}{\mathrm{area}_\rho(B)} \leq \frac{\inf_{\gamma \in \mathcal{C}_A} L_\rho(\gamma)^2}{\mathrm{area}_\rho(A)} \leq \mathrm{EL}_A(\mathcal{C}_A).$$

If the restricted area is zero, the circle-family argument in a conformal coordinate of A supplies essential curves of zero ρ-length, and the leftmost numerator is zero as well. Taking the supremum over densities on B and using (9.9) yields $1/\operatorname{mod} B \leq 1/\operatorname{mod} A$, or $\operatorname{mod} A \leq \operatorname{mod} B$. An annulus of infinite modulus contains finite essential subannuli with arbitrarily large modulus, as is immediate in its cylindrical coordinate. Applying the finite comparison to those subannuli shows that it cannot embed essentially in an annulus of finite modulus. This also covers an apparently infinite candidate in the corollary below. The inverse relation is why an essential subannulus has no greater modulus even though many of its admissible curve lengths may be larger under a chosen density.

Corollary 9.3. *Let M_γ be the supremum of the moduli of embedded essential annuli whose core is freely homotopic to a primitive simple closed geodesic γ of length ℓ. Then*

$$\frac{\pi}{\ell} - 1 < M_\gamma \leq \frac{\pi}{\ell}.$$

Proof. The embedded collar provides an admissible annulus, so $M_\gamma \geq m_c(\ell) > \pi/\ell - 1$. Every other admissible annulus lifts to the cyclic cover corresponding to the subgroup generated by γ. The covering-space lifting criterion applies because its fundamental group maps into that subgroup. The lift is injective: points with the same lift have the same image in the original surface, where the annulus inclusion is injective. It is also essential. Monotonicity of conformal modulus for essential subannuli therefore bounds its modulus by that of the full cyclic cover, π/ℓ. Taking the supremum proves the upper bound. □

The statement uses a supremum and does not assert attainment. If the surface itself is the complete cyclic cylinder, its whole annulus has modulus π/ℓ, so the upper equality occurs. A universal strict upper bound would be false without excluding that case. For a parabolic cyclic cover the modulus is infinite and the reciprocal conformal width is zero; this is the limiting cusp convention, not a positive finite annular modulus.

9.7 Uniform expansion for a symmetric annular double covering

The cylinder calculation also resolves a useful distinction in annular dynamics. Strict inclusion of one annulus in another does not, by itself, give a uniform density gap: concentric annuli can be arbitrarily close. A covering-degree condition forces a genuine conformal separation and changes the conclusion.

Theorem 9.4. *Let $A \subset A'$ be annuli in $\mathbb{C} \setminus \{0\}$ invariant under the reflection $z \mapsto 1/\overline{z}$. Suppose both contain the unit circle $\mathbb{T}$ as an essential curve, $0 < \mu' \leq \operatorname{mod} A' < \infty$, and $g : A \to A'$ is an unramified holomorphic covering of degree two. Then, for every $z \in \mathbb{T}$,*

$$\frac{\lambda_{A'}(g(z))|g'(z)|}{\lambda_{A'}(z)} = \frac{\lambda_A(z)}{\lambda_{A'}(z)} \geq \frac{1}{Q(R_{\mu'})} > 1, \qquad R_{\mu'} = \frac{\pi}{2\mu'} + \operatorname{arsinh} 1,$$

where $Q(R) < 1$ is the definite-contraction function in Proposition 7.13. No upper bound on $\operatorname{mod} A'$ *is needed for this conclusion.*

Proof. Reflection is an antiholomorphic isometry of the hyperbolic metric. Its fixed curve $\mathbb{T}$ is geodesic: a geodesic initially tangent to the fixed curve is taken to a geodesic with the same initial data, so uniqueness keeps it in the fixed set. Since the circle is essential, it is the core geodesic of A'. Write its length as $\ell = \pi/\operatorname{mod} A'$. The annulus is the complete cyclic cylinder with coordinates (9.3) and core $r = 0$.

The modulus relation for a degree-two annular covering is

$$\operatorname{mod} A = \tfrac{1}{2} \operatorname{mod} A' = \frac{\pi}{2\ell}.$$

For example, this follows by representing the target as a horizontal strip modulo one translation: the degree-two cover corresponds to the subgroup generated by twice that translation, doubling the circumference and halving the modulus.

Put $r_0 = \operatorname{arsinh} 1$. We claim that for every point z on the core,

$$\operatorname{dist}_{A'}(z, A' \setminus A) \leq \ell/2 + r_0.$$

If the distance were larger, choose $\eta > 0$ so that it exceeds $\ell/2 + r_0 + \eta$. Every point with $|r| < r_0 + \eta$ can be reached from z by following the shorter core arc, of length at most $\ell/2$, and then a normal segment of length $|r|$. Hence the entire central band $|r| < r_0 + \eta$ would lie in A. Its modulus is

$$\frac{2}{\ell} \arctan(\sinh(r_0 + \eta)) > \frac{2}{\ell} \arctan 1 = \frac{\pi}{2\ell} = \operatorname{mod} A,$$

contradicting essential-annulus monotonicity. The band is essential in A because it contains the essential core $\mathbb{T}$.

Since $\ell \le \pi/\mu'$, the claimed distance is at most $R_{\mu'}$. Apply the boundary-distance density gap of Proposition 7.13 to the inclusion $A \subset A'$. It gives $\lambda_{A'}(z)/\lambda_A(z) \le Q(R_{\mu'}) < 1$. Finally a holomorphic covering is a local hyperbolic isometry, so $\lambda_{A'}(g(z))|g'(z)| = \lambda_A(z)$. Substitution proves the statement. □

The norm in this theorem uses the ambient A' metric at both the initial and final tangent spaces. With the intrinsic A metric in the denominator, the covering norm is exactly one. The expansion comes from comparing those two source metrics. The covering-degree condition, rather than strict inclusion alone, is what supplies a uniform distance bound and hence a uniform factor greater than one.

9.8 Intersecting curves and simultaneous collars

Suppose another closed geodesic crosses γ. A lift crossing the central axis must leave its width-$w(\ell)$ strip in both time directions. A different complete geodesic could remain in the strip in one direction only if it shared an ideal endpoint with the axis, which the same discreteness argument excludes for axes of deck transformations. This crossing segment has length at most one period of the other closed geodesic. Otherwise a full period would lie inside the embedded collar, putting the entire closed geodesic in that annulus. Its homotopy class would then be a power of the collar core (or trivial); uniqueness of the geodesic axis, and the absence of a contractible closed geodesic, would contradict its crossing of the core. Along the crossing segment the signed normal coordinate changes from one boundary value to the other. Since the metric dominates $|\,\mathrm{d}r|$, that segment has length at least $2w(\ell)$. Thus every crossing closed geodesic has length at least

$$2\,\operatorname{arsinh}\frac{1}{\sinh(\ell/2)}.$$

This is a quantitative version of the fact that a very short simple geodesic is difficult for another short geodesic to cross.

For several disjoint simple closed geodesics we use the classical *simultaneous collar theorem*: their open collars of half-width $w(\ell)$ are pairwise disjoint. The equal-length trace proof above establishes the embedding of each individual collar, while this additional theorem controls different central curves. The sharp theorem and its historical attribution to Keen and Buser are recorded in Basmajian's account [4]. An immediate consequence is that disjoint simple closed geodesics of lengths at most L satisfy

$$\operatorname{dist}_S(\gamma_1, \gamma_2) \ge w(\ell_1) + w(\ell_2) \ge 2w(L).$$

Here distance is the infimum over pairs of points on the two curves. The inequality concerns distinct geodesic representatives; two copies of the same free homotopy class have the same representative and must not be counted twice.

9.9 A strip geodesic and its Euclidean comparison

Consider the strip $\Sigma = \{0 < \operatorname{Im} z < \pi\}$, whose density is $1/\sin y$. Let a hyperbolic geodesic have distinct ideal endpoints $u < v$ on the lower boundary. Write $m = (u+v)/2$ and $A = (v-u)/2$. The exponential map sends the strip to $\mathbb{H}$, and sends this geodesic to the orthogonal semicircle with endpoints e^u, e^v. The circle equation simplifies to

$$\cos y = \frac{\cosh(x-m)}{\cosh A}, \qquad u < x < v. \tag{9.10}$$

In particular $0 < y < \pi/2$. We claim that this arc lies below the Euclidean semicircle with diameter $[u, v]$:

$$(x - m)^2 + y^2 \le A^2.$$

By symmetry it suffices to put $t = x - m \in [0, A)$. Differentiation of (9.10) gives $y' = -\tanh t \cot y$, and hence

$$\frac{\mathrm{d}}{\mathrm{d}t}(t^2 + y^2) = 2(t - y \cot y \tanh t) \ge 0.$$

The last inequality follows from $0 < y \cot y < 1$ and $0 \le \tanh t \le t$. As $t \to A$, $y \to 0$, so the increasing expression has upper limit A^2. This proves the claim for every $A > 0$, without a small-diameter restriction. Reflection across the midline gives the corresponding upper-boundary statement. The calculation compares two explicit arcs; it does not depend on an unlabelled appeal to a harmonic maximum principle.

9.10 Solved investigations: length and conformal width

Investigation 9.5 (An exact annulus check)**.** For $t > 0$, determine the modulus and central hyperbolic geodesic length of $A_t = \{e^{-t} < |z| < e^t\}$.

Solution. A logarithm identifies the covering strip with $-t < \operatorname{Re} w < t$ and period $2\pi i$. Its modulus is $2t/(2\pi) = t/\pi$. In the curvature--1 normalization the density on $|z| = 1$ is $\pi/(2t)$. The central circle therefore has length $2\pi \cdot \pi/(2t) = \pi^2/t$. The product of modulus and core length is π, agreeing with the cyclic-cylinder formula. This check prevents two common normalization errors: forgetting that the annulus has logarithmic thickness $2t$, and confusing a circle's radius with its circumference.

Investigation 9.6 (A collar with an elementary width)**.** Let $\ell = 2 \operatorname{arsinh} 1$. Compute its collar half-width, collar area, and collar modulus.

Solution. Since $\sinh(\ell/2) = 1$, the half-width is $w = \operatorname{arsinh} 1$. Formula (9.6) gives $A_c = 2\ell = 4 \operatorname{arsinh} 1$, and (9.7) gives

$$m_c = \frac{2}{\ell} \arctan 1 = \frac{\pi}{4 \operatorname{arsinh} 1}.$$

The full cyclic modulus is twice this value. These are exact symbolic quantities; no numerical approximation is needed to establish the relations. The example also shows that an embedded collar need not occupy almost all the cyclic modulus unless the central geodesic is small.

Investigation 9.7 (Reciprocal estimates and their range)**.** Assume $0 < \ell < \pi$. Convert the bounds for M_γ into bounds for its reciprocal width.

Solution. All denominators are positive under the stated restriction, so inversion gives

$$\frac{\ell}{\pi} \le \frac{1}{M_\gamma} < \frac{\ell}{\pi - \ell}.$$

The upper estimate exceeds ℓ/π by a relative factor at most $\ell/(\pi - \ell)$. Thus the cyclic width ℓ/π is an accurate first approximation when ℓ is small. If $\ell \ge \pi$, the lower bound $\pi/\ell - 1$ is nonpositive and cannot be inverted to produce a positive upper bound. The exact collar modulus remains usable in that range. A valid inequality can become useless after inversion when its positive-range hypothesis is lost.

Investigation 9.8 (A crossing obstruction)**.** Suppose a primitive simple geodesic satisfies $\sinh(\ell/2) = 1/10$. Give a lower bound for the length of every closed geodesic crossing it.

Solution. Its collar half-width is $\operatorname{arsinh} 10$, so every crossing closed geodesic has length at least $2 \operatorname{arsinh} 10$. The conclusion concerns actual intersections of geodesic representatives. A curve can wind many times in the collar while staying on one side, but that does not make it cross the central geodesic. Conversely, the normal-coordinate proof is insensitive to how obliquely a crossing occurs: any such segment must change its signed distance by the full amount $2w$.

Chapter 10

Area, thin regions, and quantitative geometric control

10.1 Area is governed by topology

A complete finite-area hyperbolic surface can contain arbitrarily long thin regions, but its total area is constrained by its topology. The normalization matters here as much as it did for collars. For curvature -1, the relevant identity is

$$\operatorname{area}(S) = -2\pi\chi(S),$$

when the finite-type surface has no ordinary boundary and all its ends are cusps. We derive the constant and sign rather than treating them as a convention.

In the half-plane choose the oriented orthonormal frame $e_1 = y\partial_x$, $e_2 = y\partial_y$. Direct differentiation of the metric gives

$$\nabla_{e_1} e_1 = e_2, \qquad \nabla_{e_1} e_2 = -e_1, \qquad \nabla_{e_2} e_1 = \nabla_{e_2} e_2 = 0.$$

If a unit tangent makes angle θ with e_1, its signed geodesic curvature satisfies

$$k_g \, \mathrm{d}s = \mathrm{d}\theta + \frac{\mathrm{d}x}{y}. \tag{10.1}$$

The sign convention is that a positively oriented boundary has the region to its left. Since

$$\mathrm{d}\left(\frac{\mathrm{d}x}{y}\right) = \frac{\mathrm{d}x \wedge \mathrm{d}y}{y^2},$$

Stokes' theorem converts the connection term into area.

For a positively oriented geodesic polygon with interior angles $\alpha_1, \dots, \alpha_n$, the total turning of its tangent, including the exterior corner angles, is 2π. Along each geodesic edge $k_g = 0$. Integrating (10.1) and adding the corner turns therefore gives

$$\operatorname{area}(P) = (n-2)\pi - \sum_{j=1}^{n} \alpha_j. \tag{10.2}$$

In particular a geodesic triangle has area equal to its angle defect $\pi - (\alpha + \beta + \gamma)$. An ideal triangle has area π: exhaust it by triangles whose vertices tend to its three ideal vertices; their angles tend to zero and their areas tend to the area of the ideal triangle. This argument also makes clear that the ideal vertices themselves are infinitely distant rather than missing finite-area corners that require additional angles.

Theorem 10.1 (Gauss–Bonnet in the present normalization)**.** *Let S be a connected orientable finite-type hyperbolic surface of genus g, with n cusp ends and b compact geodesic boundary curves, and with no other ends. Then*

$$\operatorname{area}(S) = 2\pi(2g - 2 + n + b) = -2\pi\chi(S).$$

Proof. First consider a compact surface without boundary. Choose a sufficiently fine geodesic triangulation; its existence follows by subdividing coordinate disks and taking their unique short geodesic edges. If it has V vertices, E edges, and F faces, summing (10.2) over the triangular faces gives area $= \pi F - 2\pi V$, because the angles around every vertex sum to 2π. The relation $3F = 2E$ then turns this expression into $-2\pi(V - E + F)$.

For a compact surface with geodesic boundary, write V_i, V_b for the interior and boundary vertex counts, and E_i, E_b for the corresponding edge counts. The total angle sum is $2\pi V_i + \pi V_b$, since the boundary has no corners. Also $3F = 2E_i + E_b$ and $E_b = V_b$. Substitution into the same sum gives area $= -2\pi\chi$.

The same local calculation allows a smooth curved boundary. Indeed, sum (10.1) over small coordinate polygons covering the surface; the contributions from interior edges cancel with opposite orientations, and the angle bookkeeping is unchanged. The result is

$$-\operatorname{area}(K) + \int_{\partial K} k_g \, \mathrm{d}s = 2\pi\chi(K).$$

This is the form needed for cusp truncation. Cut off each cusp at a horocycle of length h_j, chosen inside an embedded cusp neighborhood. On the compact remainder K, each new boundary has geodesic curvature -1: in the normalized half-plane it is the top boundary of the retained region and is traversed from right to left. Thus its curvature integral is $-h_j$. The compact remainder has the same Euler characteristic as the punctured surface, so

$$\operatorname{area}(K) = -2\pi\chi(S) - \sum_j h_j.$$

Each removed cusp has area exactly h_j by Theorem 8.4. Adding these areas proves the formula. The topological identity $\chi(S) = 2 - 2g - n - b$ supplies its final form. □

The proof uses the elementary existence of a geodesic triangulation and the classification of the stated ends. It does not suppose that every complete hyperbolic surface has finite area. A funnel, for example, contains a half-cylinder with circumference proportional to $\cosh r$, and its area diverges. Finite topology without the end condition is insufficient.

10.2 Boundary length, doubling, and the limits of an area argument

For a compact surface with geodesic boundary, the area formula has no boundary-length term. A hyperbolic pair of pants has genus zero and three geodesic boundaries, so its area is always 2π. The classical construction by two right-angled hexagons permits arbitrary positive lengths for those three boundaries. Hence neither the total boundary length nor an individual boundary length can be bounded using Euler characteristic alone.

This observation explains the role of geometric decompositions in width identities. A formula relating total conformal width to boundary length cannot follow solely from Gauss–Bonnet. It must identify the appropriate rectangles, account for the portions of boundary they cover, and control their connecting regions. The area formula controls those connecting regions only after their geometry has been related to area with a proved estimate.

Doubling provides a useful bridge between bordered and unbordered surfaces. Near a geodesic boundary, normal coordinates have the metric $\mathrm{d}r^2 + \cosh^2 r\, \mathrm{d}\tau^2$ on one side of $r = 0$. This expression is even in r, so reflecting a second copy across the boundary gives a smooth hyperbolic metric. The double has twice the area and twice the Euler characteristic, since boundary circles have Euler characteristic zero. Applying the closed-surface area formula to the double recovers the bordered formula.

The same construction allows collar arguments near geodesic boundaries. A collar in the double restricts to a one-sided collar in either copy. However, a general geodesic arc perpendicular to the boundary need not double to a simple closed geodesic without checking its endpoints and intersections. The geometry of a particular arc family still matters. For a nonparabolic cyclic cover associated with a peripheral geodesic of length ℓ, the local reciprocal modulus is exactly ℓ/π. This local identity is immediate from the cylinder calculation; it does not by itself establish a global sum over all peripheral families.

10.3 A precise classical input for the thin part

We now describe where small injectivity radius can occur. We use one classical result in a form tailored to surfaces.

> **Small-displacement lemma for hyperbolic surfaces.** There exists $\varepsilon_M > 0$ such that, for every discrete torsion-free $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$ and $z \in \mathbb{H}$, the subgroup generated by
>
> $$\{\gamma \in \Gamma : d(z, \gamma z) < 2\varepsilon_M\}$$
>
> is trivial or cyclic.

This is the surface form of the Margulis lemma. The general theorem gives a virtually nilpotent small-displacement group; in a torsion-free discrete orientation-preserving group of hyperbolic-plane isometries such an elementary group is cyclic. The standard Lie-group argument uses a neighborhood in which repeated commutators contract, together with compactness of bounded-displacement sets modulo the point stabilizer; an exposition of this route is given in [5]. We invoke that theorem here and do not assert a numerical optimal value for ε_M.

Choose once and for all

$$0 < \varepsilon < \min\{\varepsilon_M, \operatorname{arsinh}(1/2)\}. \tag{10.3}$$

The ε-thin and thick subsets are

$$S_{<\varepsilon} = \{p : \mathrm{inj}_S(p) < \varepsilon\}, \qquad S_{\geq\varepsilon} = \{p : \mathrm{inj}_S(p) \geq \varepsilon\}.$$

The strict inequality in the thin set makes its components open. We will derive their shapes and area directly rather than attaching a picture to the word “thin.”

10.4 Exact displacement tubes

For a hyperbolic transformation of translation length ℓ, substitute (9.1) and its dilated image into (8.1). At signed distance r from the axis the displacement δ satisfies

$$\cosh\delta = 1 + 2\sinh^2(\ell/2)\cosh^2 r, \qquad \sinh(\delta/2) = \sinh(\ell/2)\cosh r. \tag{10.4}$$

A power of order $|k|$ replaces ℓ by $|k|\ell$, so a primitive generator or its inverse minimizes displacement in the cyclic group. Its displacement is less than 2ε precisely when $\ell < 2\varepsilon$ and

$$|r| < r_\varepsilon(\ell), \qquad r_\varepsilon(\ell) = \operatorname{arcosh}\frac{\sinh\varepsilon}{\sinh(\ell/2)}. \tag{10.5}$$

At $\ell = 2\varepsilon$ the strict sublevel set is empty; there is no open annulus of zero half-width. For $\ell < 2\varepsilon$, the choice (10.3) implies $r_\varepsilon(\ell) < w(\ell)$, since

$$\frac{\sinh \varepsilon}{\sinh(\ell/2)} < \frac{\cosh(\ell/2)}{\sinh(\ell/2)} = \cosh w(\ell).$$

Thus the thin tube is narrower than the standard collar whenever its central geodesic is simple.

For a primitive parabolic transformation normalized to $Tz = z + 1$, formula (8.6) says that its displacement is less than 2ε exactly on

$$y > \frac{1}{2\sinh \varepsilon}. \tag{10.6}$$

The lower bound on y is greater than one, so this region lies in the embedded unit cusp already constructed. Its boundary length and area are both $2\sinh \varepsilon$.

Theorem 10.2. *For ε satisfying* (10.3)*, every connected component of the thin part of a complete orientable hyperbolic surface without boundary is either a cusp of the form* (10.6)*, or an annular neighborhood* (10.5) *of a primitive simple closed geodesic with length $\ell < 2\varepsilon$. In either component the injectivity radius equals half the displacement of the corresponding primitive generator.*

Proof. Lift the thin set to $\mathbb{H}$. It is the union of the sublevel sets $d(z, \gamma z) < 2\varepsilon$ over all nonidentity deck transformations. Group the transformations into maximal cyclic subgroups, taking a primitive generator in each. Such generators exist: for an axis the translation lengths are discrete; for a parabolic fixed point the translation parameters are discrete. The preceding calculations show that each nonempty sublevel region is a connected tube or horoball, and that powers give no larger region.

Two regions belonging to different maximal cyclic groups cannot overlap. If they contained a common point, both generators would have displacement less than $2\varepsilon_M$ there. The small-displacement lemma would put them in a single cyclic group, contrary to their maximality. A translate of a region is another region of the same kind. Thus each is precisely invariant under its stabilizer and projects to an embedded component.

The stabilizer of a hyperbolic region preserves its unique central axis. It is the corresponding maximal cyclic group, since endpoint reversal would be elliptic. The stabilizer of a parabolic region fixes its unique ideal tangency point and is its translation group, as proved above. Quotienting gives exactly an annulus or cusp. In the annular case the central axis consequently projects without self-intersection, proving the simplicity needed for comparison with the standard collar.

At a point in one component, any additional transformation with smaller displacement would also have displacement less than 2ε. The small-displacement lemma would put it in the same cyclic group. The primitive generator already minimizes displacement within that group, so no such smaller competitor exists. Formula (8.3) now gives the asserted injectivity radius. □

The theorem separates an algebraic input from explicit geometry. The Margulis lemma prevents unrelated short loops from occurring at the same point. The cylinder and cusp calculations identify the resulting regions and their scales. A numerical plot of short displacements could illustrate these regions, but could not supply the group-theoretic exclusion used in the proof.

10.5 Finite area makes the thick part compact

Proposition 10.3. *If S is complete, has no ordinary boundary, and has finite area, then $S_{\geq\varepsilon}$ is compact for every $\varepsilon > 0$. For $0 < \delta < 2\varepsilon$, a δ-separated subset of $S_{\geq\varepsilon}$ has cardinality at most*

$$\frac{\operatorname{area}(S)}{2\pi(\cosh(\delta/2) - 1)}$$

Proof. Open balls of radius $\delta/2$ centered at distinct points of a δ-separated subset are disjoint. Each is embedded because $\delta/2 < \varepsilon$, and each has area $2\pi(\cosh(\delta/2) - 1)$ by (8.4). Summing their areas proves the cardinality bound. In particular, successively choosing points outside the union of the radius-δ balls about earlier choices must terminate after finitely many steps. The resulting finite family covers the thick set. Since this works for arbitrarily small δ, that set is totally bounded. It is closed by Proposition 8.2, and hence complete as a closed subset of the complete surface. A complete totally bounded metric space is compact. □

This proof needs no chosen fundamental polygon and no explicit surface equation. It provides a geometric finite-cover bound directly from area and an injectivity-radius threshold. For very small δ, the denominator is asymptotic to $\pi\delta^2/4$, so the count has the expected two-dimensional scale.

For a boundaryless surface of genus g with n cusp ends, the annular thin components have disjoint essential nonparallel central curves. A topological pants decomposition shows there are at most $3g - 3 + n$ such curves when $2g - 2 + n > 0$. One way to see the count is to extend the disjoint curves to a maximal family. Its complementary pieces are pairs of pants; Euler characteristic gives $2g-2+n$ pieces. Counting their boundary circles gives $3(2g-2+n) = 2k+n$, hence $k = 3g - 3 + n$. Cusps correspond to the n punctures. The numerical count is topological; their lengths and the shape of the compact thick part vary with the conformal structure.

10.6 What a pinching collar becomes

The exact area of an annular thin component follows from its normal-coordinate metric:

$$A_\varepsilon(\ell) = 2\ell \sinh r_\varepsilon(\ell) = 2\ell\sqrt{\frac{\sinh^2 \varepsilon}{\sinh^2(\ell/2)} - 1}, \quad 0 < \ell < 2\varepsilon. \tag{10.7}$$

As $\ell \to 0$, this tends to $4 \sinh \varepsilon$. Each of its two boundary curves has length

$$\ell \cosh r_\varepsilon(\ell) = \frac{\ell \sinh \varepsilon}{\sinh(\ell/2)} \longrightarrow 2 \sinh \varepsilon.$$

A single ε-thin cusp has area and boundary length $2 \sinh \varepsilon$. Thus the two sides of a pinching annulus have the correct limiting area and circumference for two cusps.

The local metric explains more than those scalar limits. On the positive side set $r = r_\varepsilon(\ell) - u$, where u measures inward distance from its boundary. For fixed bounded $u \geq 0$,

$$\ell \cosh(r_\varepsilon(\ell) - u) = \ell \cosh r_\varepsilon(\ell) \cosh u - \ell \sinh r_\varepsilon(\ell) \sinh u \longrightarrow 2 \sinh \varepsilon\, e^{-u}.$$

The metric therefore tends, uniformly on every fixed bounded u-interval, to

$$\mathrm{d}u^2 + (2 \sinh \varepsilon)^2 e^{-2u}\, \mathrm{d}t^2,$$

which is exactly the cusp metric with boundary length $2 \sinh \varepsilon$. The same calculation applies to the other side. The core recedes to infinite distance because $r_\varepsilon(\ell) \to \infty$. These are explicit local model limits. A theorem about convergence of entire marked surfaces requires additional control of their thick parts and gluing parameters; it does not follow solely from a collar calculation.

10.6.1 An explicit error estimate for the cusp limit

Put $s = \sinh \varepsilon$ and $t = \sinh(\ell/2)$, and write the circumference at depth u as $L_\ell(u)$. The preceding calculation can be organized as

$$L_\ell(u) = \frac{\ell s}{t} e^{-u} + \frac{\ell}{t}(s - \sqrt{s^2 - t^2}) \sinh u. \tag{10.8}$$

This separates the leading cusp term from an error with a definite sign. For $0 < t \le s/2$, rationalizing the square root gives

$$0 \le s - \sqrt{s^2 - t^2} = \frac{t^2}{s + \sqrt{s^2 - t^2}} \le \frac{t^2}{s}.$$

Also $\ell/t \to 2$ and $|\ell/t - 2| = O(\ell^2)$, as follows by dividing the Taylor expansion $\sinh(\ell/2) = \ell/2 + \ell^3/48 + O(\ell^5)$ by $\ell/2$. Consequently, for every fixed $U < \infty$ there is a constant $C_{\varepsilon,U}$ and a positive ℓ_0 such that

$$\sup_{0 \le u \le U} |L_\ell(u) - 2se^{-u}| \le C_{\varepsilon,U}\ell^2 \qquad (0 < \ell < \ell_0). \tag{10.9}$$

Choose ℓ_0 smaller if necessary so that $U < r_\varepsilon(\ell)$; then every point being compared actually belongs to the positive half of the collar. Differentiating (10.8) in u gives the same bound for each fixed derivative order, with a possibly different constant. The coefficients of the metrics, their area densities, and all their derivatives therefore converge on this fixed coordinate cylinder. Positivity of $2se^{-u}$ there also gives convergence of inverse metric coefficients.

The dependence on U is essential. Near the core, $u = r_\varepsilon(\ell)$, the circumference is ℓ and the two sides of the collar join. No coordinate cylinder with a fixed finite depth reaches that point in the limit. Thus smooth convergence on every bounded depth interval and a change of global topology are compatible. The estimate supplies a rate for the local model, with no claim that unspecified gluing data on an entire surface converge.

10.7 How far a closed geodesic can enter a cusp

Finite area does not bound the metric diameter of a cusp, but the length of a particular closed geodesic bounds the depth of its excursions. This is a useful example of extracting compactness from an explicit formula rather than from a picture of a truncated end.

Proposition 10.4 (Length and cusp depth)**.** *Let $\{y > Y\}/\langle T\rangle$ be an embedded cusp with embedded boundary, where $Tz = z + 1$. If a closed geodesic of length L enters this cusp, its maximum depth $u = \log(y/Y)$ satisfies*

$$u \le \log\cosh(L/2) < L/2.$$

More precisely, an excursion of length a from the boundary back to the boundary has maximum depth $\log\cosh(a/2)$.

Proof. A closed geodesic cannot be contained entirely in the cusp. Such a loop would represent either a power of its parabolic generator or the trivial class. The first has no invariant axis and the second has no nonconstant closed geodesic, by lifting to $\mathbb{H}$. Hence every component of the geodesic inside the open cusp is an excursion with endpoints on its boundary. Lift one excursion to $y > Y$. Its complete geodesic is a semicircle centered on the real axis: a vertical line could not enter and leave the same horizontal line. Write its radius, which is its maximum height, as $H > Y$. A translation makes the two intersection points

$$-\sqrt{H^2 - Y^2} + iY, \qquad \sqrt{H^2 - Y^2} + iY.$$

The segment between them is minimizing in the cover. The distance formula gives

$$\cosh a = 1 + \frac{4(H^2 - Y^2)}{2Y^2} = 2(H/Y)^2 - 1.$$

As $a > 0$, the double-angle identity implies $H/Y = \cosh(a/2)$. Taking logarithms proves the exact excursion formula. Since $a \le L$ and $\log\cosh x < x$ for $x > 0$, the announced bounds follow. Winding around the cusp does not invalidate the calculation: the lift records the full horizontal separation of the endpoints before the translation quotient identifies them. □

The same proof explains an initially surprising effect. The height of an excursion can be exponentially large in its length, while its intrinsic depth is only logarithmic in that height. Coordinates and the metric measure different quantities. In particular, a Euclidean plot drawn high in the half-plane should not be used to estimate the amount of geodesic travel without the factor $1/y$.

Corollary 10.5 (Finiteness at a fixed length)**.** *On a complete finite-type hyperbolic surface with finite area and only cusp ends, there are only finitely many primitive closed geodesic images of length at most any fixed* $0 < L < \infty$.

Proof. Choose disjoint embedded cusp neighborhoods whose complement is compact. Such a choice uses the stated finite-type cusp description. The excursion bound shows that every geodesic in question lies in the compact set K_L obtained by attaching to that complement the finite-depth pieces $0 \le u \le \log\cosh(L/2)$ of its finitely many cusps. Cover K_L by finitely many small coordinate disks with compact closures, and choose a compact lift of each closure to $\mathbb{H}$. Their finite union C is compact, and every point of K_L has at least one lift in C.

Choose a point on a geodesic of length at most L, and a lift z of that point in C. The lifted parametrized geodesic lies on the axis of a hyperbolic deck transformation γ and satisfies $d(z, \gamma z) \le L$. Thus γC meets the closed L-neighborhood C_L of C. The latter is compact because closed bounded subsets of $\mathbb{H}$ are compact. Proper discontinuity implies that only finitely many deck transformations have $\gamma C \cap C_L \neq \varnothing$. Each hyperbolic transformation has a unique axis, hence determines at most one projected primitive geodesic image. The desired family is therefore finite. □

This is finiteness on one fixed surface at one fixed length scale. It gives no uniform lower bound on the systole across all surfaces of the same topological type: a family of surfaces may contain geodesics whose lengths approach zero. Nor does it give a numerical count from the area alone. The compact set and the finite deck-orbit search used in the proof depend on the actual surface.

10.7.1 A diameter bound for each connected thick piece

There is a different quantitative conclusion that depends only on area and the chosen thickness. Let K be a connected component of $S_{\ge\varepsilon}$ on a complete finite-area surface without boundary. By Proposition 10.3, K is compact. Choose a maximal δ-separated set $p_1, \dots, p_N$ in K, where $0 < \delta < 2\varepsilon$. It exists by the finite packing bound, and maximality means that the open radius-δ balls cover K. The graph joining indices whenever those balls have a common point in K is connected. If it were not, the unions corresponding to two groups of its components would give two disjoint, nonempty relatively open sets covering K.

Every edge of the graph joins centers at distance less than 2δ. A simple graph path between two centers has at most $N - 1$ edges, and any point of K is within δ of a center. The triangle inequality and the packing estimate therefore give

$$\operatorname{diam}_S K \le 2\delta N \le \frac{\delta \operatorname{area}(S)}{\pi(\cosh(\delta/2) - 1)}. \tag{10.10}$$

Here the diameter uses the ambient distance of S. The short segments joining centers may leave the thick set, so the proof does not claim the same bound for the intrinsic path metric restricted to K. It also does not bound distances between different thick components, which may be separated by very long thin collars.

10.8 Moving endpoints and buffering a rectangle

Two simple estimates often organize otherwise cumbersome geometric comparisons. First, in any metric space,

$$|d(p,q) - d(p',q')| \le d(p,p') + d(q,q'). \tag{10.11}$$

This follows by applying the triangle inequality in both directions. If both endpoint motions are at most a, the distance changes by at most $2a$. On a complete hyperbolic surface the same estimate applies to the lengths of minimizing geodesic segments. It does not automatically apply to longer geodesics chosen in an arbitrary fixed homotopy class; their lifts and endpoint correspondence must first be specified.

Second, metric comparison deep inside a subdomain gives a useful buffer estimate. If $U \subset V$ are hyperbolic surfaces and the V-distance from $z \in U$ to $V \setminus U$ is at least $L > 0$, the disk comparison proved in the local theory gives

$$1 \le \frac{\lambda_U(z)}{\lambda_V(z)} \le \coth(L/2). \tag{10.12}$$

For clarity, the upper bound follows by taking a universal disk cover of V centered at z. The disk of hyperbolic radius L lifts into U; its Euclidean radius is $\tanh(L/2)$. Schwarz–Pick applied to this local covering disk gives the upper density ratio $1/\tanh(L/2)$. The lower bound is domain monotonicity. The relative excess is at most $2/(e^L - 1)$, which is exponentially small when L stays away from zero and tends to infinity as $L \downarrow 0$.

Here is an exact application that keeps the endpoint issue visible. Let

$$\Sigma_1 = \{0 < y < 1\}, \qquad R_a = \{0 < x < a,\ 0 < y < 1\}, \qquad a > 2b > 0.$$

The strip density is $\lambda_{\Sigma_1} = \pi/\sin(\pi y)$, so it equals π on the midline. At $x + i/2$ the strip distance to the complement of R_a is

$$L(x) = \pi \min\{x, a - x\}.$$

Indeed every path to a vertical side costs at least π times its horizontal displacement, while the horizontal midline path attains that lower bound. Let $\Lambda(a,b)$ be the R_a-hyperbolic length of the buffered midline segment $b \le x \le a - b$. Then

$$\pi(a - 2b) \le \Lambda(a,b) \le \pi(a - 2b) - 4\log(1 - e^{-\pi b}). \tag{10.13}$$

To prove the upper bound, apply (10.12) pointwise and integrate. Symmetry about $a/2$ bounds the excess by

$$2\int_b^{a/2} \frac{2\pi}{e^{\pi x} - 1}\,\mathrm{d}x \le 4\pi \int_b^{\infty} \frac{\mathrm{d}x}{e^{\pi x} - 1} = -4\log(1 - e^{-\pi b}).$$

The lower bound is domain monotonicity. For each fixed positive buffer b, the additive error is independent of the rectangle length a.

The full open midline, with its endpoints tending to the vertical boundary, has infinite intrinsic length in the rectangle. Consequently the unbuffered expression cannot be obtained by silently setting $b = 0$ in a finite-error statement. The divergence of the logarithmic error is the correct warning. To apply this estimate to a canonical rectangle embedded in a more complicated surface, one must additionally identify the appropriate ambient metric and prove that the chosen endpoints lie within controlled distance of the buffered positions. Once that has been done, (10.11) converts the remaining endpoint adjustments into an explicit additive constant.

10.9 Solved investigations: area, degeneration, and buffers

Investigation 10.6 (A punctured torus and its cusp)**.** A complete once-punctured torus has an embedded cusp bounded by a horocycle of length h. Find its total area and the area of the compact complement. Explain why the answer does not determine its conformal structure.

Solution. Here $g = 1$, $n = 1$, so Gauss–Bonnet gives total area 2π. The cusp area is h, and the complement has area $2\pi - h$. The unit-cusp theorem guarantees admissible lengths $0 < h < 1$, and possibly larger ones depending on the surface. These area values do not determine the lengths of essential closed geodesics or their twist parameters. The family of hyperbolic structures on a punctured torus has genuine moduli even though its total area is fixed. Area supplies a topological normalization, not a complete geometric coordinate.

Investigation 10.7 (Finite chart counts on a genus-two surface)**.** Let S be a closed genus-two hyperbolic surface. Give a bound for the number of radius-δ disks needed to cover its ε-thick part, where $0 < \delta < 2\varepsilon$.

Solution. The area is 4π. Choose a maximal δ-separated family in the thick part. Its radius-δ disks cover that part, while its radius-$\delta/2$ disks are embedded and pairwise disjoint. Thus its cardinality N satisfies

$$N \leq \frac{2}{\cosh(\delta/2) - 1}.$$

Since N is an integer one may take the floor of the right-hand side. The bound is deliberately coarse; it uses no information about the particular length spectrum. It establishes a finite local atlas with a controlled scale without claiming that the displayed number is optimal or that a particular set of centers has been found.

Investigation 10.8 (A long boundary does not force large area)**.** Explain how a pair of pants can have a boundary component longer than any prescribed number while retaining area 2π. Identify the extra geometric work needed before using area to bound a total width.

Solution. The right-angled-hexagon construction realizes a hyperbolic pair of pants with any three prescribed positive boundary lengths. Choose one of them larger than the given number and keep the other two fixed. Gauss–Bonnet still gives 2π because the boundary is geodesic and the topology is unchanged. The surrounding geometry adjusts through thinner connecting regions. An estimate for a width sum must isolate annuli or rectangles carrying that width, prove how much boundary they account for, and control the geometry of the complement. Replacing those steps by "the area is fixed" would not bound the arbitrary prescribed perimeter.

Investigation 10.9 (Two cusp areas from one thin annulus)**.** Fix an admissible ε and let $\ell \downarrow 0$ in (10.7). Prove the limiting area directly and explain its factor of two.

Solution. Rewrite the expression as

$$A_\varepsilon(\ell) = 2\frac{\ell}{\sinh(\ell/2)}\sqrt{\sinh^2\varepsilon - \sinh^2(\ell/2)}.$$

The first ratio tends to 2, and the square root tends to $\sinh\varepsilon$, so the limit is $4\sinh\varepsilon$. A thin cusp cut at injectivity radius ε has area $2\sinh\varepsilon$. The factor two records the two sides of the central geodesic. The preceding normal-coordinate calculation shows that each side has its own cusp limit; it is not merely an arithmetic coincidence of the area formula.

Investigation 10.10 (Prescribing a cusp excursion)**.** In a cusp normalized by $Tz = z + 1$ with boundary $y = Y$, consider a geodesic segment joining iY to $k + iY$, where $k > 0$. Compute its length and maximum depth. What changes if k is an integer?

Solution. The semicircle has center $k/2$ and radius $H = \sqrt{Y^2 + k^2/4}$. The distance formula gives

$$a = 2\,\operatorname{arsinh}\frac{k}{2Y}, \qquad u = \log(H/Y) = \frac{1}{2}\log\left(1 + \frac{k^2}{4Y^2}\right).$$

These expressions satisfy $u = \log\cosh(a/2)$ exactly. For integer k, the endpoints project to the same point on the horocycle, and the segment becomes a based loop winding k times. It has a corner at its basepoint and is not a smooth closed geodesic. Thus the existence of such geodesic segments does not contradict the absence of a closed geodesic in a parabolic free homotopy class. As $k \to \infty$, both the depth and half the length grow logarithmically, with $a/2 - u \to \log 2$. The strict difference records the nonvertical part of the excursion.

Investigation 10.11 (A thick set and an intrinsic diameter)**.** Explain why the proof of (10.10) bounds an ambient diameter even when it supplies no bound for the length of paths constrained to the thick set itself.

Solution. The covering graph supplies a finite list of centers with successive ambient distances less than 2δ. A minimizing segment realizing one of those distances lies in the complete surface, but no step of the construction says that it remains in $S_{\geq\varepsilon}$. Concatenating these segments is therefore an admissible competitor for the surface distance and may fail to be an admissible path for the restricted intrinsic distance. The connectedness argument uses relatively open subsets of the thick component, not geodesic convexity of that component. To obtain an intrinsic bound one would need, for example, controlled replacement paths inside the thick set. The distinction is the same one that appears when an isometrically embedded coordinate disk permits shorter paths through the rest of a quotient surface: a local metric formula does not prescribe which global paths may be used in a minimization.

Investigation 10.12 (A unit buffer and a long rectangle)**.** Take $b = 1$ and $a > 2$ in (10.13). Show that the relative error in the strip approximation tends to zero as $a \to \infty$, and explain why the same conclusion fails if the buffer is allowed to shrink too quickly.

Solution. The additive error is bounded by the fixed constant $C = -4\log(1 - e^{-\pi})$. Dividing by the strip length $\pi(a-2)$ gives a relative error at most $C/[\pi(a-2)]$, which tends to zero. If $b = b(a) \downarrow 0$, however, the bound is asymptotic to $4\log(1/(\pi b(a)))$. With $b(a) = e^{-a}$ that bound has order a, so it no longer supplies a vanishing relative error. This does not assert that the upper bound is sharp in every such regime; it identifies exactly what the estimate does and does not justify. Long rectangles alone do not remove boundary singularities.

Investigation 10.13 (Area under an unramified finite covering)**.** Let $p : \widetilde{S} \to S$ be an unramified holomorphic covering of finite degree k between complete finite-area hyperbolic surfaces. Relate their areas and Euler characteristics.

Solution. The covering is a local isometry for the hyperbolic metrics, and almost every point has exactly k preimages. Integrating the area form over evenly covered coordinate pieces gives $\operatorname{area}(\widetilde{S}) = k\operatorname{area}(S)$. Equivalently, lift a finite cell decomposition after truncating the cusps: every cell has k lifts, giving $\chi(\widetilde{S}) = k\chi(S)$. Gauss–Bonnet makes the two identities consistent. The qualification "unramified" is necessary. A branched map has critical points and changes both the local metric relation and the Euler-characteristic formula; the degree alone is then insufficient.

The three global chapters connect the local Schwarz theory to the geometry of surfaces. A cusp has an exact exponential metric, a short simple geodesic has a quantitative embedded collar, and finite area gives a compact thick part. Modulus, length, and area interact through explicit formulas, while the named global inputs specify where topology and discrete-group theory enter the argument.

Chapter 11

Normal families and pointed conformal domains

Conformal geometry has two kinds of limits. A sequence of maps may converge on compact subsets of a fixed coordinate disk, while the regions occupied by their images may change shape, develop narrow passages, or move away from the chosen observation point. The relation between these limits is precise, but only after the observation point and the normalization of the maps have been specified. This chapter develops the analytic compactness tools. The next proves the kernel theorem and applies it to moving domains.

Throughout these two chapters a *domain* is a nonempty connected open subset of $\mathbb{C}$. A conformal disk is a simply connected proper plane domain; the word disk does not require its Euclidean boundary to be a circle or even a Jordan curve. We write $K \Subset U$ when K is a compact subset of U. No regularity of ∂U is included in this notation.

11.1 The topology of local uniform convergence

Pointwise convergence is too weak for most analytic arguments. For example, z^n tends to zero at every point of $\mathbb{D}$, but its behavior near the unit circle cannot be controlled by a bound that is uniform in both n and z. Local uniform convergence records exactly the compact subsets on which limits, contour integrals and derivatives can be interchanged.

Choose an increasing compact exhaustion (K_m) of a domain U, with every compact subset of U contained in some K_m. One possible choice is

$$K_m = \{z \in U : |z| \le m,\ \operatorname{dist}(z, \mathbb{C} \setminus U) \ge 1/m\}.$$

For $U = \mathbb{C}$ the distance condition is omitted. Empty initial sets cause no problem; their supremum is defined as zero below. On continuous functions on U put

$$d_U(f,g) = \sum_{m=1}^{\infty} 2^{-m} \min\{1, \sup_{K_m} |f - g|\}. \tag{11.1}$$

This is a metric. Its triangle inequality follows from the triangle inequality for each supremum and $\min(1, s+t) \le \min(1,s) + \min(1,t)$. If it vanishes, the functions agree on each K_m, hence everywhere. Convergence in this metric is equivalent to uniform convergence on every compact subset: one implication follows by examining each nonnegative summand; the other follows by first making the tail of the series small and then treating finitely many compact sets. Thus the particular exhaustion changes the formula, but not the topology.

Lemma 11.1 (Analytic limits and derivatives)**.** *If holomorphic functions f_n on U converge locally uniformly to f, then f is holomorphic, and $f_n^{(j)} \to f^{(j)}$ locally uniformly for every integer $j \geq 0$.*

Proof. Take a closed disk $\overline{B(a,R)} \subset U$ and $r < R$. Uniform convergence on its boundary allows passage to the limit in the Cauchy formula. The resulting integral is holomorphic for $|z-a| < R$ and equals $f(z)$. Moreover, the differentiated formula gives, for $|z-a| \leq r$,

$$|f_n^{(j)}(z) - f^{(j)}(z)| \leq \frac{j!R}{(R-r)^{j+1}} \sup_{|\zeta-a|=R} |f_n(\zeta) - f(\zeta)|.$$

The right side tends to zero. Finitely many smaller disks cover any given compact set, proving the assertion there. □

The same argument shows that the space of holomorphic functions is complete for (11.1). A Cauchy sequence is uniformly Cauchy on each K_m; the compatible limits form a continuous function on U, and the lemma makes it holomorphic. Completeness by itself does not give compactness. For that we need local bounds.

Theorem 11.2 (Locally bounded normal families)**.** *Let $\mathcal{F}$ be a family of holomorphic functions on U. Suppose that for every $K \Subset U$ there is $M_K < \infty$ such that $|f(z)| \leq M_K$ for all $f \in \mathcal{F}$ and $z \in K$. Every sequence in $\mathcal{F}$ has a subsequence converging locally uniformly to a finite holomorphic function.*

Proof. First fix a compact set $K \Subset U$. Cover it by finitely many disks whose closures, enlarged by a fixed positive amount, still lie in U. Cauchy's estimate on the enlarged disks gives a common bound for $|f'|$ on the smaller disks. Integration along line segments inside those disks gives equicontinuity on K. The assumed bound also gives uniform boundedness.

Here is the compactness step explicitly. Choose a countable dense subset of K. Repeatedly select subsequences on which the values converge at its first, second, and subsequent points; bounded subsets of $\mathbb{C}$ have convergent subsequences. The diagonal sequence converges at every selected point. Given $\varepsilon > 0$, equicontinuity and a finite sufficiently fine net in K reduce the uniform Cauchy condition to the Cauchy condition at finitely many of these points. Thus the diagonal sequence converges uniformly on K. Apply this construction successively to an exhaustion of U and take another diagonal sequence. Its limit is holomorphic by Lemma 11.1. □

The qualification "finite" matters. A family such as the constant functions $f_n = n$ has a spherical limit at infinity but no finite holomorphic limit. The theorem above concerns Euclidean local uniform convergence. Spherical normality for meromorphic maps is a different theorem with different hypotheses.

11.1.1 The finite-net step in the normal-family proof

It is useful to spell out how local boundedness becomes uniform compactness. Fix a nonempty $K \Subset U$ and choose $d > 0$ such that the closed Euclidean d-neighborhood K_d is compact in U. Let M bound all members of the family on K_d. Cauchy's derivative estimate on disks of radius $d/2$ gives

$$|f'(z)| \leq 2M/d \qquad (z \in K_{d/2}).$$

If $z, w \in K$ and $|z-w| < d/2$, their joining line segment stays inside $K_{d/2}$, and integration gives $|f(z) - f(w)| \leq (2M/d)|z-w|$. The segment need not stay in K, which is why the larger neighborhood was chosen. When $M = 0$, every function vanishes on K_d and there is nothing to prove.

For $M > 0$ and a desired error $\eta > 0$, choose $\delta < \min\{d/2, \eta d/(6M)\}$. From a countable dense subset of K, select finitely many points $x_1, \dots, x_N$ whose open δ-balls cover K. In a subsequence converging at each of these points, two sufficiently late functions f_j, f_k satisfy $|f_j(x_i) - f_k(x_i)| < \eta/3$ for every i. For any $x \in K$, choose one covering point and estimate

$$|f_j(x) - f_k(x)| \le |f_j(x) - f_j(x_i)| + |f_j(x_i) - f_k(x_i)| + |f_k(x_i) - f_k(x)| < \eta.$$

Thus convergence at finitely many net points controls the entire compact set. The number of points depends on K and δ; no fixed finite sampling set is being asserted to determine an arbitrary holomorphic function.

To pass from one compact set to the domain, choose successive subsequences converging uniformly on $K_1, K_2, \dots$, each subsequence contained in its predecessor. Take the jth element of the jth subsequence, choosing indices increasingly. For any fixed m, all but finitely many members of this diagonal sequence belong to the subsequence selected for K_m, so they converge uniformly there. Compatible limits on overlaps define one function on U. This final indexing step explains why the limit is uniform on every compact subset while there need not be one index that works simultaneously on the whole noncompact domain.

11.2 Zeros, injectivity and limits

Uniform convergence on a contour preserves a winding-number calculation when the limiting function has no zero on that contour. This observation is the useful content of Hurwitz's theorem in the present setting.

Lemma 11.3 (Persistence of isolated zeros)**.** *Suppose $f_n \to f$ locally uniformly on U, with f holomorphic and not identically zero. If f has a zero of multiplicity k at a, choose $r > 0$ so that $\overline{B(a,r)} \subset U$ and a is its only zero in that closed disk. For all sufficiently large n, f_n has exactly k zeros in $B(a,r)$, counted with multiplicity. Consequently, a locally uniform limit of zero-free holomorphic functions is either zero-free or identically zero.*

Proof. The minimum of $|f|$ on $|z - a| = r$ is positive. Eventually $|f_n - f| < |f|$ there. Rouché's theorem gives the equality of zero counts. If the f_n are zero-free, the existence of an isolated zero of a nonzero limit would contradict that equality. □

Proposition 11.4 (Limits of univalent maps)**.** *A locally uniform limit of univalent holomorphic functions on a domain is either constant or univalent.*

Proof. Fix $w \in U$ and remove the singularity at w in

$$q_n(z) = \frac{f_n(z) - f_n(w)}{z - w}, \qquad q_n(w) = f_n'(w).$$

Each q_n is zero-free. Indeed, injectivity deals with $z \neq w$, and a univalent holomorphic function has nonzero derivative. To recall why the latter assertion holds, a local power series whose first nonzero term has degree $k \ge 2$ takes nearby regular values k times in a small disk, by Rouché's theorem, contradicting injectivity.

The functions q_n converge locally uniformly to the analogous quotient q for f. This follows directly away from w and, near w, from Cauchy's formula for the divided difference. If $f(z) = f(w)$ for a point $z \neq w$, then q has a zero. Lemma 11.3 forces $q \equiv 0$, which makes f constant on U. A nonconstant limit is therefore injective. □

We will repeatedly use one additional elementary rule. If $f_n \to f$ locally uniformly and $z_n \to z \in U$, then $f_n(z_n) \to f(z)$. All the z_n eventually belong to a fixed compact neighborhood

of z, on which the convergence is uniform; continuity of f handles the moving argument. More generally, if $g_n \to g$ locally uniformly and g maps a compact set into the interior of U, then $f_n \circ g_n \to f \circ g$ uniformly on that compact set whenever the compositions are defined. A compact neighborhood of the image provides the needed common domain. This last qualification will prevent a hidden domain error when inverse maps are used below.

11.3 The basepoint and the stable component

Let (U_n, a_n) be pointed domains, where $a_n \in U_n$ and $a_n \to a \in \mathbb{C}$. The eventual interior of the sequence is

$$G = \bigcup_{N=1}^{\infty} \operatorname{int}\left(\bigcap_{n \geq N} U_n \right). \tag{11.2}$$

A point belongs to G precisely when some neighborhood of it is contained in every sufficiently late domain. The index at which this happens can depend on the neighborhood. Every compact subset of G is eventually contained in all the U_n, by a finite covering argument.

Definition 11.5 (Pointed kernel)**.** If $a \in G$, the pointed kernel of (U_n, a_n) is the connected component of G containing a. If $a \notin G$, the kernel is the degenerate symbol $\{a\}$, which is not being regarded as a domain. The sequence *converges in the kernel sense* to (U, a) if $a_n \to a$ and every subsequence has pointed kernel U. Degenerate convergence to $\{a\}$ is defined in the same way.

The nondegenerate kernel has an equivalent maximality description. It is the largest connected open set V containing a such that every compact subset of V lies in all sufficiently late U_n. For if V has this property, every point of V has a closed disk in V, and therefore lies in G. Connectedness puts all of V in the component containing a. Conversely, compact subsets of that component have the eventual inclusion property by (11.2). This description allows large regions to be recovered by progressively larger compact subsets; it does not demand that the entire kernel lie in one fixed tail domain.

Example 11.6 (A passage closes)**.** Let

$$U_n = B(0,1) \ \cup \ B(3,1) \ \cup \ \{x + iy : 0 < x < 3, \ |y| < 1/n\}, \qquad a_n = 0.$$

These are simply connected proper domains. Every vertical section is an interval containing its point on the real axis. Moving vertically to the axis and then along the axis contracts the domain to 0, so it is contractible. The two round chambers are joined by a passage that becomes arbitrarily narrow.

The eventual interior is $B(0,1) \cup B(3,1)$. A point in the intervening real segment has no neighborhood eventually contained in the domains, whereas every compact subset of either open chamber remains present. The pointed kernel is $B(0,1)$. The same statement holds for every subsequence, so there is kernel convergence to that disk. If the basepoint is instead 3, the limit is $B(3,1)$.

In particular, the small disk $B(3, 1/2)$ lies in every U_n but does not belong to the kernel based at 0. A definition requiring every disk contained in infinitely many domains to lie in the pointed limit would give the wrong answer. The component attached to the limiting basepoint is indispensable.

Example 11.7 (A kernel without convergence)**.** Set $U_{2n} = B(0,1)$ and $U_{2n+1} = B(0,2)$, with basepoint 0. The kernel of the entire sequence is $B(0,1)$. The odd subsequence has kernel $B(0,2)$, so the sequence does not converge in the kernel sense. The normalized conformal maps are alternately z and $2z$, which also do not converge. Passing from "has a kernel" to "converges to its kernel" is a substantive additional condition.

For increasing domains with a common basepoint the kernel is their union. Indeed, a compact subset of the union is covered by finitely many of the increasing open domains and hence lies in one of them. Every subsequence has the same union. For decreasing domains, the kernel is the component containing the basepoint of the interior of their intersection, or is degenerate if no such component exists. Taking interiors is necessary: a nested sequence of thin rectangles can have a line segment as its set-theoretic intersection, but no nondegenerate conformal domain as its kernel.

Translations by convergent vectors preserve this notion of convergence. For clarity, if $b_n \to b$, then (U_n, a_n) and $(U_n - b_n, a_n - b_n)$ have corresponding kernels U and $U - b$. To check the assertion, surround a compact subset of a proposed limit domain by a slightly larger compact neighborhood still in that domain. Eventually $|b_n - b|$ is smaller than the available margin, so inclusion for the larger neighborhood gives inclusion for the translated compact set. Apply the same argument with the inverse translations and to every subsequence. In particular, moving basepoints converging to a finite limit may be reduced to a fixed basepoint by translation.

11.4 Conformal radius as a normalization parameter

For a conformal disk U and $a \in U$, the Riemann mapping theorem (proved below as Theorem 12.4) and a rotation provide a unique conformal bijection

$$f_{U,a} : \mathbb{D} \longrightarrow U, \qquad f_{U,a}(0) = a, \qquad f'_{U,a}(0) > 0.$$

Uniqueness follows from Schwarz's lemma applied to the transition map and its inverse: a disk automorphism fixing zero is a rotation, and a positive derivative removes that rotation. Define

$$c(U, a) = f'_{U,a}(0), \qquad r(U, a) = \operatorname{dist}(a, \mathbb{C} \setminus U).$$

These are the conformal radius and the inner Euclidean radius, respectively. Thus $c(U, a) = r_U(a)$ in the notation of Chapter 3; the two-argument symbol $r(U, a)$ here denotes the Euclidean inner radius. They are different measurements of the same local scale.

Proposition 11.8 (Radius comparison and monotonicity)**.** *For every pointed conformal disk,*

$$r(U, a) \le c(U, a) \le 4r(U, a). \tag{11.3}$$

If $U \subset V$ are conformal disks containing a, then $c(U, a) \le c(V, a)$, with equality only if $U = V$.

Proof. For $0 < s < r(U, a)$, the function $\zeta \mapsto f_{U,a}^{-1}(a + s\zeta)$ maps $\mathbb{D}$ into itself and fixes zero. Schwarz's derivative inequality gives $s/c(U, a) \le 1$. Let $s \uparrow r(U, a)$. Conversely, the Koebe quarter theorem (Theorem 3.4) puts $B(a, c(U, a)/4)$ inside U, proving the other inequality. For monotonicity apply Schwarz's lemma to $f_{V,a}^{-1} \circ f_{U,a}$. Equality in the derivative inequality makes this transition map a rotation, hence onto $\mathbb{D}$; its images then give $U = V$. □

The growth estimate from Theorem 3.6 takes the form

$$|f_{U,a}(z) - a| \le c(U, a) \frac{|z|}{(1 - |z|)^2}. \tag{11.4}$$

Thus bounds on the basepoint and the conformal radius give locally uniform bounds on the normalized maps even when the image domains are unbounded. This is where a geometric theorem about univalent functions enters the compactness argument. Bounded derivatives at one point do not give such a conclusion for arbitrary holomorphic functions.

11.5 Worked investigations in analytic compactness

Investigation 11.9 (Why one derivative is not enough)**.** Construct holomorphic maps on $\mathbb{D}$ with $f_n(0) = 0$ and $f_n'(0) = 1$ that have no locally uniformly convergent subsequence. Identify the hypothesis that prevents the same example in the normalized univalent family.

Solution. Take $f_n(z) = z + nz^2$. Their values at $z = 1/2$ tend to infinity, so no subsequence can converge locally uniformly to a finite function. For $n \geq 1$, their derivative vanishes at $-1/(2n) \in \mathbb{D}$, so they are not univalent. In the univalent family the growth estimate bounds all values on each smaller disk by a constant determined by the one derivative. The distinction is global injectivity, not greater smoothness.

Investigation 11.10 (Conformal radius of an off-center disk)**.** Let $U = B(b, R)$ and $a \in U$. Find $c(U, a)$ and compare it with $r(U, a)$.

Solution. Put $\alpha = (a - b)/R$. The map $z \mapsto b + R(z + \alpha)/(1 + \overline{\alpha}z)$ takes 0 to a and has positive derivative $R(1 - |\alpha|^2)$. Therefore

$$c(U, a) = R - \frac{|a - b|^2}{R}, \qquad r(U, a) = R - |a - b|,$$

and $c/r = 1 + |a - b|/R$. This ratio ranges from 1 at the center toward 2 near the boundary. The universal factor 4 in (11.3) is not approached by round disks; slit geometry is more extreme.

Investigation 11.11 (Pointwise information plus normality)**.** Suppose a locally bounded sequence of holomorphic functions on U converges at every point of a subset $E \subset U$ having an accumulation point in U. Prove that the sequence converges locally uniformly.

Solution. Normality supplies a locally uniform cluster function f. Any other cluster function has the same values on E, because those scalar sequences already converge. The identity theorem makes the two cluster functions equal. If the original sequence failed to converge to f on some compact set, a subsequence would stay a fixed positive distance away there. Normality would give that subsequence a further subsequence converging locally uniformly to f, a contradiction. Without local boundedness the cluster argument has no starting point.

Investigation 11.12 (A degenerate pointed limit)**.** Let $U_n = B(0, 1)$ and $a_n = 1 - 1/n$, with $n \geq 2$. Determine the pointed kernel and the locally uniform limit of the positively normalized maps.

Solution. The limiting basepoint is 1, which is not in the eventual interior $\mathbb{D}$. Hence the pointed kernel is the degenerate symbol $\{1\}$. The normalized maps are

$$f_n(z) = \frac{z + a_n}{1 + a_n z}, \qquad f_n'(0) = 1 - a_n^2 > 0.$$

They converge locally uniformly on $\mathbb{D}$ to the constant 1, because the limiting denominator $1 + z$ is nonzero there. The unpointed domains have not changed at all. The degeneration comes entirely from the observer approaching their boundary.

Chapter 12

The planar uniformization theorem with its existence proof

The disk metric on a simply connected plane domain presupposes that a conformal coordinate exists. Uniqueness follows quickly from Schwarz's lemma; existence requires a different argument. This chapter supplies that argument using the normal-family tools just established. It proves the classical planar Riemann mapping theorem, discussed also in [3]. General uniformization of arbitrary Riemann surfaces is a separate global theorem and is not a consequence of this planar proof.

The construction has three stages. First, simple connectivity turns local primitives into global logarithms. Second, a square root and an omitted disk produce at least one bounded injective holomorphic function. Third, maximizing a derivative produces a map whose image cannot omit any point of the disk. We give the topological and limiting steps explicitly so that none of these stages assumes the desired conformal map.

12.1 From local primitives to a global logarithm

A domain is simply connected here if every continuous loop in it is homotopic to a constant loop while remaining in the domain. Plane domains are locally path connected, and a connected open set is polygonally connected: the set reachable from a fixed point by polygonal paths is both open and relatively closed, since a small disk about any point can be joined by line segments. This supplies the piecewise smooth paths used in contour integration. The homotopies themselves need only be continuous.

Lemma 12.1 (Homotopy invariance from local primitives)**.** *Let g be holomorphic on a domain U. Its integral along a piecewise C^1 path is unchanged by a fixed-endpoint homotopy within U, provided the other endpoint path is also piecewise C^1. Consequently, if U is simply connected, g has a holomorphic primitive on U.*

Proof. Every sufficiently small disk contained in U has a primitive of g by Proposition 1.3. A continuous path γ on a compact interval can be divided into finitely many subintervals whose images lie in such disks. On a subinterval $[t_j, t_{j+1}]$, define its increment to be $G_j(\gamma(t_{j+1})) - G_j(\gamma(t_j))$, where G_j is a local primitive on the selected disk. Sum these increments. This definition makes sense even when the path has no derivative.

The sum does not depend on the chosen division or disks. Refining a division simply splits one difference into a telescoping sum. To compare two choices, refine both divisions simultaneously. On each small piece the image lies in the intersection of the two selected disks. That intersection is convex and therefore connected. The two local primitives have the same derivative there, so

their difference is constant and they assign the same increment. If the path is piecewise C^1, the chain rule and fundamental theorem of calculus identify the increment with the usual contour integral. Reversing a path reverses its increment; concatenating paths adds their increments.

Now let $H : [0,1]^2 \to U$ be the homotopy. Its image is compact. Cover that image by finitely many disks with local primitives. The Lebesgue-number property for a finite open cover of a compact set, together with uniform continuity of H, gives a square grid fine enough that the image of each closed grid cell lies in one of those disks. The increment around that cell is zero: in its local primitive the four endpoint differences telescope. Add over all cells. Each interior edge cancels with the oppositely oriented copy in its adjacent cell. Only the outer boundary remains. The two vertical boundary paths are constant because endpoints are fixed, so their increments vanish. The two horizontal paths therefore have equal increments. This proves homotopy invariance without differentiating the homotopy or integrating an arbitrary continuous curve in the classical sense.

For a homotopy through loops, the basepoint may move. In that case the two vertical boundary paths coincide and are traversed in opposite directions, so their increments still cancel. A freely contractible loop therefore has the same zero increment as a constant loop.

If U is simply connected, every loop has zero increment. Fix $p \in U$ and define $G(z)$ by integrating from p to z along any polygonal path. Two such paths differ by a loop, so the value is independent of the path. Near any point, append a short line segment in a small disk. The resulting difference of G is the difference of a local primitive. Thus G is holomorphic there with derivative g. Since the point was arbitrary, $G' = g$ throughout U. □

Corollary 12.2 (Logarithms and roots on simply connected domains)**.** *If U is simply connected and h is holomorphic and nowhere zero on U, then there is a holomorphic L with $e^L = h$. For every positive integer m, $\exp(L/m)$ is a nowhere-zero holomorphic mth root. Specifying one permitted value at one point fixes the logarithm or the root uniquely.*

Proof. Apply the lemma to h'/h to obtain a primitive G. The derivative of he^{-G} is zero, so connectedness makes it a nonzero constant c. Choose $\ell \in \mathbb{C}$ with $e^\ell = c$ and put $L = G + \ell$. Then $e^L = h$, and the formula for an mth root follows by exponentiation. Two logarithms differ by a continuous function taking values in $2\pi i\mathbb{Z}$, hence by one constant from that set. Two roots have a continuous quotient taking values among the finitely many mth roots of unity, hence have constant quotient. Their prescribed point values remove precisely this freedom. □

This extends the convex-domain logarithm of Proposition 1.10. The local primitive theorem is analytic; homotopy invariance is what permits continuation through a nonconvex domain. The hypothesis is substantive: $h(z) = z$ on a punctured disk has no logarithm, as its logarithmic derivative has nonzero integral around a circle about zero.

12.2 Producing a bounded injective map

Lemma 12.3 (A nonempty normalized map family)**.** *Let $U \subsetneq \mathbb{C}$ be a simply connected domain and $p \in U$. There exists an injective holomorphic map $F : U \to \mathbb{D}$ with $F(p) = 0$ and $F'(p) > 0$ real.*

Proof. Choose $a \in \mathbb{C} \setminus U$. The function $z - a$ is nowhere zero on U, so Corollary 12.2 gives a holomorphic function h satisfying $h(z)^2 = z - a$. It is injective: equality of two values implies equality of their squares and hence of their arguments. It is nonzero, and differentiating gives $h' = 1/(2h)$, which is also nonzero.

The open sets $h(U)$ and $-h(U)$ are disjoint. Indeed if $h(z_1) = -h(z_2)$, squaring gives $z_1 = z_2$, whence $h(z_1) = 0$, an impossibility. Put $b = h(p)$. Openness gives an $\varepsilon > 0$ such that

$B(b,\varepsilon) \subset h(U)$; choose it smaller than $|b|$. Reflection through zero now gives $B(-b,\varepsilon) \subset -h(U)$, disjoint from $h(U)$. Thus $|h(z)+b| \geq \varepsilon$ for every $z \in U$.

The reciprocal transformation

$$G(z) = \frac{\varepsilon}{2(h(z)+b)}$$

is holomorphic and injective, with $|G(z)| \leq 1/2$. In particular its image lies in $\mathbb{D}$ and its derivative does not vanish. Compose with $m_{G(p)}$ to send the basepoint to zero, then multiply by a constant of modulus one to make the derivative at p positive real. Both operations preserve the disk and injectivity. This produces F. □

The omitted object in this argument is a disk, not just a point. Inverting about an omitted point alone need not give a bounded map, because the image might approach that point arbitrarily closely. The reflected open disk supplies the positive lower bound on the denominator that is needed. No boundary regularity of U was used, and U may be unbounded.

12.3 The extremal derivative and surjectivity

Theorem 12.4 (Planar Riemann mapping theorem)**.** *For every simply connected proper plane domain U and every $p \in U$, there is a unique conformal bijection $F : U \to \mathbb{D}$ satisfying $F(p) = 0$ and $F'(p) > 0$ real. Its inverse $\varphi : \mathbb{D} \to U$ satisfies $\varphi(0) = p$ and $\varphi'(0) > 0$ real.*

Proof. Let $\mathcal{A}$ be the injective holomorphic maps $f : U \to \mathbb{D}$ with $f(p) = 0$ and $f'(p) > 0$ real. It is nonempty by Lemma 12.3. Choose $\rho > 0$ with $\overline{B(p,\rho)} \subset U$. Cauchy's estimate gives $f'(p) \leq 1/\rho$ for every $f \in \mathcal{A}$. Hence

$$A = \sup\{f'(p) : f \in \mathcal{A}\}$$

is finite and strictly positive. For each n choose $f_n \in \mathcal{A}$ with $f_n'(p) > A - 1/n$. The family is bounded by one on every compact subset of U, so Theorem 11.2 gives a locally uniformly convergent subsequence with holomorphic limit F.

Derivative convergence, Lemma 11.1, gives $F'(p) = A > 0$. Also $F(p) = 0$ and $|F| \leq 1$. The positive derivative excludes constancy, so the maximum principle gives $|F(z)| < 1$ at every point of U. Proposition 11.4 makes F injective. Thus the limit belongs to $\mathcal{A}$ and attains the supremum; the normalization did not disappear in the limit.

It remains to prove $F(U) = \mathbb{D}$. Suppose instead that $b \in \mathbb{D} \setminus F(U)$. Since $F(p) = 0$, $b \neq 0$. Set $g = m_b \circ F$. This is an injective map into $\mathbb{D}$ that never vanishes. Since U is simply connected, choose a holomorphic square root h with $h^2 = g$. It is injective, because equality of its values implies equality of g; also $|h| < 1$. Put $c = h(p)$. Then $c^2 = -b$, so $|c| = \sqrt{|b|} \in (0,1)$. The map $H = m_c \circ h$ is injective, maps U to $\mathbb{D}$, and vanishes at p. Its derivative has modulus

$$\begin{aligned} |H'(p)| &= \frac{|h'(p)|}{1-|c|^2} = \frac{|g'(p)|}{2|c|(1-|c|^2)} \\ &= \frac{1-|b|^2}{2\sqrt{|b|}(1-|b|)} F'(p) = \frac{1+|b|}{2\sqrt{|b|}} A. \end{aligned}$$

The first identity uses the derivative of m_c at c, the second uses $2hh' = g'$, and the third uses $m_b'(0) = 1-|b|^2$. The last factor exceeds one, since $1+|b|-2\sqrt{|b|} = (1-\sqrt{|b|})^2 > 0$. Multiplying H by a phase makes its derivative positive without changing its modulus. The resulting map lies in $\mathcal{A}$ and has derivative greater than A, a contradiction. Thus F is onto.

An injective holomorphic map has nonzero derivative, so its local holomorphic inverses agree on overlaps and define a holomorphic global inverse. This proves existence of φ and the derivative normalization $\varphi'(0) = 1/A$.

Finally, if F_1, F_2 are two normalized bijections, then $F_2 \circ F_1^{-1}$ is a disk automorphism fixing zero. Schwarz's lemma applied to it and to its inverse makes it a rotation. Its derivative at zero is the positive real number $F_2'(p)/F_1'(p)$; a positive real number of modulus one is one. The rotation is the identity, so $F_1 = F_2$. □

Several possible failures have been excluded separately. The Cauchy bound prevents the optimizing derivatives from diverging. The initially constructed map prevents the supremum from being zero. The positive limit derivative prevents collapse to a constant. Hurwitz-type univalent limit control prevents self-intersections. The square-root improvement prevents omission of a target point. None of those conclusions follows from boundedness of the map family alone.

The improvement factor has a useful quantitative interpretation. If $r = |b|$, then

$$\frac{1+r}{2\sqrt{r}} - 1 = \frac{(1-\sqrt{r})^2}{2\sqrt{r}}.$$

It is strictly positive for every omitted interior point, but tends to zero quadratically as that point approaches the unit circle. The argument therefore proves surjectivity of an exact extremizer. It does not give a uniform iteration count for a numerical procedure that repeatedly chooses arbitrary omitted points. Computing such a point and controlling the limiting map from finite data would require additional information.

12.4 Conformal radius as an extremal quantity

Proposition 12.5 (Extremal derivative and conformal radius)**.** *Let F be the map of Theorem 12.4, and define $r_U(p) = 1/F'(p) > 0$. For every holomorphic $g : U \to \mathbb{D}$ with $g(p) = 0$,*

$$|g'(p)| \le \frac{1}{r_U(p)}.$$

Equality holds precisely when $g = e^{i\theta}F$ for a real θ. At any $q \in U$ the hyperbolic density and conformal radius satisfy

$$r_U(q) = \frac{1-|F(q)|^2}{|F'(q)|}, \qquad \lambda_U(q) = \frac{2}{r_U(q)}.$$

Proof. The map $g \circ F^{-1} : \mathbb{D} \to \mathbb{D}$ fixes zero. Schwarz's derivative inequality gives $|g'(p)|/F'(p) \le 1$, and its equality case gives the claimed rotation characterization. Injectivity of g was not assumed; it follows in the equality case.

To normalize at q, compose F with $m_{F(q)}$ and then with a rotation making the new derivative positive real. The derivative modulus is $|F'(q)|/(1-|F(q)|^2)$. The uniqueness theorem identifies its reciprocal as $r_U(q)$. Pulling back the disk density $2/(1-|w|^2)$ gives the displayed formula for λ_U. Changing the original normalized map by a disk automorphism leaves this density unchanged, by the metric identity in Chapter 2. □

For an inclusion $U \subset V$ of simply connected proper plane domains, restrict a normalized disk-valued map on V to U and apply the extremal inequality. It gives $r_U(p) \le r_V(p)$ and hence $\lambda_U(p) \ge \lambda_V(p)$. If equality holds at one point, the restricted map attains equality and maps U onto the disk. Since the map on V is injective and already maps V onto that disk, this forces $U = V$. Thus a proper simply connected inclusion has strict density monotonicity at every interior point. This argument gives the same order as Schwarz–Pick and explains it as an extremal problem.

The theorem concerns interior coordinates only. It supplies no continuous extension to an arbitrary boundary, no boundary smoothness, and no description of how a rough boundary is

parametrized. Those are separate questions. Likewise the proof for a plane domain does not construct the universal cover of a general surface. The later surface arguments retain their explicitly stated covering and global uniformization inputs.

12.5 Solved investigations in existence and normalization

Investigation 12.6 (A shifted disk)**.** For $U = B(c, R)$ and $p \in U$, write the normalized conformal map to $\mathbb{D}$ and calculate $r_U(p)$. Verify the density at a point approaching the boundary along a radius.

Solution. Set $a = (p - c)/R$. The map $F(z) = m_a((z - c)/R)$ sends p to zero and has derivative $1/(R(1 - |a|^2)) > 0$ there, so it is already normalized. Therefore $r_U(p) = R(1 - |a|^2) = (R^2 - |p - c|^2)/R$ and $\lambda_U(p) = 2R/(R^2 - |p - c|^2)$. If $d = R - |p - c|$, then $\lambda_U(p) = 2R/(d(2R - d))$ and $d\lambda_U(p) \to 1$ as $d \downarrow 0$. This exact disk limit has a sharper constant than the general two-sided boundary-distance bounds for arbitrary domains.

Investigation 12.7 (Why the whole plane is excluded)**.** Show that no nonconstant entire function maps $\mathbb{C}$ into $\mathbb{D}$, using only the estimates already proved. Identify the stage of the existence proof that fails when $U = \mathbb{C}$.

Solution. For any $p \in \mathbb{C}$ and any $R > 0$, Cauchy's estimate gives $|f'(p)| \le 1/R$. Letting $R \to \infty$ gives $f'(p) = 0$, and connectedness gives constancy. In Lemma 12.3 there is no omitted point a from which to start the square-root construction. Correspondingly, no positive normalized derivative is available, so an extremal-family proof cannot even begin.

Investigation 12.8 (Why a puncture changes the conclusion)**.** Prove that the punctured disk $\mathbb{D} \setminus \{0\}$ is not conformally equivalent to $\mathbb{D}$, without assuming the classification of multiply connected domains.

Solution. Suppose $F : \mathbb{D} \setminus \{0\} \to \mathbb{D}$ were a conformal bijection, with inverse $G : \mathbb{D} \to \mathbb{D} \setminus \{0\}$. Boundedness makes the singularity of F at zero removable. The extension is nonconstant and has modulus at most one, so the maximum principle places $F(0)$ strictly inside $\mathbb{D}$. On the punctured disk, $G(F(z)) = z$. Continuity now gives $G(F(0)) = 0$, contradicting the fact that G omits zero. The obstruction is not a bad normalization: no conformal bijection exists at all.

Investigation 12.9 (A derivative certificate for a disk coordinate)**.** Let $U \subsetneq \mathbb{C}$ be simply connected, and let $g : U \to \mathbb{D}$ be holomorphic with $g(p) = 0$. Suppose $|g'(p)| = 1/r_U(p)$. Must g be onto even if injectivity was not part of the hypothesis? Explain why replacing equality by a strict near-equality changes the conclusion.

Solution. Yes. Proposition 12.5 gives $g = e^{i\theta}F$, so g is a conformal bijection. For near-equality take $g = tF$ with $0 < t < 1$ tending to one. Then $|g'(p)| = t/r_U(p)$ approaches the extremal value, while the image is only the disk of radius t. An arbitrarily small derivative deficit does not establish exact surjectivity. A quantitative closeness claim would need its own norm and compact subset, as in the earlier rigidity estimates.

Chapter 13

Kernel convergence and maps on changing domains

We can now state the relation between geometry and normalized conformal coordinates. The form below includes collapse to a point and distinguishes it from expansion to the whole plane. It develops the kernel-convergence questions of Lyubich, Section 7.7, with the connected-component and normalization requirements made explicit. A related normalized treatment is given in Tao's complex-analysis notes[1].

13.1 The normalized kernel theorem

Theorem 13.1 (Carathéodory kernel convergence)**.** *Let U_n be conformal disks, $a_n \in U_n$, and $a_n \to a \in \mathbb{C}$. Let $f_n : \mathbb{D} \to U_n$ be their conformal bijections normalized by $f_n(0) = a_n$ and $f_n'(0) > 0$.*

1. *If (U_n, a_n) converges in the kernel sense to a nondegenerate proper domain (U, a), then f_n converges locally uniformly to the normalized conformal bijection $f : \mathbb{D} \to U$.*
2. *Conversely, if f_n converges locally uniformly to a nonconstant finite function f, then there is kernel convergence to $(f(\mathbb{D}), a)$, and f is its normalized conformal bijection.*
3. *There is kernel convergence to $\{a\}$ if and only if $f_n \to a$ locally uniformly. There is kernel convergence to $\mathbb{C}$ if and only if $f_n'(0) \to \infty$.*

In the nondegenerate case the inverse maps converge locally uniformly on U to f^{-1}, and so do all their derivatives on compact subsets.

Proof. Translations by a_n reduce the proof to $a_n = a = 0$. Write $c_n = f_n'(0) > 0$. We first prove the forward geometric consequences of convergence of the maps, since they will identify all possible cluster limits in the converse direction.

A nonconstant limit contains every compact part of its image. Suppose $f_n \to f$ locally uniformly and f is nonconstant. By Proposition 11.4, f is univalent. Its derivative at zero is the limit of positive real numbers and is nonzero, so it is positive. Set $V = f(\mathbb{D})$. If $L \Subset V$, choose $r < 1$ such that $f^{-1}(L) \subset B(0, r)$. The disjoint compact sets L and $f(\{|z| = r\})$ have a positive distance ε. For large n, $|f_n - f| < \varepsilon$ on that circle. For every $w \in L$, Rouché's theorem says that $f_n - w$ and $f - w$ have the same one zero inside it. Thus $L \subset U_n$ for all sufficiently large n, uniformly for $w \in L$. This also holds along every subsequence. Consequently V lies in each subsequential pointed kernel.

[1] T. Tao, *246C notes 3: Univalent functions, the Loewner equation, and the Bieberbach conjecture*, Theorem 12, author's notes, 2 May 2018.

No larger stable component is possible. Let W be a connected open neighborhood of 0 whose compact subsets are eventually in the domains of some subsequence U_{n_j}. The inverse maps $g_j = f_{n_j}^{-1}$ are defined on every such compact set for all sufficiently large j and satisfy $|g_j| < 1$. The diagonal normal-family argument on an exhaustion of W gives a further subsequence with $g_j \to g$ locally uniformly. Here $g(0) = 0$ and $|g| \leq 1$; the maximum principle implies $|g| < 1$ everywhere on W. For each $w \in W$, the moving-argument rule now applies inside $\mathbb{D}$ and gives

$$w = f_{n_j}(g_j(w)) \longrightarrow f(g(w)).$$

Thus $W \subset V$. This proves that every subsequential kernel is exactly V, which establishes the second assertion. It also proves that every locally uniform cluster limit of the inverses on V is f^{-1}. Normality and uniqueness of cluster limits give convergence of the entire inverse sequence, and Lemma 11.1 gives derivative convergence.

A constant limit corresponds to collapse. If $f_n \to 0$ and a subsequence had a nondegenerate pointed kernel, some disk $B(0, \varepsilon)$ would be contained in all sufficiently late domains of that subsequence. Apply Schwarz's lemma to $\zeta \mapsto f_n^{-1}(\varepsilon\zeta)$ to obtain $c_n \geq \varepsilon$. This contradicts $c_n \to 0$. Therefore every subsequential kernel is $\{0\}$.

Kernel convergence supplies upper and lower scales. Now assume kernel convergence to a nondegenerate proper domain U. A closed disk about 0 contained in U is eventually in U_n. The same inverse-map estimate gives $c_n \geq \varepsilon > 0$ eventually. The sequence c_n is also bounded above. Otherwise a subsequence would have $c_{n_j} \to \infty$. The Koebe quarter theorem would then give $B(0, c_{n_j}/4) \subset U_{n_j}$, so this subsequence would have kernel $\mathbb{C}$, contrary to the assumed kernel convergence to U.

The upper bound and (11.4) make (f_n) locally bounded. Every subsequence therefore has a further locally uniformly convergent subsequence. Its limit f has $f'(0) \geq \varepsilon$, hence is nonconstant. The direction already proved identifies its image as U. The positive derivative normalization makes this conformal map unique. All cluster limits agree, so the full sequence converges. This proves the first assertion and also proves, rather than assumes, that the nondegenerate proper kernel in this setting is simply connected.

The two limiting scales. If the kernels converge to $\{0\}$, unbounded c_n would again give a subsequence with kernel $\mathbb{C}$. Hence c_n is bounded above and the maps form a normal family. A nonconstant cluster map would give a nondegenerate subsequential kernel, which is impossible. Every cluster map is therefore the constant 0, and the entire sequence converges to 0.

Finally, if the kernel convergence is to $\mathbb{C}$, every closed disk $\overline{B(0, R)}$ is eventually in U_n, giving $c_n \geq R$. As R is arbitrary, $c_n \to \infty$. Conversely, that derivative divergence and the quarter theorem put each compact subset of $\mathbb{C}$ in every sufficiently late U_n, along every subsequence as well. This is kernel convergence to $\mathbb{C}$. It admits no finite locally uniform limit of the normalized maps, since their derivatives diverge. □

There are three distinct outcomes: a finite nonconstant conformal map, a constant map at the basepoint, and an unbounded conformal scale. The theorem does not assert that every unnormalized sequence of maps has one of these limits without taking a subsequence. Alternating scales or phases can still prevent convergence.

13.2 A quantitative inverse-map step

The inverse convergence in the kernel theorem used compactness and uniqueness. A direct estimate is available on a fixed interior region. It reveals which separation constants are being used whenever a numerical or symbolic approximation to a conformal coordinate is turned into an approximation to its inverse.

Lemma 13.2 (Uniform inverse stability)**.** *Let f and f_n be univalent on $\mathbb{D}$, with $f_n \to f$ locally uniformly. Suppose $L \Subset f(\mathbb{D})$, and choose $r < 1$ with $f^{-1}(L) \subset B(0,r)$. Define*

$$\eta = \operatorname{dist}(L, f(\{|z| = r\})), \qquad E_n = \max_{|z|=r} |f_n(z) - f(z)|.$$

There is a constant $m_r > 0$, depending only on f and r, such that whenever $E_n < \eta$, the inverse f_n^{-1} is defined on L and

$$\sup_{w \in L} |f_n^{-1}(w) - f^{-1}(w)| \le E_n / m_r. \tag{13.1}$$

Proof. Both sets defining η are compact and disjoint, so $\eta > 0$. For every $w \in L$, the inequality $|f_n - f| < |f - w|$ holds on $|z| = r$. Rouché's theorem puts the unique preimage $z_n = f_n^{-1}(w)$ in $B(0,r)$, since $f - w$ has exactly one zero there. The same choice of n works for every $w \in L$; this is where the distance between two compact sets is needed.

On the product of two closed radius-r disks, extend the divided difference by

$$Q(z,\zeta) = \begin{cases} (f(z) - f(\zeta))/(z - \zeta), & z \ne \zeta, \\ f'(z), & z = \zeta. \end{cases}$$

It is jointly continuous. For example, the identity $Q(z,\zeta) = \int_0^1 f'(\zeta + t(z-\zeta))\,\mathrm{d}t$ verifies continuity also on the diagonal. Injectivity of f and nonvanishing of f' show that Q has no zero. Compactness now gives $m_r = \min |Q| > 0$. This is a global lower bound for divided differences on the chosen compact disk, not merely a lower bound for $|f'|$ at individual points.

Put $z = f^{-1}(w)$. The maximum principle bounds $|f_n - f|$ throughout $|z| \le r$ by E_n. Consequently

$$m_r |z_n - z| \le |f(z_n) - f(z)| = |f(z_n) - f_n(z_n)| \le E_n.$$

Taking the supremum over L proves (13.1). □

The proof explains why replacing m_r by $\min |f'|$ without further argument is unsafe. A derivative lower bound is a local statement; integrating complex derivatives along a segment can involve cancellation. The nonzero divided difference controls two points at once, and injectivity makes that control possible on a compact set. Neither constant is asserted to be uniform as $r \uparrow 1$.

Derivative estimates follow with a second margin. Choose a compact neighborhood L^+ of L in $f(\mathbb{D})$ and apply the lemma on L^+. If every point of L has a closed Euclidean disk of radius $s > 0$ inside L^+, Cauchy's formula gives

$$\sup_{w \in L} |(f_n^{-1})^{(j)}(w) - (f^{-1})^{(j)}(w)| \le \frac{j!}{s^j} \sup_{w \in L^+} |f_n^{-1}(w) - f^{-1}(w)|.$$

The neighborhood must be chosen before the estimates are applied; a compact set with empty interior does not itself provide the contours needed to control derivatives. These estimates are analytic implications of stated sup-norm bounds, rather than claims that such bounds have been measured for an unspecified approximation.

13.3 A topology and a compactness theorem

The kernel criterion can be turned into a topology without guessing a collection of geometric neighborhoods. Associate to each pointed conformal disk (U, a) its unique map $f_{U,a}$. On $\mathbb{D}$ take the compact sets $\{|z| \le 1 - 1/(m+1)\}$ in (11.1), and pull that metric back to pointed domains. This gives a Hausdorff metric topology. The kernel theorem says that its convergent sequences are precisely the nondegenerate kernel-convergent sequences. The derivative normalization is part of this correspondence: images alone cannot distinguish z from $e^{i\theta}z$.

Proposition 13.3 (Continuity of the two radii)**.** *If (U_n, a_n) converges in the kernel sense to a pointed conformal disk (U, a), then*

$$c(U_n, a_n) \to c(U, a), \qquad r(U_n, a_n) \to r(U, a).$$

Proof. The first statement follows from derivative convergence of the normalized maps. For the second, first translate the basepoints to zero and write $r_n = r(U_n, 0)$, $r = r(U, 0)$. Every closed disk of radius less than r is eventually in U_n, so $\liminf r_n \geq r$. If a subsequence had $r_{n_j} \geq r + \varepsilon$, its pointed kernel would contain the disk $B(0, r + \varepsilon)$: every compact subset of that disk is in every domain of the subsequence. Kernel convergence would put this disk in U, contradicting the definition of r. Thus $\limsup r_n \leq r$. Translations preserve both radii and the argument applies to moving basepoints as well. □

Theorem 13.4 (Compact families of pointed disks)**.** *Let $A \subset \mathbb{C}$ be compact and $0 < m \leq M < \infty$. The family of pointed conformal disks satisfying*

$$a \in A, \qquad m \leq c(U, a) \leq M$$

is compact in the topology just defined. The same conclusion holds for the family defined instead by $m \leq r(U, a) \leq M$.

Proof. For a sequence in the first family, select a subsequence with $a_n \to a$ in A. The growth bound and $c_n \leq M$ make the normalized maps locally bounded. A further subsequence converges to f, with $m \leq f'(0) \leq M$. The limit is univalent. Its image is proper: otherwise its inverse would be a bounded nonconstant entire function, contradicting Liouville's theorem. The kernel theorem identifies this image as the pointed limit. It has basepoint in A and conformal radius in $[m, M]$, so the family is sequentially compact. In a metric space sequential compactness is equivalent to compactness.

For the inner-radius family, (11.3) gives $m \leq c_n \leq 4M$, so the same extraction works. The inner-radius continuity proposition ensures that the limit still has inner radius in the closed interval $[m, M]$. □

The basepoint condition is necessary. The disks $B(n, 1)$ based at n have both radii equal to 1, but their basepoints have no finite convergent subsequence. Likewise, replacing a closed parameter interval by a strict inequality can destroy compactness. Radius bounds prevent collapse and expansion about the observer; they do not prevent the observer from escaping, and strict bounds need not survive a limit. This is the precise adjustment needed when using radius-based compactness statements such as Lyubich's Corollary 7.29.

There is also a useful metric consequence. If $g_n = f_n^{-1}$ and $g = f^{-1}$, the hyperbolic densities with curvature -1 are

$$\rho_{U_n}(w) = \frac{2|g_n'(w)|}{1 - |g_n(w)|^2}, \qquad \rho_U(w) = \frac{2|g'(w)|}{1 - |g(w)|^2}.$$

On a compact subset of U, the values of g stay a positive distance from the unit circle. Inverse and derivative convergence therefore give locally uniform convergence of these densities. In particular, $\rho_{U_n}(a_n) = 2/c(U_n, a_n) \to 2/c(U, a)$. This is a local density statement; comparing distances between points also involves the paths over which lengths are minimized.

13.4 Distances and geodesics under kernel convergence

Local convergence of a metric density does not, for an arbitrary sequence of Riemannian manifolds, immediately imply convergence of global distances: a minimizing path might leave every fixed

compact set. In conformal disks the common normalized coordinate resolves this problem. It puts the entire minimization in the fixed complete disk, where both distance and the minimizing segment are explicit.

Theorem 13.5 (Intrinsic distance convergence)**.** *Suppose pointed conformal disks (U_n, a_n) converge in the nondegenerate kernel sense to (U, a). If $K \Subset U$, then*

$$\sup_{v,w\in K} |d_{U_n}(v,w) - d_U(v,w)| \longrightarrow 0. \tag{13.2}$$

If $v_n \to v \in U$ and $w_n \to w \in U$, the constant-speed minimizing geodesics in U_n from v_n to w_n, parametrized on $[0,1]$, converge uniformly to the corresponding geodesic in U. The constant path is included when the limiting endpoints coincide.

Proof. Let f_n, f be the normalized maps and g_n, g their inverses. The kernel theorem puts K in U_n eventually and gives $g_n \to g$ uniformly on a compact neighborhood of K. Thus all the sets $g_n(K)$, for sufficiently large n, lie in one disk $|z| \le r < 1$. Conformal invariance gives

$$d_{U_n}(v,w) = 2\,\operatorname{artanh}\left|\frac{g_n(v) - g_n(w)}{1 - \overline{g_n(w)}\,g_n(v)}\right|.$$

The denominator is bounded below by $1 - r^2$. The displayed pseudohyperbolic distance between two points of the radius-r disk is at most $2r/(1+r^2) < 1$, by the disk distance formula or Schwarz's automorphism calculation. Hence the outer function artanh has bounded derivative on the entire range being used. Uniform convergence of g_n proves (13.2). This argument covers the diagonal $v = w$ without differentiating the absolute value there.

For the paths, put $\alpha_n = g_n(v_n)$, $\beta_n = g_n(w_n)$. The moving-argument rule gives $\alpha_n \to \alpha = g(v)$ and $\beta_n \to \beta = g(w)$. For any $\alpha, \beta \in \mathbb{D}$, set $b = T_\alpha(\beta)$ and define

$$P(\alpha,\beta,t) = T_\alpha^{-1}\left(\frac{b}{|b|}\tanh\bigl(t\,\operatorname{artanh}|b|\bigr)\right) \quad (0 \le t \le 1),$$

with the inner expression interpreted as zero when $b = 0$. The expression is continuous also at $b = 0$, since its modulus is at most $|b|$. It is the radial constant-speed geodesic after sending α to zero. It therefore gives the unique minimizing disk segment, with the constant segment at coincident endpoints.

The pairs (α_n, β_n) and their limit lie in a compact subset of $\mathbb{D} \times \mathbb{D}$. Continuity of P on its product with $[0,1]$ gives uniform convergence of the disk paths. Their union is contained in one compact subdisk: the continuous image of that compact parameter set is compact and lies in $\mathbb{D}$. Local uniform convergence of f_n now implies

$$f_n(P(\alpha_n,\beta_n,t)) \longrightarrow f(P(\alpha,\beta,t)) \quad \text{uniformly for } t \in [0,1].$$

These are exactly the required geodesics in the changing domains. The compact coordinate set also shows that their images eventually lie in a fixed compact neighborhood of the limiting path inside U. Thus no escaping shortcut has been hidden in the argument. □

Smooth metric convergence is available on every fixed compact interior set. The inverse derivatives of every order converge by Cauchy's formula. On such a set g' has positive minimum modulus and $1 - |g|^2$ has positive minimum. The density $2|g'|/(1-|g|^2)$ is a smooth function of these real and imaginary coordinates, with no zero denominator. Repeated real differentiation therefore proves convergence of all derivatives of ρ_{U_n} to those of ρ_U. The same holds for the metric tensor and its inverse. This claim concerns compact subsets of the limiting open domain; it asserts no convergence of curvature distributions or boundary data across a collapsing neck.

13.4.1 What an exhaustion does to the metric

For an increasing exhaustion $U_n \uparrow U$, domain monotonicity gives $\rho_{U_n} \downarrow \rho_U$ pointwise, and the preceding argument makes the convergence uniform on compact subsets. The distances also decrease to the limiting distance, by Theorem 13.5. An explicit bound is possible when a buffer is known. If a compact path γ in U_n has U-distance at least $L > 0$ from $U \setminus U_n$ at every point, then (10.12) gives

$$\operatorname{length}_U(\gamma) \le \operatorname{length}_{U_n}(\gamma) \le \coth(L/2) \operatorname{length}_U(\gamma).$$

In particular, if the U-geodesic from v to w has that buffer,

$$d_U(v, w) \le d_{U_n}(v, w) \le \coth(L/2) d_U(v, w).$$

The lower bound uses domain monotonicity for every path; the upper bound uses the specified limiting geodesic as a competitor. For fixed compact endpoints in an exhaustion, arbitrarily large buffers around their compact family of disk-coordinate geodesics eventually lie in U_n. Indeed, a closed bounded neighborhood of a compact set in the complete hyperbolic disk is compact, and its conformal image is compact in U. The exhaustion then contains that neighborhood. This supplies a geometric explanation for the vanishing distance error, although a rate in n still requires information about how quickly the domains exhaust U.

13.5 An alternative proof for increasing domains

An exhaustion permits a shorter argument with a different source of boundedness. Suppose $U_n \subset U_{n+1}$, all contain a, and their union is a proper conformal disk U. Let $F : \mathbb{D} \to U$ be normalized and put $q_n = F^{-1} \circ f_n$. These maps take $\mathbb{D}$ into itself, fix zero, and have positive derivative. Normality now follows from the elementary bound $|q_n| \le 1$, without applying a growth theorem to the f_n.

The inverse maps $h_n = q_n^{-1}$ are defined on the increasing domains $V_n = F^{-1}(U_n)$, whose union is $\mathbb{D}$. Select convergent subsequences of q_n on $\mathbb{D}$ and of h_n on compact subsets of $\mathbb{D}$. Their limits q, h fix zero and have values in $\mathbb{D}$ by the maximum principle. Taking limits in $q_n(h_n(w)) = w$ gives $q \circ h = \mathrm{id}$. Similarly, on a compact neighborhood of each $q(z)$, eventual inclusion in V_n and convergence of h_n justify $h \circ q = \mathrm{id}$. Hence q is a disk automorphism fixing zero with positive derivative; it is the identity. Every cluster limit is the identity, so $q_n \to \mathrm{id}$ and $f_n \to F$.

This proof uses a fixed outer uniformizing map. The general kernel theorem cannot assume such an outer domain in advance; its growth and quarter estimates replace that missing common geometry.

13.6 Holomorphic functions on varying domains

Let $(U_n, a_n) \to (U, a)$ in the nondegenerate kernel sense. A sequence of holomorphic functions $h_n : U_n \to \mathbb{C}$ converges locally uniformly on U if, for every $K \Subset U$, the functions are defined on K for all sufficiently large n and converge uniformly there. The omission of finitely many early functions depends on K and is harmless.

Proposition 13.6 (Varying-domain normality)**.** *Suppose that for every $K \Subset U$ the quantities $\sup_K |h_n|$ are bounded for all sufficiently large n. Then a subsequence converges locally uniformly on U to a holomorphic function. All derivatives converge locally uniformly along the same subsequence.*

Proof. Choose an exhaustion by compact sets with successively larger compact neighborhoods in U. Each neighborhood lies in all sufficiently late U_n. Cauchy estimates give equicontinuity on the smaller sets, and the diagonal construction in Theorem 11.2 applies after discarding the finitely many undefined terms at each stage. Cauchy's formula on disks compactly contained in U proves holomorphy and derivative convergence exactly as before. □

When the domains are conformal disks, this convergence agrees with convergence in normalized coordinates. If $f_n : \mathbb{D} \to U_n$ and $f : \mathbb{D} \to U$ are normalized, then

$$h_n \to h \text{ locally on } U \quad \Longleftrightarrow \quad h_n \circ f_n \to h \circ f \text{ locally on } \mathbb{D}. \tag{13.3}$$

For the forward implication, a compact image $f(K)$ has a compact neighborhood in U, and $f_n(K)$ lies in that neighborhood eventually. For the reverse implication use $g_n = f_n^{-1} \to f^{-1} = g$ and $h_n(w) = (h_n \circ f_n)(g_n(w))$. The compact neighborhoods needed for both compositions are furnished by the kernel theorem. Merely writing down a formal composition would not justify either limit.

A bounded target gives local bounds but does not automatically preserve the open target. For example, the disk-valued constants $h_n = 1 - 1/n$ converge to the boundary value 1. If instead $h_n(a_n) \to b \in \mathbb{D}$, then a disk-valued sequence has limits in $\mathbb{D}$: any holomorphic limit has modulus at most one and takes the interior value b at a, so the maximum principle excludes boundary values everywhere. This simple normalization often makes the difference between a limit map into the desired space and a boundary constant.

13.7 Worked investigations in changing geometry

Investigation 13.7 (A complete compact annulus model)**.** For $m > 0$ and $t \in \mathbb{R}$, set

$$A(m,t) = \{e^{t-\pi m} < |z| < e^{t+\pi m}\}, \qquad a = e^t.$$

Show that the pointed family with $\mu \le m \le 1/\mu$ and $|t| \le L$ is compact, where $0 < \mu \le 1$ and $L < \infty$. Explain the roles of the two parameters.

Solution. With the convention $\operatorname{mod} A = (2\pi)^{-1} \log(R/r)$, the modulus is m. The parameter t fixes the Euclidean scale and puts the basepoint on the logarithmic middle circle. Any parameter sequence has a convergent subsequence $(m_n, t_n) \to (m, t)$ in the closed rectangle. Its inner and outer radii converge to distinct positive finite radii. A compact set in the limiting annulus has a positive radial margin from both circles, so it lies in every sufficiently late annulus. A point outside the closed limiting annulus has a neighborhood eventually outside; a point on a limiting circle has no neighborhood eventually inside. Thus the eventual interior is exactly the limiting annulus, for every subsequence as well. The basepoints converge to e^t. This proves sequential compactness in the parameter topology and the stated kernel convergence. A rotating basepoint $e^{t+i\theta}$ can also be allowed, since the extra parameter lies in the compact circle.

This proves a compactness result for a specified family of round, normalized annuli. Modulus alone would allow $t \to \infty$ or $t \to -\infty$. A strict basepoint range would also fail to be closed: for example $t_n \uparrow L$ would reach an excluded endpoint.

Investigation 13.8 (Degree does not normalize a target)**.** Fix a round annulus $A = \{r < |z| < R\}$ and an integer $d \ge 1$. Consider $F_n(z) = nz^d$ as a map onto its image annulus. Determine its degree and explain why it has no finite locally uniform subsequential limit. Give a corrected compact parameter family of this form.

Solution. The target is $\{nr^d < |w| < nR^d\}$. Each target point has exactly d distinct roots in A, and $F_n'(z) = ndz^{d-1} \neq 0$ there. Thus it is an unbranched degree-d covering. Nevertheless

$|F_n(z)| \to \infty$ at every point of A. Fixed source geometry and fixed degree do not control the Euclidean scale of the target.

For maps $F(z) = cz^d$, restrict c to a compact annulus $0 < c_- \le |c| \le c_+ < \infty$, and, if the source varies, restrict (m, t) as in Exercise 13.7. A subsequence has convergent coefficients and source parameters. The resulting map is still cz^d with $c \neq 0$, has the same degree, and converges uniformly on every compact subset of the limiting source. This is a corrected statement for this explicit covering family; it does not assert a compactness theorem for all coverings without additional normalization.

Investigation 13.9 (Why boundary Hausdorff convergence needs an observer)**.** Give pointed disks whose boundaries are identical but which do not converge to a nondegenerate pointed disk. Then give a simple situation in which boundary and kernel convergence do agree.

Solution. The disks and basepoints in the last exercise of Chapter 11 have identical boundaries, yet their pointed limit is degenerate. Boundary convergence by itself has not kept the basepoint in the interior. In contrast, if $U_n = B(b_n, R_n)$, $b_n \to b$, $R_n \to R > 0$, and $a_n \to a$ with $|a - b| < R$, then compact radial margins show kernel convergence to $(B(b, R), a)$. Their boundaries also converge in Hausdorff distance, since corresponding points $b_n + R_n e^{i\theta}$ and $b + Re^{i\theta}$ are uniformly within $|b_n - b| + |R_n - R|$. The conclusion uses the limiting side of the circle and the interior basepoint, not an unqualified boundary-set criterion.

Investigation 13.10 (Continuity of normalized inverse coordinates)**.** Let $(U_n, a_n) \to (U, a)$ be nondegenerate kernel convergence and let $w_n \to w \in U$, with $w_n \in U_n$. Prove $f_n^{-1}(w_n) \to f^{-1}(w)$ and identify the additional fact needed if w were instead a boundary point of U.

Solution. A closed neighborhood of w is compactly contained in U, contains all sufficiently late w_n, and lies eventually in U_n. The inverse maps converge uniformly there, so the moving-argument rule gives the claim. At a boundary point no such compact neighborhood is available. One would need a separate boundary-extension and boundary-convergence theorem with suitable hypotheses. The kernel theorem alone supplies neither continuity of f^{-1} at that point nor uniform control up to the varying boundaries.

Chapter 14

Holomorphic iteration and boundary attraction

Schwarz–Pick gives a Lipschitz constant of one, which alone does not imply an interior fixed point. Iteration may approach an interior point or the boundary, or rotate indefinitely. We distinguish these possibilities and prove Denjoy–Wolff, including boundary attraction.

We use the disk automorphisms and distance

$$T_p(z) = \frac{z-p}{1-\overline{p}z}, \qquad \delta(z,w) = |T_w(z)|, \qquad d_{\mathbb{D}}(z,w) = 2\operatorname{artanh}\delta(z,w).$$

The factor two gives curvature -1; changing this factor would not affect any nonexpansion argument below. We write $f^0 = \mathrm{id}$ and $f^{n+1} = f \circ f^n$. An elliptic automorphism means a disk automorphism with an interior fixed point; it includes the identity. Conjugating such an automorphism by T_p makes it a rotation about zero. For classical context and further boundary theory, see Abate's author-posted monograph[1].

14.1 An interior fixed point

Suppose $f : \mathbb{D} \to \mathbb{D}$ is holomorphic and $f(p) = p$. Set $g = T_p \circ f \circ T_p^{-1}$. Schwarz's lemma gives $|g(z)| \le |z|$ and $|g'(0)| \le 1$. If equality holds at a nonzero point or in the derivative inequality, g is a rotation. Otherwise the inequality is strict away from zero and can be made uniformly strict on any fixed smaller disk.

Theorem 14.1 (Attraction to an interior fixed point)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic, with $f(p) = p$. If f is not an elliptic automorphism, then p is its unique fixed point and $f^n \to p$ locally uniformly on $\mathbb{D}$. More precisely, for every $0 < r < 1$ there is $0 \le q_r < 1$ such that*

$$|T_p(f^n(z))| \le q_r^n |T_p(z)| \quad \text{whenever } |T_p(z)| \le r. \tag{14.1}$$

If f is an elliptic automorphism, then $f^n = T_p^{-1} \circ (\lambda^n T_p)$ for some $|\lambda| = 1$.

Proof. Remove the singularity of $q(z) = g(z)/z$ at zero by setting $q(0) = g'(0)$. Schwarz's lemma gives $|q| \le 1$. Unless g is a rotation, the maximum principle gives $|q(z)| < 1$ throughout $\mathbb{D}$. Hence $q_r = \max_{|z|\le r} |q(z)| < 1$. Since $|g(z)| \le |z|$, an orbit starting in $\overline{B(0,r)}$ stays there; induction gives $|g^n(z)| \le q_r^n |z|$. Conjugation gives (14.1) and local uniform convergence. A second fixed point would give a nonzero fixed point of g, contradicting strict Schwarz contraction. In the rotation case the displayed formula for the iterates follows directly from conjugation. □

[1] M. Abate, *Iteration Theory of Holomorphic Maps on Taut Manifolds*, Mediterranean Press, 1989, Chapters 1.2–1.3. The boundary Schwarz chapter and the author's chapter index provide the primary reference. Our statements below specify the disk normalization used in this volume.

There is a useful second way to see attraction. The numbers $|g^n(z)|$ decrease to some $\ell \geq 0$. If $\ell > 0$, an orbit subsequence converges to a nonzero point w in a compact subdisk. Continuity gives both $|w| = \ell$ and $|g(w)| = \ell$, contradicting $|g(w)| < |w|$. This argument proves convergence of each orbit without first estimating q_r. The maximum-on-a-subdisk argument supplies the additional uniformity and an explicit geometric form of the error. Its constant depends on the compact subdisk and on the map; it is not a universal contraction factor on the whole disk.

For a nonidentity rotation, $\lambda^n z$ cannot converge when $z \neq 0$. Indeed, a hypothetical limit would have modulus $|z| > 0$ and would satisfy $L = \lambda L$, forcing $\lambda = 1$. Finite-order rotations give periodic orbits, and the identity leaves every point fixed. These are exactly the interior exceptions that must be excluded from a theorem asserting convergence to a single constant.

14.2 Linearizing an attracting fixed point

For a nonzero attracting multiplier, the classical Koenigs construction gives an exactly linear local coordinate. Its extension to the disk need not be injective. Abate's survey gives the classical local form[2].

Theorem 14.2 (Normalized Koenigs coordinate)**.** *Let $g : \mathbb{D} \to \mathbb{D}$ be holomorphic, with $g(0) = 0$ and $\lambda = g'(0)$ satisfying $0 < |\lambda| < 1$. Then*

$$h(z) = \lim_{n\to\infty} \frac{g^n(z)}{\lambda^n} \tag{14.2}$$

exists locally uniformly on $\mathbb{D}$. It is the unique holomorphic function on $\mathbb{D}$ satisfying

$$h(0) = 0, \qquad h'(0) = 1, \qquad h(g(z)) = \lambda h(z).$$

It is univalent on some neighborhood of zero, and it is univalent on all of $\mathbb{D}$ if and only if g is univalent on $\mathbb{D}$.

Proof. Choose a local contraction strong enough for the error series. Since $0 < |\lambda| < 1$, the interval $(|\lambda|, \sqrt{|\lambda|})$ is nonempty. Choose q in that interval. Taylor expansion gives a radius $r > 0$ and $M < \infty$ such that, for $|z| \leq r$,

$$|g(z)| \leq q|z|, \qquad |g(z) - \lambda z| \leq M|z|^2.$$

Shrink r to keep the closed disk inside $\mathbb{D}$. The first inequality makes that disk forward invariant and gives $|g^n(z)| \leq q^n|z|$. The choice $q^2 < |\lambda|$ is the part that will make the normalized second-order error summable.

Sum the successive differences. Put $h_n = g^n/\lambda^n$, with $h_0(z) = z$. For $|z| \leq r$,

$$\begin{aligned} |h_{n+1}(z) - h_n(z)| &= \frac{|g(g^n(z)) - \lambda g^n(z)|}{|\lambda|^{n+1}} \\ &\leq \frac{M}{|\lambda|} \left(\frac{q^2}{|\lambda|}\right)^n |z|^2. \end{aligned}$$

Writing $\theta = q^2/|\lambda| < 1$, the geometric series proves uniform convergence on this closed disk and gives the tail estimate

$$|h(z) - h_n(z)| \leq \frac{M|z|^2}{|\lambda|(1-\theta)} \theta^n. \tag{14.3}$$

[2]M. Abate, *Discrete holomorphic local dynamical systems*, Section 2, author's survey, arXiv:0903.3289. The argument here gives the explicit telescoping bound and then extends the coordinate using the already proved disk attraction.

Analytic-limit and derivative convergence give $h(0) = 0$ and $h'(0) = 1$, because $h_n'(0) = \lambda^n/\lambda^n = 1$. In particular, the limit has not collapsed to zero. The identity $h_n(g(z)) = \lambda h_{n+1}(z)$ passes to the limit on the invariant disk, proving the functional equation there. The holomorphic inverse function theorem gives injectivity on a smaller neighborhood of zero.

Extend to every compact set of the disk. By Theorem 14.1, for each $K \Subset \mathbb{D}$ there is an N such that $g^N(K) \subset B(0, r)$; the same is true for a compact neighborhood of K. For $n \geq N$ write

$$\frac{g^n(z)}{\lambda^n} = \lambda^{-N} h_{n-N}(g^N(z)).$$

The right side converges uniformly on K. Thus the original sequence h_n converges locally uniformly on the entire disk and defines a single holomorphic function. This construction also proves that the expression $\lambda^{-N} h(g^N(z))$ is independent of which sufficiently large N is chosen. The functional equation holds everywhere by local uniform convergence, or by the identity theorem from the neighborhood of zero.

Check uniqueness rather than assuming it from normalization. If k is another holomorphic solution with the same normalization, then $k(w) = w + O(w^2)$ near zero. For $|z| \leq r$, the functional equation gives

$$k(z) = \lambda^{-n} k(g^n(z)) = h_n(z) + O(|\lambda|^{-n} |g^n(z)|^2).$$

The error is bounded by a constant times $\theta^n |z|^2$ and tends to zero. Hence $k = h$ near zero, and the identity theorem extends equality to $\mathbb{D}$. The normalization of the derivative is essential: without it, nonzero scalar multiples are also solutions.

Separate local and global injectivity. If g is univalent, every iterate g^n and every h_n is univalent. Their limit has derivative one at zero, so Proposition 11.4 makes h univalent on $\mathbb{D}$. Conversely, if h is globally univalent and $g(z) = g(w)$, then $\lambda h(z) = h(g(z)) = h(g(w)) = \lambda h(w)$. Since $\lambda \neq 0$, injectivity of h implies $z = w$. Thus g must be univalent. □

The conjugacy interpretation now follows on a small neighborhood: $h \circ g \circ h^{-1}(\zeta) = \lambda\zeta$ wherever the two sides are defined. On the whole disk one has a semiconjugacy unless global injectivity has been established. The distinction matters for maps with critical points or multiple preimages. A holomorphic change of coordinate cannot remove a global failure of injectivity.

Corollary 14.3 (The actual first-order orbit scale)**.** *Under the hypotheses of Theorem 14.2, for every $K \Subset \mathbb{D}$,*

$$g^n(z) = \lambda^n h(z) + O_K(|\lambda|^{2n}) \quad (z \in K). \tag{14.4}$$

Moreover, $h(z) = 0$ if and only if some iterate of z is zero. If no iterate is zero, then $g^{n+1}(z)/g^n(z) \to \lambda$ and $|g^n(z)|^{1/n} \to |\lambda|$.

Proof. Choose a neighborhood on which h is univalent. Its inverse has expansion $h^{-1}(\zeta) = \zeta + O(\zeta^2)$. For all sufficiently large n, attraction puts $g^n(K)$ in this neighborhood, and $h(g^n(z)) = \lambda^n h(z)$. The set $h(K)$ is bounded. Applying the inverse expansion uniformly to $\lambda^n h(K)$ proves (14.4).

If $g^N(z) = 0$, the functional equation gives $\lambda^N h(z) = 0$, so $h(z) = 0$. Conversely, if $h(z) = 0$, choose N so large that $g^N(z)$ is in the neighborhood where h is injective. There $h(g^N(z)) = 0 = h(0)$ implies $g^N(z) = 0$. When $h(z) \neq 0$, divide the asymptotic formula by $\lambda^n h(z)$ to get a factor tending to one. Ratios of successive iterates and nth roots of their moduli give the two limits. For an orbit eventually equal to zero these ratios are undefined; it would be incorrect to state the ratio limit without that exception. □

For an attracting fixed point $p \neq 0$, apply this result to $g = T_p \circ f \circ T_p^{-1}$. Its multiplier is still $f'(p)$, because the derivatives of the two coordinate changes cancel at the fixed point. The inverse

coordinate has expansion $T_p^{-1}(\zeta) = p + (1-|p|^2)\zeta + O(\zeta^2)$, so the same argument translates the orbit asymptotic back to the original disk. The prefactor depends on the starting point through $h(T_p(z))$; the multiplier controls the exponential scale. This is stronger than the earlier estimate with an arbitrary compact-set factor q_r, which was sufficient for convergence but need not be sharp.

14.2.1 Continuity of the normalized linearizing coordinate

The limit in (14.2) involves infinitely many iterations. Continuity in the map cannot be obtained by simply interchanging that limit with a parameter limit. The geometric tail estimate provides the uniform control needed to justify the interchange, as long as the limiting multiplier stays in the nonzero attracting range.

Proposition 14.4. *Let $g_j : \mathbb{D} \to \mathbb{D}$ be holomorphic maps fixing zero, and suppose $g_j \to g$ locally uniformly, where $0 < |g'(0)| < 1$. For all sufficiently large j, the normalized Koenigs function h_j of g_j exists, and $h_j \to h$ locally uniformly on $\mathbb{D}$, with locally uniform convergence of every derivative.*

Proof. Derivative convergence gives $\lambda_j = g_j'(0) \to \lambda = g'(0)$. Choose constants a, b, q such that

$$0 < a < |\lambda| < b < q < 1, \qquad q^2 < a.$$

These choices are possible by first taking q strictly between $|\lambda|$ and $\sqrt{|\lambda|}$, then choosing a between q^2 and $|\lambda|$, and b between $|\lambda|$ and q. Eventually $a \le |\lambda_j| \le b$.

On a fixed closed disk inside $\mathbb{D}$, Cauchy's estimates give a common bound for the second derivatives of g_j, because the maps are bounded by one on a slightly larger disk. Taylor's integral remainder consequently gives one M with $|g_j(z) - \lambda_j z| \le M|z|^2$ there for every sufficiently large j. Choose a smaller $r > 0$ with $b + Mr \le q$. Then all the maps satisfy $|g_j(z)| \le q|z|$ on $|z| \le r$, and the telescoping proof of Theorem 14.2 gives the uniform tail

$$\sup_{|z|\le r} |h_j(z) - \lambda_j^{-n} g_j^n(z)| \le \frac{Mr^2}{a(1-\vartheta)}\vartheta^n, \qquad \vartheta = q^2/a < 1.$$

The limiting map satisfies the same estimates. For each fixed n, finite composition and $\lambda_j \to \lambda \neq 0$ imply $\lambda_j^{-n} g_j^n \to \lambda^{-n} g^n$ uniformly on $|z| \le r$. Given an error tolerance, first choose n making both uniform tails small, and only then choose j for this finite composition. The triangle inequality proves $h_j \to h$ on the small disk.

Now fix $K \Subset \mathbb{D}$. Attraction of g gives an N with $g^N(K) \subset B(0, r/2)$. Finite-iterate convergence puts $g_j^N(K)$ in $B(0, 3r/4)$ for large j. The exact identities

$$h_j(z) = \lambda_j^{-N} h_j(g_j^N(z)), \qquad h(z) = \lambda^{-N} h(g^N(z))$$

transfer the already proved local convergence to K. Cauchy's formula on compact neighborhoods gives convergence of all derivatives. The order of the choices, first r and the common tail, then N for K, and finally j, avoids assuming a uniform number of entry iterates over the whole open disk. □

For maps whose attracting fixed points move, first use Proposition 14.8 to obtain $p_j \to p$, and conjugate by T_{p_j}. The explicit disk automorphisms converge locally uniformly with their inverses, so the proposition applies whenever the limiting multiplier is nonzero. A multiplier tending to zero changes the appropriate normal form, and a multiplier tending to the unit circle destroys the strict common tail bound. Neither limiting case is included in the statement.

14.2.2 A zero multiplier has a different scale

The division by λ^n is unavailable when $g'(0) = 0$. There is nevertheless an elementary quantitative conclusion. Suppose g is nonconstant, fixes zero, and has first nonzero Taylor term az^m with $m \geq 2$. Choose $r > 0$ and $A > 0$ so that

$$|g(z)| \leq A|z|^m \quad (|z| \leq r), \qquad A^{1/(m-1)}r < 1.$$

Such choices are possible by the Taylor expansion and then shrinking the radius. Put $v_n = A^{1/(m-1)}|g^n(z)|$. The displayed estimate implies $v_{n+1} \leq v_n^m$, and $v_0 < 1$ ensures that all iterates remain in the same disk. Induction gives

$$|g^n(z)| \leq A^{-1/(m-1)}\big(A^{1/(m-1)}|z|\big)^{m^n} \qquad (|z| \leq r). \tag{14.5}$$

For a compact set of initial points elsewhere in the disk, first use attraction to enter a smaller such disk after one common finite number of iterates. Formula (14.5) then applies uniformly to the remaining iterates. The exponent m^n describes a much faster scale than a fixed geometric factor. The constant map $g \equiv 0$ is simpler and reaches the fixed point in one step; it has no first nonzero Taylor term and is handled separately. We have proved a decay estimate, not asserted the additional existence of a Böttcher conjugacy.

14.3 Horodisks and a sequential Julia lemma

For $\tau \in \partial\mathbb{D}$ define

$$H_\tau(z) = \frac{|\tau - z|^2}{1 - |z|^2}, \qquad z \in \mathbb{D}. \tag{14.6}$$

The set $E(\tau, R) = \{H_\tau < R\}$, where $R > 0$, is a Euclidean disk: completing the square gives

$$E(\tau, R) = B\left(\frac{\tau}{1+R}, \frac{R}{1+R}\right). \tag{14.7}$$

Its boundary is internally tangent to the unit circle at τ. Its closure in $\overline{\mathbb{D}}$ meets $\partial\mathbb{D}$ at that one point. We call it a horodisk; its boundary circle inside $\mathbb{D}$, together with the limiting tangency point when discussing Euclidean closure, is a horocycle. Formula (14.7) fixes which parameter is being used: R is not its Euclidean radius.

The Cayley coordinate based at τ is

$$C_\tau(z) = \frac{\tau + z}{\tau - z} = \frac{1 + \overline{\tau}z}{1 - \overline{\tau}z}. \tag{14.8}$$

It maps $\mathbb{D}$ bijectively onto the right half-plane $\mathcal{H} = \{\operatorname{Re} w > 0\}$ and satisfies

$$\operatorname{Re} C_\tau(z) = \frac{1 - |z|^2}{|\tau - z|^2} = H_\tau(z)^{-1}. \tag{14.9}$$

Thus nested horodisks become right half-planes bounded by vertical lines. This coordinate will turn boundary attraction into elementary growth toward infinity.

Lemma 14.5 (Sequential Julia inequality)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic. Suppose $a_j \to \tau \in \partial\mathbb{D}$, $f(a_j) \to \eta \in \partial\mathbb{D}$, and*

$$\frac{1 - |f(a_j)|^2}{1 - |a_j|^2} \longrightarrow \alpha < \infty.$$

Then $\alpha > 0$ and

$$H_\eta(f(z)) \leq \alpha H_\tau(z) \qquad (z \in \mathbb{D}). \tag{14.10}$$

Proof. Schwarz–Pick and the identity

$$1-\delta(u,v)^2=\frac{(1-|u|^2)(1-|v|^2)}{|1-\overline{v}u|^2}$$

give, after rearrangement of positive factors,

$$\frac{|1-\overline{f(a_j)}f(z)|^2}{1-|f(z)|^2}\le\frac{1-|f(a_j)|^2}{1-|a_j|^2}\frac{|1-\overline{a_j}z|^2}{1-|z|^2}.$$

For fixed $z\in\mathbb{D}$, every denominator in this formula is positive. Take $j\to\infty$ and use $|\eta|=|\tau|=1$ to obtain (14.10). Its left side is strictly positive, because $f(z)$ is interior and η is on the circle. Therefore α cannot be zero. □

This statement uses one specified sequence and one specified finite limit. It makes no assumption that f extends to the boundary, and it does not silently identify α with a derivative at a boundary point. The existence and interpretation of angular derivatives require a further theorem. The sequential form suffices for iteration.

Lemma 14.6 (Wolff's invariant horodisks)**.** *If a holomorphic self-map f of $\mathbb{D}$ has no fixed point in $\mathbb{D}$, there are $\tau\in\partial\mathbb{D}$ and $0<\alpha\le 1$ such that*

$$H_\tau(f(z))\le\alpha H_\tau(z)\qquad(z\in\mathbb{D}).\tag{14.11}$$

In particular, every horodisk $E(\tau,R)$ is mapped into itself.

Proof. For $0<r<1$ consider $f_r=rf$. It has a fixed point p_r in $\mathbb{D}$. To prove existence without a boundary-continuity assumption, choose $r<\rho<1$. On $|z|=\rho$ one has $|rf(z)|<r<\rho$, so Rouché's theorem says that $z-rf(z)$ has exactly one zero in $B(0,\rho)$. It cannot have a second fixed point elsewhere in the disk: a disk self-map with two distinct interior fixed points is the identity by Schwarz's equality case, while $rf(\mathbb{D})\subset B(0,r)$ is not the identity image. The fixed point satisfies $p_r=rf(p_r)$ and $|p_r|<r$.

Take $r_j\uparrow 1$ and select a subsequence with $p_{r_j}\to\tau\in\overline{\mathbb{D}}$. An interior limit would satisfy $f(\tau)=\tau$, contrary to the hypothesis. Thus $|\tau|=1$, and $f(p_{r_j})=p_{r_j}/r_j\to\tau$. Moreover,

$$0<\beta_j=\frac{1-|p_{r_j}|^2/r_j^2}{1-|p_{r_j}|^2}\le 1.$$

Take a further subsequence on which $\beta_j\to\alpha\in[0,1]$. Lemma 14.5 shows that $\alpha>0$ and gives (14.11). The horodisk inclusion follows directly by comparing with R. □

The approximating maps rf have interior fixed points, but those points move to the circle as $r\uparrow 1$. Their limiting geometry is a horodisk, rather than a hyperbolic ball with an interior center. This is how an interior Schwarz inequality gives information about boundary dynamics without evaluating f on the boundary.

14.4 From an invariant horodisk to an attracting point

Suppose first that f has no interior fixed point and is not an automorphism. Fix $z_0\in\mathbb{D}$ and put $z_n=f^n(z_0)$. The consecutive hyperbolic distances

$$s_n=d_{\mathbb{D}}(z_n,z_{n+1})$$

form a decreasing sequence, by Schwarz–Pick, and hence have a finite limit $s \geq 0$. The orbit cannot have an interior accumulation point. Indeed, if $z_{n_j} \to p \in \mathbb{D}$, continuity gives $z_{n_j+1} \to f(p)$ and $z_{n_j+2} \to f^2(p)$, whence

$$d_{\mathbb{D}}(p, f(p)) = s = d_{\mathbb{D}}(f(p), f^2(p)).$$

Here $p \neq f(p)$ because there is no fixed point. Equality in Schwarz–Pick for that pair would make f an automorphism, a contradiction. Notice that no assumption $s > 0$ was needed; if $s = 0$ the first equality itself would already give a fixed point.

By Wolff's lemma the entire orbit lies in the closed Euclidean horodisk determined by $H_\tau(z) \leq H_\tau(z_0)$. That compact subset of $\overline{\mathbb{D}}$ meets the unit circle only at τ. Since there are no interior accumulation points, every accumulation point of the orbit is τ. Compactness of $\overline{\mathbb{D}}$ now proves $z_n \to \tau$.

The convergence of one orbit is automatically locally uniform for all starting points. If $K \Subset \mathbb{D}$, then $q = \max_{z \in K} \delta(z, z_0) < 1$. Put $x_n = f^n(z_0)$ and $y_n = f^n(z)$. Schwarz–Pick gives $|y_n - x_n| \leq q|1 - \overline{x_n} y_n|$. Since

$$|1 - \overline{x_n} y_n| \leq 1 - |x_n|^2 + |x_n|\,|y_n - x_n|,$$

we obtain the uniform bound

$$\sup_{z \in K} |f^n(z) - x_n| \leq \frac{q}{1-q}(1 - |x_n|^2) \longrightarrow 0. \tag{14.12}$$

Thus $f^n \to \tau$ locally uniformly. This estimate is also useful when one orbit can be computed exactly but the others cannot.

An automorphism needs a separate argument because equality in Schwarz–Pick is then expected. Let f have no interior fixed point and let τ be supplied by Wolff's lemma. In the Cayley coordinate,

$$F = C_\tau \circ f \circ C_\tau^{-1} : \mathcal{H} \longrightarrow \mathcal{H}$$

is a Möbius automorphism. Equations (14.9) and (14.11) imply $\operatorname{Re} F(w) \geq \operatorname{Re} w$. In particular, $F(x) \to \infty$ as $x \to +\infty$ along the real axis. A Möbius map with this property fixes infinity, so it has the form $Aw + B$. Positivity of its real part for every $x + iy$ with $x > 0$ forces A to be positive real; otherwise varying y would violate positivity, or the map would fail to be invertible. Surjectivity onto the right half-plane forces $\operatorname{Re} B = 0$. Consequently

$$F(w) = sw + it, \qquad s > 0, \quad t \in \mathbb{R}. \tag{14.13}$$

The inequality gives $s \geq 1$. If $s = 1$, then $t \neq 0$, since the identity would have interior fixed points. In that case $F^n(w) = w + nit \to \infty$ locally uniformly in the spherical sense, or equivalently its modulus tends to infinity on compact subsets. If $s > 1$, then

$$F^n(w) = s^n \left(w + \frac{it}{s-1} \right) - \frac{it}{s-1},$$

whose real part is $s^n \operatorname{Re} w \to \infty$, uniformly on compact subsets of $\mathcal{H}$. In both cases $C_\tau^{-1}(F^n(w))$ tends locally uniformly to τ. This proves the boundary case for automorphisms as well.

Theorem 14.7 (Denjoy–Wolff for the disk)**.** *Let $f : \mathbb{D} \to \mathbb{D}$ be holomorphic and not an elliptic automorphism. There is a unique $\tau \in \overline{\mathbb{D}}$ such that*

$$f^n(z) \longrightarrow \tau$$

locally uniformly on $\mathbb{D}$. If f has an interior fixed point, then τ is that fixed point. Otherwise $\tau \in \partial\mathbb{D}$, and every horodisk tangent at τ is invariant under f.

Proof. The interior case is Theorem 14.1. If there is no interior fixed point, Wolff's lemma and the two preceding arguments prove convergence for nonautomorphisms and automorphisms, respectively. A sequence cannot have two different constant limits, so τ is unique. □

The statement includes constant maps: if $f \equiv p \in \mathbb{D}$, every iterate after the zeroth is p. It excludes all elliptic automorphisms, including the identity, for exactly the reasons given earlier. No extension of f to $\partial\mathbb{D}$ occurs in the hypotheses.

14.5 What boundary attraction does and does not say

The invariant horodisk inequality gives a boundary value along nontangential approaches. More precisely, suppose

$$z \to \tau, \qquad |z-\tau| \le M(1-|z|)$$

for a fixed M. Then

$$H_\tau(z) = \frac{|z-\tau|^2}{(1-|z|)(1+|z|)} \le \frac{M|z-\tau|}{1+|z|} \longrightarrow 0.$$

The inequality $H_\tau(f(z)) \le \alpha H_\tau(z)$ implies $|f(z)-\tau|^2 \le H_\tau(f(z)) \to 0$. Thus $f(z) \to \tau$ along every such approach. This proves a boundary fixed value in the nontangential sense. It does not establish continuity along every possible tangential approach, nor does it establish an angular derivative formula.

If the constant in (14.11) can be taken with $\alpha < 1$, the same proof gives a quantitative estimate. Iteration of (14.9) yields

$$\operatorname{Re} C_\tau(f^n(z)) \ge \alpha^{-n} \operatorname{Re} C_\tau(z).$$

Since $C_\tau^{-1}(w) = \tau(w-1)/(w+1)$, it follows that

$$|f^n(z)-\tau| \le \frac{2}{1+\alpha^{-n}\operatorname{Re} C_\tau(z)}. \tag{14.14}$$

On a compact subset of the disk the real part in the denominator has a positive minimum. This is an exponential Euclidean estimate under the extra strict horodisk factor. When only $\alpha = 1$ is available, the estimate does not give decay; the convergence theorem remains true, but a separate argument is needed to measure its rate.

For a proper simply connected domain U, choose a conformal map $\psi : \mathbb{D} \to U$ and conjugate a self-map $F : U \to U$ to $f = \psi^{-1} \circ F \circ \psi$. The disk theorem describes iteration in the coordinate ψ^{-1}. If its attracting point lies inside $\mathbb{D}$, then F^n converges locally uniformly to the corresponding interior point of U. If it lies on $\partial\mathbb{D}$, the conclusion that F^n converges to a particular Euclidean boundary point of U requires further information about ψ at that point. For example, if ψ has a continuous extension there, composition gives the claimed boundary limit. Without such a hypothesis, a conformal coordinate approaching the unit circle is not itself a proof of a unique Euclidean landing point on an arbitrary domain boundary. The prime-end compactification is designed to record the relevant boundary access; its construction is a separate topological subject.

14.6 Stability of an attracting interior point

Normal-family arguments also explain which dynamical features persist under a small analytic perturbation. The interior case has a particularly clean answer because its fixed point is a simple zero of $f(z) - z$.

Proposition 14.8 (Persistence of an attracting fixed point)**.** *Let holomorphic self-maps $f_j : \mathbb{D} \to \mathbb{D}$ converge locally uniformly to f. Suppose f has an interior fixed point p and is not an elliptic automorphism. Then, for all sufficiently large j, f_j has a unique interior fixed point p_j, $p_j \to p$, and p_j is attracting. For each compact set of initial points, convergence of f_j^n to p_j has a geometric bound in normalized disk coordinates whose factor is less than one and independent of sufficiently large j.*

Proof. Schwarz rigidity after conjugation gives $|f'(p)| < 1$, so $f'(p) - 1 \neq 0$. Thus $f(z) - z$ has a simple zero at p. Choose a small closed disk around p with no other zero and no boundary zero. Rouché's theorem gives exactly one zero of $f_j(z) - z$ inside it for large j. Taking successively smaller disks proves $p_j \to p$. Derivative convergence gives $f_j'(p_j) \to f'(p)$, so eventually $|f_j'(p_j)| < 1$. In particular, these maps are not the identity; Schwarz's equality case rules out a second interior fixed point.

Put $g_j = T_{p_j} \circ f_j \circ T_{p_j}^{-1}$ and $g = T_p \circ f \circ T_p^{-1}$. The explicit formulas and convergence of the parameters show $g_j \to g$ locally uniformly. The quotients $g_j(z)/z$, continued at zero by $g_j'(0)$, converge locally uniformly to $g(z)/z$. Given a compact set $K \Subset \mathbb{D}$, choose $r < 1$ such that $T_{p_j}(K) \subset B(0, r)$ for all large j. The maximum of $|g(z)/z|$ on $|z| \leq r$ is less than one. Therefore the corresponding maxima for all large j are bounded by one common $q < 1$. The induction in (14.1) gives

$$|T_{p_j}(f_j^n(z))| \leq q^n |T_{p_j}(z)| \qquad (z \in K),$$

as required. □

This result controls two limits without interchanging them unconditionally: first choose a compact set of starting points, then take j sufficiently large to obtain a common factor, and finally iterate. The hypothesis excludes a limiting rotation, where the factor can approach one and the attracting point can cease to be determined by the limit map.

14.7 Worked investigations in disk dynamics

Investigation 14.9 (Three exact boundary models)**.** Use $C(z) = (1 + z)/(1 - z)$ to transfer the following maps of the right half-plane to disk self-maps: $w \mapsto sw$ with $s > 1$; $w \mapsto w + b$ with $b > 0$; and $w \mapsto w + it$ with $t \neq 0$. Find their iterates, their attracting disk boundary point, and their Euclidean convergence rates for a fixed initial point.

Solution. All three have attracting disk point 1, since $C^{-1}(w) = 1 - 2/(w + 1)$. Their disk iterates are respectively

$$1 - \frac{2}{s^n C(z) + 1}, \qquad 1 - \frac{2}{C(z) + nb + 1}, \qquad 1 - \frac{2}{C(z) + nit + 1}.$$

The first error is of order s^{-n}, and the last two are of order n^{-1}, with nonzero leading constants for each fixed $z \in \mathbb{D}$. The first and third maps are automorphisms. The second has image $\{\operatorname{Re} w > b\}$ in half-plane coordinates and is not onto. For the vertical translation, the real part remains constant, so the orbit stays on a single horocycle and approaches 1 tangentially. For the positive real translation, the real part tends to infinity and the orbit enters arbitrarily small horodisks. Thus the same Euclidean order of convergence can accompany different approach geometry. The n^{-1} examples also show why a geometric rate cannot be concluded from the disk iteration theorem alone.

Investigation 14.10 (A nonlinear map with an exact linearizing coordinate)**.** Let $0 < \lambda < 1$ and $g(z) = \lambda z/[1 + (1 - \lambda)z]$. Prove that g maps $\mathbb{D}$ to itself, find its normalized Koenigs function, and compute its iterates. Does the Koenigs image have to be bounded?

Solution. For $|z| = r < 1$, the denominator has modulus at least $1 - (1 - \lambda)r$, so

$$|g(z)| \le \frac{\lambda r}{1 - (1 - \lambda)r} < 1.$$

The last inequality is equivalent to $r < 1$, and the denominator cannot vanish in $\mathbb{D}$. Thus g is a holomorphic disk self-map, with $g(0) = 0$ and $g'(0) = \lambda$. Direct substitution gives

$$h(z) = \frac{z}{1 + z}, \qquad h(g(z)) = \lambda h(z), \qquad h(0) = 0, \quad h'(0) = 1.$$

Uniqueness in Theorem 14.2 identifies this rational function with the limit of normalized iterates. Solving $h^{-1}(w) = w/(1 - w)$ yields

$$g^n(z) = \frac{\lambda^n z}{1 + (1 - \lambda^n)z}.$$

The denominator has no zero in the disk for any n. Expanding it on a compact subset verifies $g^n(z) = \lambda^n h(z) + O_K(\lambda^{2n})$ directly. Although the original disk is bounded, h is unbounded as $z \to -1$ from inside it. The theorem controls compact subsets and does not bound the image of its change of coordinate. In this example g and h are both univalent, as the global criterion predicts.

Investigation 14.11 (A semiconjugacy with more than one zero)**.** For the nonlinear disk map $g(z) = z(a + z)/(1 + az)$, $0 < a < 1$, show that its normalized Koenigs function cannot be injective. Explain how its zero set affects the orbit asymptotic.

Solution. The factor $(a + z)/(1 + az)$ is a disk automorphism, and multiplication by z keeps the image in $\mathbb{D}$. The multiplier at zero is a, so the Koenigs function exists. Both 0 and $-a$ map to zero. The functional equation gives $h(-a) = 0 = h(0)$, proving failure of injectivity without needing a formula for h. More generally, every point that maps to zero in finitely many steps is a zero of h, and the corollary proves the converse. For these exceptional points the orbit is eventually exactly zero. For every other starting point the leading coefficient $h(z)$ is nonzero and $g^{n+1}(z)/g^n(z) \to a$. Thus a statement about the typical multiplier scale must still specify the finite-preimage exception.

Investigation 14.12 (Two limits of an attracting family)**.** Let $g_j(z) = \lambda_j z$ with $0 < \lambda_j < 1$ and $\lambda_j \uparrow 1$. Compare the limits obtained by first taking the iteration number to infinity or first taking j to infinity.

Solution. For every fixed j, $g_j^n(z) = \lambda_j^n z \to 0$ as $n \to \infty$. Taking j afterward still gives zero. For every fixed n, however, $g_j^n(z) \to z$ as $j \to \infty$, and taking n afterward leaves z. At a nonzero point the two iterated limits differ. The Koenigs functions themselves are all $h_j(z) = z$, so their continuity alone cannot give a uniform attracting rate as the multiplier approaches the unit circle. The common bound $q < 1$ in Proposition 14.8 and the strict parameter range in Proposition 14.4 are the hypotheses that prevent this loss of control in their respective statements.

Investigation 14.13 (A boundary point becomes undetermined)**.** Let $0 < r_j < 1$ tend to one and let $\tau_j \in \partial\mathbb{D}$. Study $f_j(z) = r_j z + (1 - r_j)\tau_j$. What happens if τ_j alternates between 1 and -1?

Solution. The triangle inequality gives $|f_j(z)| \le r_j|z| + (1 - r_j) < 1$, so each map is a disk self-map. Its only fixed point in $\mathbb{C}$ is τ_j, which is on the boundary, and direct induction gives

$$f_j^n(z) = \tau_j + r_j^n(z - \tau_j).$$

Its Denjoy–Wolff point is therefore τ_j. On the other hand, $|f_j(z) - z| \le (1 - r_j)(1 + |z|) \to 0$ locally uniformly, independently of the choices of τ_j. Alternating these choices makes the attracting points fail to converge although the maps tend to the identity. This does not contradict Proposition 14.8: the limiting map is precisely an excluded elliptic automorphism.

Investigation 14.14 (An interior nonlinear example)**.** For $0 < a < 1$, let $g(z) = z(a+z)/(1+az)$. Prove that all iterates converge to zero, find the multiplier there, and give a bound of the form (14.1) on $|z| \le r < 1$.

Solution. The factor $(a + z)/(1 + az)$ is a disk automorphism, so its modulus is less than one on $\mathbb{D}$. Hence g is a disk self-map fixing zero, with $g'(0) = a$, and it is not a rotation. A direct calculation gives

$$\left|\frac{a+z}{1+az}\right| \le \frac{a+r}{1+ar} \quad (|z| \le r).$$

For completeness, the maximum on the circle $|z| = r$ follows by writing the squared ratio as $(a^2 + r^2 + 2ar\cos\theta)/(1 + a^2r^2 + 2ar\cos\theta)$; it increases with $\cos\theta$, because the difference between denominator and numerator is $(1 - a^2)(1 - r^2) > 0$. The maximum principle covers the interior. Thus

$$|g^n(z)| \le \left(\frac{a+r}{1+ar}\right)^n |z| \quad (|z| \le r).$$

The factor approaches one as $r \uparrow 1$, even though every compact subdisk has a strict geometric rate. At zero the derivative of the nth iterate is a^n by the chain rule.

Investigation 14.15 (An alternative local-uniformity argument)**.** Suppose the iterates of a disk self-map converge pointwise on $\mathbb{D}$ to one boundary point τ. Prove local uniform convergence without using estimate (14.12).

Solution. The iterates are uniformly bounded by one, so every subsequence has a locally uniformly convergent further subsequence. Its holomorphic limit has value τ at every point by pointwise convergence, and is therefore the constant τ. If the original sequence failed to converge uniformly on a compact set, a subsequence separated from τ there would contradict the normal-family conclusion. This proof is short and qualitative. Estimate (14.12) additionally relates the error on a compact set to the location of one chosen orbit.

Investigation 14.16 (The approximating fixed points have one limit)**.** For a fixed-point-free disk self-map, let p_r be the fixed point of rf constructed in Lemma 14.6. Show that p_r tends to the Denjoy–Wolff point as $r \uparrow 1$, without selecting a subsequence in the final assertion.

Solution. Any sequence $r_j \uparrow 1$ has a subsequence for which p_{r_j} converges in $\overline{\mathbb{D}}$. Its limit is on the circle, and the construction in Wolff's lemma produces invariant horodisks tangent at that limit. The proof of the boundary iteration theorem then forces every orbit to converge to that same point. Uniqueness of the orbit limit identifies it with the Denjoy–Wolff point τ. Thus every convergent subsequence of the compact family of points has limit τ, which proves $p_r \to \tau$. For the affine map $f(z) = az + (1 - a)$, $0 < a < 1$, one can check the conclusion directly: $p_r = r(1 - a)/(1 - ra) \to 1$.

Investigation 14.17 (Transport to a half-plane and the meaning of infinity)**.** Consider $F(w) = w + 1$ on the right half-plane. Describe the limit of its iterates in Euclidean and spherical terms, and compare it with the disk theorem after conjugation.

Solution. One has $F^n(w) = w + n$. There is no finite Euclidean limit, but the modulus tends uniformly to infinity on every compact subset of the half-plane. Thus the iterates tend to the point at infinity in the Riemann sphere. Under C^{-1} they tend to the finite disk boundary point 1. The two descriptions are compatible because C^{-1} extends continuously at infinity on the sphere. On a general uniformized domain such an extension at the attracting coordinate point is an additional boundary assertion, not a consequence of local uniform convergence inside the disk.

Chapter 15

Harmonic analysis and the hyperbolic metric

The same density that bounds holomorphic derivatives also bounds the relative variation of positive harmonic functions. It appears again as the diagonal size of a reproducing kernel and as the conformal factor in a nonlinear curvature equation. These links explain why the disk metric is useful beyond the theory of maps between surfaces. We develop them with the normalizations fixed, so that their constants can be transported to other domains.

15.1 Harmonic conjugates and the Poisson formula

A real C^2 function u is harmonic if $\Delta u = 0$. On a convex domain the differential form $-u_y\,dx + u_x\,dy$ is closed, because the difference of its cross derivatives is Δu. Green's theorem on triangles, followed by the same segment construction as for complex primitives, gives a function v with $v_x = -u_y$ and $v_y = u_x$. The function $u + iv$ satisfies the Cauchy–Riemann equations and is holomorphic. Its imaginary additive constant is the only freedom. Thus the local harmonic theory follows from the analytic foundations already proved, together with elementary real differentiation and integration.

For $0 \le r < 1$ put

$$P_r(\theta) = \frac{1 - r^2}{1 - 2r\cos\theta + r^2} = \operatorname{Re}\frac{1 + re^{i\theta}}{1 - re^{i\theta}} = 1 + 2\sum_{n \ge 1} r^n \cos(n\theta).$$

The series converges uniformly for r in any compact subinterval of $[0, 1)$. In particular $P_r > 0$ and its mean on a period is one. If u is harmonic near the closed disk of radius R, then

$$u(re^{i\theta}) = \frac{1}{2\pi}\int_0^{2\pi} P_{r/R}(\theta - t)\,u(Re^{it})\,dt, \qquad r < R. \tag{15.1}$$

One proof writes the holomorphic function with real part u as a power series on a slightly larger disk. Integration against the displayed Fourier series multiplies its nth angular mode by $(r/R)^n$. The constant mode is unchanged, giving exactly its value on the smaller circle. Uniform convergence justifies all integrations. In particular the center value is the average of the boundary values.

For continuous boundary data φ on the unit circle, the same integral defines a harmonic function in the open disk and extends continuously to the prescribed data. Harmonicity follows by differentiating the locally uniformly convergent kernel series. For the boundary limit, given an angular distance $\delta > 0$, split the integral into $|t - \theta| < \delta$ and its complement. Uniform continuity of φ controls the first part. On the complement, $1 - 2r\cos(t - \theta) + r^2$ is bounded below away

from zero for r near one, while $1-r^2$ tends to zero. The total mass there therefore tends to zero. Positivity and unit mass control the remaining constant factors. This proves uniform convergence to φ.

The preceding approximate-identity argument can also be quantified. Let $\omega_\varphi(\delta)$ be the maximum oscillation of φ between circle points of angular separation at most δ, where $0 < \delta \le \pi$. On the complementary arc,

$$1 - 2r\cos(t-\theta) + r^2 = (1-r)^2 + 4r\sin^2((t-\theta)/2) \ge 4r\sin^2(\delta/2).$$

For $0 < r < 1$, positivity and unit kernel mass therefore give

$$\sup_\theta |u(re^{i\theta}) - \varphi(e^{i\theta})| \le \omega_\varphi(\delta) + \frac{\|\varphi\|_\infty (1-r^2)}{2r\sin^2(\delta/2)}.$$

First choose δ so that the oscillation is small, then let $r \uparrow 1$ with that δ fixed. This order of choices avoids assuming a boundary convergence rate that continuity alone does not provide. The estimate is not intended at $r = 0$, where the Poisson formula itself has no singularity and supplies the center average.

Uniqueness follows from the maximum principle for harmonic functions. For completeness, the mean-value formula shows that an interior maximum forces the function to be constant on every sufficiently small circle and then on a neighborhood. Alternatively a harmonic conjugate gives a holomorphic exponential whose modulus has a maximum. Connectedness propagates constancy. A continuous harmonic function on the closed disk with zero boundary data must consequently vanish.

15.2 Harnack inequalities as metric statements

Theorem 15.1 (Harnack comparison)**.** *If $u > 0$ is harmonic on $\mathbb{D}$, then for every $z, w \in \mathbb{D}$,*

$$e^{-d_{\mathbb{D}}(z,w)} \le \frac{u(z)}{u(w)} \le e^{d_{\mathbb{D}}(z,w)}.$$

In particular, with $|z| = r$,

$$\frac{1-r}{1+r}u(0) \le u(z) \le \frac{1+r}{1-r}u(0).$$

The constants cannot be improved uniformly over positive harmonic functions.

Proof. Apply the Poisson formula on the radius-R disk, $r < R < 1$. The kernel is between $(R-r)/(R+r)$ and $(R+r)/(R-r)$, while the mean of the positive boundary values is $u(0)$. Integrating and letting $R \uparrow 1$ proves the centered bounds. Compose with a disk automorphism sending zero to w; harmonicity is preserved by a conformal change of coordinates, as follows from the Laplacian chain rule. Since $d_{\mathbb{D}}(z,w) = \log((1+|m_w(z)|)/(1-|m_w(z)|))$, the general statement follows. The positive harmonic function $u(z) = \operatorname{Re}\frac{1+z}{1-z}$ attains the upper bound on the positive radius and the lower bound on the negative radius. Rotating it gives sharpness in every radial direction. □

The logarithm $\log u$ is therefore a function with Lipschitz constant one for hyperbolic distance. It need not be harmonic. Its differential version is

$$|\nabla \log u(z)| \le \frac{2}{1-|z|^2}. \tag{15.2}$$

To prove this directly, choose a holomorphic $h = u + iv$ on the disk. The range of h is in the right half-plane, so its Schwarz derivative bound is $|h'| \le 2u/(1-|z|^2)$. The Cauchy–Riemann

equations give $|h'| = |\nabla u|$, proving (15.2). Integrating along paths gives the preceding theorem by a second route. On a simply connected hyperbolic domain U, conformal transfer gives $|\nabla \log u| \le \lambda_U$ in a planar coordinate. The density and gradient both transform by the same factor, so this is intrinsic.

Investigation 15.2 (Positive data on an interior disk)**.** Suppose $u > 0$ is harmonic on the unit disk and $u(0) = 3$. Give the best universal upper and lower bounds on $|z| \le 1/2$. Determine whether the same numbers control an arbitrary harmonic function with center value three.

Solution. The Harnack bounds are $1 \le u \le 9$, and both are attained on the half-radius by scaled and rotated Poisson kernels. Positivity is essential. The harmonic functions $u_N(z) = 3 + N \operatorname{Re} z$ all have center value three, but their values at $\pm 1/2$ escape in opposite directions. Knowing a center value supplies no oscillation bound without a sign, boundary bound or another normalization.

15.3 Compactness of normalized positive harmonic fields

The positivity assumption has a second consequence: one value controls an entire family on every compact subset. This is useful when boundary data become concentrated and no uniform bound on the boundary survives.

Proposition 15.3 (Harmonic compactness)**.** *Let $u_j \ge 0$ be harmonic on $\mathbb{D}$ and suppose $u_j(0) \le M$ for one finite constant M. There is a subsequence converging uniformly on compact subsets to a nonnegative harmonic function u. If every $u_j(0) = 1$, the limit is strictly positive and $u(0) = 1$.*

Proof. The Harnack estimate also holds for nonnegative functions: apply it to $u_j + \epsilon$ and let $\epsilon \downarrow 0$. On $|z| \le R < 1$ it therefore gives

$$0 \le u_j(z) \le M \frac{1+R}{1-R} =: M_R.$$

Similarly, the differential estimate applied to $u_j + \epsilon$ gives $|\nabla u_j(z)| \le 2u_j(z)/(1-|z|^2)$ after taking the limit. Consequently the gradients on the closed radius-R disk are bounded by $2M_R/(1-R^2)$, independently of j. Integrating along straight segments in that disk supplies a common Lipschitz constant. We have both uniform boundedness and equicontinuity, not merely bounded values on a chosen grid.

Here is the compactness step explicitly. Choose a countable dense subset of the radius-R disk. Repeatedly extract convergent subsequences of the bounded values at its points and take the diagonal subsequence. For any tolerance, a finite sufficiently fine net and the common Lipschitz bound turn convergence at the net points into uniform Cauchy convergence on the whole disk. Completeness of the real numbers gives a continuous uniform limit. Repeat on radii increasing to one and diagonalize again, obtaining a locally uniform limit u on $\mathbb{D}$.

To prove harmonicity without differentiating a uniform limit, take any closed disk $\overline{B(a,\rho)} \subset \mathbb{D}$. The Poisson formula for u_j on that disk passes to the limit under the boundary integral, because convergence on the boundary is uniform and the kernel is integrable. Thus u inside this disk equals the Poisson extension of its continuous boundary values. That extension is harmonic by the kernel calculation above. Since such disks cover $\mathbb{D}$, u is harmonic. Nonnegativity passes to the limit pointwise. In the normalized case $u(0) = 1$, and Harnack gives the strictly positive lower bound $u(z) \ge (1-|z|)/(1+|z|)$. □

For example, let $|\xi_j| = 1$ and suppose $\xi_j \to \xi$. The fields $P_{\xi_j}(z) = (1-|z|^2)/|\xi_j - z|^2$ have center value one and converge locally uniformly to P_ξ. On $|z| \le R$, every denominator is at least $(1-R)^2$, so this convergence also follows directly from the formula. Their values close to their

respective boundary poles are unbounded. The conclusion is deliberately local inside the disk; compactness here does not mean uniform convergence on the closed disk or convergence of finite boundary values at the pole.

The same reasoning applies on a connected domain with one fixed normalization point, provided compact sets can be joined to it by finite chains of interior disks. Harnack comparisons along a chain propagate the initial bound. For the disk proof above no such covering argument was needed, because the centered estimate already controls every compact subset. Without a normalization, even the constant functions $u_j = j$ have no finite locally uniform subsequential limit.

Locally uniform convergence of harmonic functions also controls every fixed order of their interior derivatives. Indeed, take concentric disks with radii $r < R$ whose closed outer disk lies in the domain. The difference $u_j - u$ equals its boundary Poisson integral on the outer disk. Every Cartesian derivative of the kernel is uniformly bounded for points in the inner disk, because the source point remains a positive distance from the boundary. Differentiation under this fixed, smooth integral bounds each derivative of $u_j - u$ by a constant times its boundary supremum. That supremum tends to zero. The argument explains why convergence of harmonic fields has stronger interior consequences than uniform convergence of arbitrary continuous functions; the constants deteriorate as the inner disk approaches the outer circle.

15.4 Horocycles and a differential boundary coordinate

Fix an ideal boundary point ξ with $|\xi| = 1$ and define

$$P_\xi(z) = \frac{1 - |z|^2}{|\xi - z|^2}, \qquad b_\xi(z) = -\log P_\xi(z).$$

The first function is the Poisson kernel directed toward ξ. Its level sets are circles tangent to the unit circle at ξ; these are horocycles. The sublevel sets of b_ξ are horodisks. For $\xi = 1$, the upper-half-plane Cayley map $C(z) = i(1 + z)/(1 - z)$ sends them to horizontal lines and upper horoballs in the half-plane, because $P_1(z) = \operatorname{Im} C(z)$. Thus $b_1 = -\log y$ there.

The hyperbolic Laplacian in a planar conformal coordinate is $\Delta_h = \lambda^{-2}\Delta$, with the convention that Δ has positive second derivatives. In the half-plane, $\Delta_h = y^2(\partial_x^2 + \partial_y^2)$. Direct calculation gives

$$|\nabla_h b_\xi| = 1, \qquad \Delta_h b_\xi = 1.$$

For $b = -\log y$, the Euclidean gradient has size $1/y$ and the Euclidean Laplacian is $1/y^2$; multiplying by the metric factors proves both identities. Conformal invariance transports them to the disk. In contrast, $P_\xi = e^{-b_\xi}$ is harmonic because

$$\Delta_h e^{-b} = e^{-b}(|\nabla_h b|^2 - \Delta_h b) = 0.$$

These formulas connect an ideal boundary direction, a family of tangent circles and a positive harmonic function. They also explain the horodisks in the boundary dynamics of disk maps without postulating Euclidean boundary differentiability of the map.

For real s, the same computation gives

$$\Delta_h(P_\xi^s) = s(s-1)P_\xi^s. \tag{15.3}$$

This is an exact local eigenfunction identity. It is not a statement that P_ξ^s is square integrable over the whole hyperbolic plane, nor a classification of a self-adjoint operator's spectrum. Boundary conditions and integrability determine the operator domain and must be added before such a spectral statement is meaningful.

15.5 The Bergman kernel from a complete orthogonal expansion

Let $A^2(\mathbb{D})$ consist of holomorphic functions square integrable with respect to Euclidean area, with $\|f\|_2^2 = \int_{\mathbb{D}} |f|^2 \, dx\, dy$. For $f(z) = \sum_{n\ge 0} a_n z^n$, orthogonality on circles and monotone convergence as their radii increase to one give

$$\|f\|_2^2 = \pi \sum_{n\ge 0} \frac{|a_n|^2}{n+1}. \tag{15.4}$$

The formula is valid with either side infinite. It follows first on a smaller disk from uniform convergence of the Taylor series. The resulting nonnegative coefficient sum increases to the displayed sum. Conversely a sequence with finite displayed sum defines a holomorphic function: the Cauchy–Schwarz inequality gives uniform absolute convergence on each smaller disk. Thus this coefficient correspondence is an isometry onto a weighted sequence space, proving completeness without assuming it separately.

For each $a \in \mathbb{D}$, Cauchy–Schwarz gives

$$|f(a)|^2 \le \|f\|_2^2 \sum_{n\ge 0} \frac{n+1}{\pi} |a|^{2n} = \frac{\|f\|_2^2}{\pi(1-|a|^2)^2}.$$

The representing function for evaluation is

$$K_a(z) = \frac{1}{\pi(1-\bar{a}z)^2}, \qquad f(a) = \int_{\mathbb{D}} f(z)\overline{K_a(z)}\, dx\, dy.$$

The integral identity follows coefficient by coefficient, with convergence in the Hilbert norm. The two-variable kernel is $K_{\mathbb{D}}(z,w) = 1/(\pi(1-z\bar{w})^2)$. Equality in the evaluation bound holds precisely for scalar multiples of K_a, including the zero function. This gives a genuine extremal problem with a complete solution, not just an upper estimate.

Proposition 15.4 (Minimum area norm with a point constraint)**.** *Among $f \in A^2(\mathbb{D})$ with $f(a) = c$, the minimum squared norm is*

$$\pi(1-|a|^2)^2|c|^2,$$

attained uniquely by $f(z) = c(1-|a|^2)^2/(1-\bar{a}z)^2$.

Proof. The evaluation inequality gives the lower bound. The displayed function has the prescribed value and is proportional to K_a, so it attains the bound. Uniqueness follows from the equality case of Cauchy–Schwarz, or from the orthogonal decomposition into the evaluation kernel and its orthogonal complement. □

Notice the choice of area. We use Euclidean area for $A^2(\mathbb{D})$, not hyperbolic area. The constant function belongs to this space, whereas the hyperbolic area of the entire disk is infinite. Interchanging these measures would change the function space and invalidate the displayed kernel.

15.6 Several values and an exact minimum-norm solution

The one-point extremal problem extends to finitely many values with a fully explicit positive matrix. This matrix is a Gram matrix in the area-norm space; it is different from the Pick matrix for a supremum bound developed in Chapter 17.

Take distinct $a_1, \dots, a_N \in \mathbb{D}$ and prescribed values $y_1, \dots, y_N \in \mathbb{C}$. Set

$$G_{ij} = K_{\mathbb{D}}(a_i, a_j) = \frac{1}{\pi(1 - a_i\overline{a_j})^2}.$$

This matrix is positive definite. To see all parts of that assertion, give the area space the inner product $\langle f, g\rangle = \int_{\mathbb{D}} f\overline{g}$, linear in its first argument. For any coefficient column c,

$$\left\| \sum_j c_j K_{a_j} \right\|_2^2 = c^* G c \geq 0.$$

If this norm vanishes, the holomorphic function inside it vanishes identically: a nonzero value would give, by continuity, a small disk with positive area integral. Pair this zero function with a Lagrange polynomial p_k satisfying $p_k(a_j) = \delta_{jk}$. Such a polynomial is explicitly $p_k(z) = \prod_{j\neq k}(z - a_j)/(a_k - a_j)$ and belongs to $A^2(\mathbb{D})$. Reproduction then gives $0 = \langle p_k, \sum_j c_j K_{a_j}\rangle = \overline{c_k}$. All coefficients vanish. Thus G is invertible.

Put $c = G^{-1}y$ and $f_* = \sum_j c_j K_{a_j}$. Evaluation at each node gives $f_*(a_i) = (Gc)_i = y_i$, so all value constraints are met. If f is any other admissible function, $h = f - f_*$ vanishes at every a_i. Therefore

$$\langle h, f_*\rangle = \sum_j \overline{c_j} h(a_j) = 0, \qquad \|f\|_2^2 = \|f_*\|_2^2 + \|h\|_2^2.$$

The unique minimum is consequently

$$\|f_*\|_2^2 = y^* G^{-1} y. \tag{15.5}$$

This is existence, uniqueness and the exact optimal value in one calculation. A finite polynomial already interpolates arbitrary values; the kernel solution answers the additional norm-minimization question. There is no feasibility obstruction analogous to a failed Pick matrix unless a numerical norm ceiling is imposed.

Example 15.5 (Two constraints becoming one derivative constraint)**.** Require $f(0) = 0$ and $f(a) = 1$, with $0 < |a| < 1$, and write $t = |a|^2$. Here

$$G = \frac{1}{\pi}\begin{pmatrix} 1 & 1 \\ 1 & (1-t)^{-2} \end{pmatrix}.$$

Its inverse, or the orthogonal projection of K_a away from the constant kernel, gives

$$f_*(z) = \frac{(1 - \overline{a}z)^{-2} - 1}{(1-t)^{-2} - 1}, \qquad \|f_*\|_2^2 = \frac{\pi(1-t)^2}{t(2-t)}.$$

The denominator is positive because $0 < t < 1$. As $a \to 0$ this norm diverges like $\pi/(2|a|^2)$. Two fixed, different values cannot be maintained at colliding nodes with a bounded area norm.

Instead prescribe $f(a) = ca$ with a fixed $c \in \mathbb{C}$. Multiplying the preceding interpolant by ca gives minimum squared norm tending to $\pi|c|^2/2$. Its formula converges locally uniformly to cz: the numerator is $2\overline{a}z + O(|a|^2)$ on a fixed compact disk, and the denominator is $2|a|^2 + O(|a|^4)$. This agrees with the derivative problem $f(0) = 0$, $f'(0) = c$, whose coefficient norm (15.4) has minimum $\pi|c|^2/2$ attained by cz. The scaled data, rather than the unscaled data, have a finite derivative limit.

15.6.1 A value and a derivative away from the center

One can solve a first-derivative area-norm problem without introducing a new reproducing kernel. Fix $a \in \mathbb{D}$ and require $f(a) = c$ and $f'(a) = d$. Write $s = 1 - |a|^2 > 0$ and let $F(z) = m_a^{-1}(z) = (a + z)/(1 + \overline{a}z)$. The weighted composition $g = (f \circ F)F'$ preserves the area norm by the Jacobian change of variables, and differentiation gives

$$g(0) = cs, \qquad g'(0) = ds^2 - 2\overline{a}\, cs,$$

since $F'(0) = s$ and $F''(0) = -2\overline{a}s$. The weight F' is essential: plain composition would not preserve Euclidean area norm, and would omit the second term of the derivative constraint.

The coefficient formula (15.4) now proves that the minimum squared norm is exactly

$$\pi|cs|^2 + \frac{\pi}{2}|ds^2 - 2\overline{a}\, cs|^2. \tag{15.6}$$

The unique minimizing transformed function is the affine polynomial $g_*(z) = cs + (ds^2 - 2\overline{a}\, cs)z$; all higher coefficients add nonnegative terms to the norm and are unnecessary for the two constraints. Transforming back gives the explicit unique minimizer

$$f_*(w) = g_*(m_a(w))m_a'(w).$$

All denominators are nonzero inside the disk. The chain rule confirms the two original data, and norm preservation confirms the displayed optimal value. This proves the result without supposing that a boundary integral exists for an arbitrary member of A^2.

As a consistency check, if the derivative is free, minimizing (15.6) over d chooses $d = 2\overline{a}c/s$ and leaves $\pi s^2|c|^2$, exactly the one-point minimum. At $a = 0$, $c = 1$, $d = 2$, the minimizing function is $1 + 2z$ and its squared area norm is 3π. It is not a Schur function: a Schur function taking the value one at zero must be constant. Area-norm interpolation and supremum-norm interpolation impose different admissibility conditions even for the same first jet.

15.7 Conformal transfer and a kernel expression for the metric

If $F : \mathbb{D} \to U$ is a conformal bijection, the map

$$T : A^2(U) \longrightarrow A^2(\mathbb{D}), \qquad (Tg)(z) = g(F(z))F'(z)$$

is unitary. Change of variables gives equality of norms, since the real Jacobian of F is $|F'|^2$. Its inverse is the corresponding weighted composition with F^{-1}. Applying the evaluation representation through this isometry gives

$$K_U(F(z), F(w)) = \frac{K_{\mathbb{D}}(z, w)}{F'(z)\overline{F'(w)}}. \tag{15.7}$$

One can check the conjugates in this transformation without guessing them from the diagonal case. For $g \in A^2(U)$, reproduction in the disk states

$$g(F(w))F'(w) = \int_{\mathbb{D}} g(F(z))F'(z)\overline{K_{\mathbb{D}}(z, w)}\, dA(z).$$

Substitute $\zeta = F(z)$ and $dA(z) = dA(\zeta)/|F'(z)|^2$, then divide by $F'(w)$. The resulting kernel in the integral is the conjugate of $K_{\mathbb{D}}(z, w)/(F'(z)\overline{F'(w)})$, exactly as in (15.7). This candidate belongs to the area space by the same unitary change of variables. Uniqueness follows because two representing vectors with the same pairing against every area-integrable holomorphic function have zero squared difference norm when paired against their difference. The calculation thus

establishes the full two-variable identity, including its phase factors, rather than only matching the positive diagonal values. In particular,

$$K_U(a,a) = \frac{\lambda_U(a)^2}{4\pi} \tag{15.8}$$

for every proper simply connected plane domain. One can also see this directly at $z = 0$: both sides equal $1/(\pi|F'(0)|^2)$. The equality is a special feature of the simply connected setting; the Bergman kernel of an arbitrary multiply connected surface is not obtained by substituting its hyperbolic density in this formula.

With $\partial_z = (\partial_x - i\partial_y)/2$, define here the kernel metric by $ds_B^2 = 2\partial_z\partial_{\bar z}\log K_U(z,z)\,|dz|^2$. On the disk, $\log K_{\mathbb{D}} = -\log\pi - 2\log(1-|z|^2)$, so

$$2\partial_z\partial_{\bar z}\log K_{\mathbb{D}}(z,z) = \frac{4}{(1-|z|^2)^2} = \lambda_{\mathbb{D}}(z)^2.$$

The factor two is part of the stated convention. Under a conformal change, the additional term $-\log|F'|^2$ in (15.7) is harmonic, so the metric transforms correctly. We therefore recover the complete curvature -1 metric on a simply connected domain from an extremal problem in a Hilbert space of holomorphic functions. This is a concrete bridge between the functional analysis of Volume V and the present geometry.

Investigation 15.6 (A disk with physical scale)**.** For $U = B(a,R)$ compute $K_U(w,w)$ and the minimum Euclidean area norm of a holomorphic function with value one at the center.

Solution. Using $F(z) = a + Rz$ gives $K_U(w,w) = R^2/[\pi(R^2-|w-a|^2)^2]$. At the center it is $1/(\pi R^2)$, so the minimum squared norm is πR^2, attained by the constant function one. This checks both the Jacobian factor and the units: a kernel diagonal has inverse-area units, and its reciprocal has the units of an area norm squared for a dimensionless function value.

15.8 Dirichlet energy and conformal invariance in two dimensions

For a real C^1 function u on a planar domain U, define $E_U(u) = \frac{1}{2}\int_U |\nabla u|^2\,dx\,dy$, allowing infinity. Under a conformal diffeomorphism $F : V \to U$,

$$|\nabla(u\circ F)|^2 = |F'|^2|\nabla u|^2 \circ F.$$

The same $|F'|^2$ occurs in the area Jacobian, so $E_V(u\circ F) = E_U(u)$. This cancellation is particular to dimension two for the quadratic energy. It is the analytical reason that logarithms and conformal coordinates are natural in planar electrostatic models.

Now assume U is bounded, connected and smooth, and that all functions under discussion are smooth up to its boundary. For a harmonic u with fixed smooth boundary values, Green's formula shows it minimizes this energy among smooth competitors with those values. If $v = u + h$ and h vanishes on the boundary, then

$$E(v) = E(u) + \int_U \nabla u\cdot\nabla h + E(h) = E(u) + E(h),$$

because the middle integral is $-\int_U h\Delta u + \int_{\partial U} h\partial_n u = 0$. Equality forces $\nabla h = 0$, hence $h = 0$ on a connected domain with nonempty boundary. This proves uniqueness within the stated regularity class. Existence for arbitrary rough domains or weak boundary traces is a separate question; the disk Poisson formula supplies existence for the continuous disk data treated here.

If real boundary data have Fourier expansion $\varphi(\theta) = a_0 + \sum_{n\ge1}(a_n\cos n\theta + b_n\sin n\theta)$, their harmonic extension has radial factors r^n. For a finite sum, orthogonality and integration give

$$E_{\mathbb{D}}(u) = \frac{\pi}{2}\sum_{n\ge1} n(a_n^2 + b_n^2). \tag{15.9}$$

For each mode the radial and angular derivative contributions sum to $2\pi n^2 r^{2n-2}(a_n^2 + b_n^2)$ after angular integration. The radial integral contributes $1/(2n)$, and the energy definition contributes $1/2$. For an infinite expansion with finite displayed sum, the finite extensions form a Cauchy sequence in energy; their local analytic convergence identifies the limit with the Poisson extension. Formula (15.9) then follows by Hilbert-space orthogonality. It quantifies why high-frequency boundary oscillations cost more energy even though they decay rapidly in the interior.

Investigation 15.7 (Controlled truncation of a boundary field)**.** Suppose $|a_n| + |b_n| \le Mq^n$ for $n \ge 1$, where $0 < q < 1$. Give a uniform error bound on $|z| \le r < 1$ after retaining modes $n \le N$, and an energy bound for the omitted modes.

Solution. The absolute value of the omitted field is at most $M\sum_{n>N}(qr)^n = M(qr)^{N+1}/(1-qr)$. Since $a_n^2 + b_n^2 \le M^2 q^{2n}$, the omitted energy is at most

$$\frac{\pi M^2}{2}\sum_{n>N} nq^{2n} = \frac{\pi M^2 q^{2N+2}}{2}\left(\frac{N+1}{1-q^2} + \frac{q^2}{(1-q^2)^2}\right).$$

The first bound controls a supremum norm on a specified interior disk; the second controls the whole-disk energy. They are different error criteria. Neither is obtained by inspecting a finite sample of values.

15.9 Boundary flux and the cost of one Fourier mode

For smooth real boundary data φ, let u be their harmonic extension to the disk. Define the outward boundary flux operator by $\Lambda\varphi = \partial_r u|_{r=1}$. Smoothness here can be read concretely from Fourier coefficients: repeated integration by parts in the coefficient integrals gives decay faster than any fixed inverse power of n. Hence the Fourier series below and the required derivative series converge uniformly. There is no exchange of a boundary derivative with a merely pointwise convergent series.

Differentiating each radial factor r^n gives

$$\Lambda\varphi(\theta) = \sum_{n\ge1} n(a_n\cos n\theta + b_n \sin n\theta).$$

Green's identity, with $\Delta u = 0$, yields

$$E_{\mathbb{D}}(u) = \frac{1}{2}\int_0^{2\pi} \varphi(\theta)\Lambda\varphi(\theta)\, d\theta. \tag{15.10}$$

This recovers (15.9) by the orthogonality of the trigonometric functions. More generally, if u, v extend smooth data φ, ψ, respectively, integration by parts gives

$$\int_0^{2\pi} \varphi\Lambda\psi\, d\theta = \int_{\mathbb{D}} \nabla u\cdot\nabla v\, dx\, dy = \int_0^{2\pi} \psi\Lambda\varphi\, d\theta.$$

The operator is therefore symmetric and nonnegative on these real smooth data. Its kernel consists precisely of constants: zero pairing in (15.10) forces $\nabla u = 0$, and the connected disk then forces constant boundary data. This is a statement on the specified smooth class, without an unstated assertion about an unbounded operator's full Hilbert-space domain.

For the individual datum $\varphi(\theta) = A\cos(n\theta)$, $n \ge 1$, the extension is $Ar^n\cos(n\theta)$ and $E = \pi n A^2/2$. The flux is $nA\cos(n\theta)$. Thus equally large boundary amplitudes can produce arbitrarily large flux and energy as n increases, even though at a fixed interior radius their amplitudes decrease like r^n. A small field well inside the domain does not imply a small energy concentrated near its boundary.

Investigation 15.8 (Finite amplitude and infinite energy)**.** Show that the boundary series $\varphi(\theta) = \sum_{n\ge 2} \cos(n\theta)/(n\log n)$ converges at every point except possibly $\theta = 0$ but does not supply a continuous bounded datum there. Then give a different, continuous boundary datum whose harmonic extension has infinite energy.

Solution. For $\theta \notin 2\pi\mathbb{Z}$, the partial sums of $\sum_{n=2}^{N} e^{in\theta}$ are bounded by the finite geometric-series formula. Summation by parts and the decrease of $1/(n\log n)$ to zero give convergence. At zero the positive series diverges by the integral test, so this proposed datum cannot be described as continuous on the circle. To obtain the requested continuous example, use instead

$$\psi(\theta) = \sum_{k\ge 1} 2^{-k} \cos(4^k\theta).$$

The sum converges uniformly by comparison with $\sum 2^{-k}$, hence is continuous. Its Poisson extension is the uniformly convergent sum of $2^{-k} r^{4^k} \cos(4^k\theta)$. On each radius-$R$ disk, orthogonality gives energy $\frac{\pi}{2} \sum_{k\ge 1} 4^k 2^{-2k} R^{2\cdot 4^k}$. Letting $R \uparrow 1$ makes this sum diverge, since every limiting summand is $\pi/2$. Continuous boundary values ensure existence of a continuous harmonic extension; they do not ensure finite Dirichlet energy.

Chapter 16

Curvature equations, stability and geometric field models

Hyperbolic geometry also appears as a nonlinear field equation for a conformal factor. This chapter derives that equation, an associated variational problem and a quantitative residual estimate. The models are mathematical: identifying their variables with physical fields requires the modeling assumptions stated for the intended application. No claim about an experiment follows from the solvability or stability of the differential equation.

16.1 The conformal connection and its curvature

Let $ds^2 = e^{2u}(dx^2 + dy^2)$ on an open set, with $u \in C^2$. The metric coefficients are $g_{ij} = e^{2u}\delta_{ij}$ and their inverse is $g^{ij} = e^{-2u}\delta_{ij}$. Substituting into the connection formula gives

$$\Gamma^k_{ij} = \delta_{jk}u_i + \delta_{ik}u_j - \delta_{ij}u_k.$$

For example $\Gamma^1_{11} = u_x$, $\Gamma^1_{12} = u_y$, $\Gamma^1_{22} = -u_x$, and the other three distinct entries are $\Gamma^2_{11} = -u_y$, $\Gamma^2_{12} = u_x$ and $\Gamma^2_{22} = u_y$. To check the sign and the cancellation in the curvature calculation, use the convention

$$R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z.$$

The coefficient of ∂_x in $R(\partial_x, \partial_y)\partial_y$ is

$$\partial_x\Gamma^1_{22} - \partial_y\Gamma^1_{12} + \sum_{m=1}^{2}\Gamma^m_{22}\Gamma^1_{1m} - \sum_{m=1}^{2}\Gamma^m_{12}\Gamma^1_{2m}.$$

The derivative terms give $-u_{xx} - u_{yy}$. The first sum is $-u_x^2 + u_y^2$ and the second is $u_y^2 - u_x^2$; they cancel exactly. Taking the metric inner product with ∂_x multiplies by e^{2u}, while the squared area of the coordinate parallelogram is $g_{11}g_{22} - g_{12}^2 = e^{4u}$. Their quotient therefore yields

$$K = -e^{-2u}\Delta u. \tag{16.1}$$

This is the coordinate formula used earlier in the curvature contraction proof. Consequently a metric has curvature $-\kappa$, $\kappa > 0$, exactly when

$$\Delta u = \kappa e^{2u}.$$

The disk solution for $\kappa = 1$ is $u(z) = \log 2 - \log(1 - |z|^2)$. Scaling the density by $\kappa^{-1/2}$ subtracts $\frac{1}{2}\log\kappa$ from u and changes the curvature to $-\kappa$. A constant added to u changes a geometric scale and cannot be discarded as an arbitrary additive normalization in the curvature equation.

If f is a holomorphic local diffeomorphism and u is such a solution, the pulled-back conformal factor is $v = u \circ f + \log|f'|$. The Laplacian chain rule and the harmonicity of $\log|f'|$ give $\Delta v = \kappa e^{2u\circ f}|f'|^2 = \kappa e^{2v}$. At a critical point the pulled-back density vanishes and is no longer a positive smooth metric. This is why the Ahlfors comparison treats critical points separately instead of taking a global logarithm there.

16.2 A strictly convex Dirichlet variational problem

Let $\Omega \subset \mathbb{C}$ be a bounded smooth domain and fix a smooth boundary value for u. On smooth functions with that trace define

$$\mathcal{J}(u) = \int_\Omega \left(\tfrac{1}{2}|\nabla u|^2 + \tfrac{\kappa}{2}e^{2u}\right) dx\, dy.$$

For a smooth perturbation h vanishing on the boundary,

$$\frac{d}{dt}\mathcal{J}(u+th)\Big|_{t=0} = \int_\Omega (-\Delta u + \kappa e^{2u})h\, dx\, dy.$$

The Euler equation is precisely the constant negative-curvature equation. The second variation in the same direction is

$$\frac{d^2}{dt^2}\mathcal{J}(u+th) = \int_\Omega \left(|\nabla h|^2 + 2\kappa e^{2(u+th)}h^2\right) dx\, dy > 0$$

for nonzero h. Hence the functional is strictly convex along every nonconstant admissible line. A smooth critical point is its unique smooth minimizer with those boundary values: integrate the positive second derivative along the segment from the critical point to a competitor. This establishes uniqueness if a solution exists. It does not prove existence for every possible boundary datum or for a singular boundary condition; such existence requires a separate compactness or PDE argument.

The disk solution restricted to $\overline{B(0,r)}$, $r<1$, supplies an explicit example with smooth finite boundary values and finite action. The complete unit-disk density diverges at the unit circle. Its conformal factor is not a finite smooth boundary datum for this bounded-domain problem, and its full-disk action is infinite. Treating it as an admissible finite-energy minimizer would conflate two different variational settings.

16.2.1 An exact action gap and its linearized equation

There is a stronger form of the minimization calculation that identifies which norm the functional controls. Suppose u is a smooth solution and h is smooth with zero boundary trace. Expand the gradient square and integrate its cross term by parts. Using $\Delta u = \kappa e^{2u}$ gives the exact identity

$$\mathcal{J}(u+h) - \mathcal{J}(u) = \frac{1}{2}\int_\Omega |\nabla h|^2 + \frac{\kappa}{2}\int_\Omega e^{2u}(e^{2h} - 1 - 2h). \tag{16.2}$$

Every integral in this paragraph is with respect to Euclidean area. The elementary inequality $e^x \geq 1 + x$ shows nonnegativity of the second term. Hence the energy difference always bounds half the squared gradient norm. No smallness assumption on h was used.

If both u and $u+h$ are at least m on the closure, the whole segment $u+th$, $0 \leq t \leq 1$, is also at least m. Taylor's formula with integral remainder gives

$$e^{2(u+h)} - e^{2u} - 2e^{2u}h = 4h^2\int_0^1 (1-t)e^{2(u+th)}\, dt \geq 2e^{2m}h^2.$$

Consequently the gap is at least $\frac{1}{2}\|\nabla h\|_2^2 + \kappa e^{2m}\|h\|_2^2$. This quantitative convexity compares actual admissible functions. It does not infer their existence from a formal second derivative.

For a differentiable family of smooth solutions u_t with the same κ but possibly varying boundary values, differentiation of the equation at $t = 0$ gives, for $h = \partial_t u_t|_0$,

$$(\Delta - 2\kappa e^{2u_0})h = 0.$$

The parameter differentiation is justified here by the stipulated smooth dependence in the interior, not by an existence theorem for such a family. With zero boundary variation, multiply by h and integrate to obtain $\int |\nabla h|^2 + 2\kappa \int e^{2u_0}h^2 = 0$, hence $h = 0$. The linearized equation has no nonzero smooth solution with homogeneous Dirichlet data. With nonzero boundary data, the same maximum argument as below gives $\|h\|_\infty \le \|h\|_{\infty,\partial\Omega}$: a positive interior maximum above the boundary bound would have $\Delta h \le 0$ but $2\kappa e^{2u_0}h > 0$.

Another comparison keeps the boundary value fixed and varies curvature. Suppose u_1, u_2 solve their respective equations for $0 < \kappa_1 < \kappa_2$. A positive interior maximum of $u_2 - u_1$ would satisfy

$$\Delta(u_2 - u_1) = \kappa_2 e^{2u_2} - \kappa_1 e^{2u_1} > 0,$$

contradicting the second derivative test. Thus $u_2 \le u_1$. Here the finite boundary trace has been held fixed. Scaling a complete metric, which changes that trace or its asymptotics, is a different comparison. Specifying what remains fixed is necessary for interpreting the direction of this inequality.

16.3 An explicit finite-boundary curvature problem

The uniqueness statements above do not prove existence for arbitrary data. A disk with constant finite boundary value provides a useful case where existence, smoothness and limiting behavior can all be checked directly. Fix $R > 0$, $\kappa > 0$ and a real boundary value b. For a parameter $0 < a < 1/R$, put

$$u_a(z) = \log\frac{2a}{\sqrt{\kappa}} - \log(1 - a^2|z|^2), \qquad |z| \le R.$$

For the radial coordinate $\rho > 0$,

$$u_a'(\rho) = \frac{2a^2\rho}{1 - a^2\rho^2}, \qquad u_a''(\rho) + \frac{u_a'(\rho)}{\rho} = \frac{4a^2}{(1 - a^2\rho^2)^2} = \kappa e^{2u_a(\rho)}.$$

At zero the Cartesian formula is smooth, and the displayed radial Laplacian has the same limiting value $4a^2$ as its Cartesian Laplacian. Thus the apparent division by ρ introduces no singularity in the solution.

The boundary condition $u_a(R) = b$ is equivalent, on writing $K = \sqrt{\kappa}e^b > 0$, to $KR^2a^2 + 2a - K = 0$. Its positive root can be written without a subtraction of nearly equal positive numbers as

$$a = \frac{K}{1 + \sqrt{1 + K^2R^2}}. \tag{16.3}$$

It is positive, and $aR < 1$ because $KR < 1 + \sqrt{1 + K^2R^2}$. The second quadratic root is negative and is outside the chosen parameter range. This constructs a smooth solution on the closed disk for every finite real b. The maximum comparison, or the strict convexity argument, proves uniqueness among all smooth solutions with this boundary value, not just among the radial ansatz used to find it.

Now let $b \to +\infty$. Then $K \to \infty$ and $a \to 1/R$. On every compact subset of the open disk, u_a and its fixed-order derivatives converge to

$$u_\infty(z) = \log\frac{2R}{\sqrt{\kappa}} - \log(R^2 - |z|^2).$$

The limiting density is the complete curvature $-\kappa$ disk density. Completeness can be checked along any approach to the boundary using $e^{u_\infty} \geq 1/[\sqrt{\kappa}(R-|z|)]$; the radial lower bound has divergent integral. Each finite-b density is bounded up to the boundary and has finite distance to it. Completeness of the limit therefore does not imply completeness of the approximating metrics. A finite-boundary approximation cannot be treated as if it already had the limiting boundary condition.

In the opposite limit $b \to -\infty$, formula (16.3) gives $a \sim K/2$ and $u_a - b \to 0$ uniformly on the closed radius-R disk. The densities tend to zero. Their curvatures all equal $-\kappa$, but their pointwise metric limit is degenerate, so it is not another positive metric of that curvature. These two limits illustrate why convergence of formulas, nondegeneracy and boundary completeness must be checked as separate assertions.

16.4 A whole-domain error bound from a curvature residual

A small residual at a finite set of points says little about an equation between the points. The following estimate explains what a uniform residual bound, together with a boundary bound, would actually imply. It applies to any bounded domain where the listed functions extend continuously to the closure; no discretization is assumed.

Theorem 16.1 (Residual comparison)**.** *Let $u, v \in C^2(\Omega) \cap C(\overline{\Omega})$, where Ω is bounded, and suppose $\Delta u = \kappa e^{2u}$ with $\kappa > 0$. Assume throughout Ω that*

$$|\Delta v - \kappa e^{2v}| \leq \varepsilon, \qquad \kappa e^{2v} \geq A > \varepsilon \geq 0,$$

and $|u - v| \leq \delta$ on $\partial\Omega$. Set

$$t = \max\left\{\delta, -\tfrac{1}{2}\log(1 - \varepsilon/A)\right\}.$$

Then $|u - v| \leq t$ on $\overline{\Omega}$ and the two conformal densities obey $e^{-t} \leq e^u/e^v \leq e^t$ everywhere.

Proof. Write $F(w) = \Delta w - \kappa e^{2w}$ and $r = F(v)$. For the upper comparison function $v + t$,

$$F(v+t) = r - \kappa e^{2v}(e^{2t} - 1) \leq \varepsilon - A(e^{2t} - 1) \leq 0.$$

The last inequality follows since $e^{2t} \geq (1 - \varepsilon/A)^{-1} \geq 1 + \varepsilon/A$. If $u - (v + t)$ had a positive maximum, it would be attained in the interior because the boundary difference is nonpositive. There

$$\Delta(u - v - t) = \kappa(e^{2u} - e^{2(v+t)}) - F(v+t) > 0,$$

contrary to the second derivative test. Thus $u \leq v + t$. For the lower comparison function,

$$F(v-t) = r + \kappa e^{2v}(1 - e^{-2t}) \geq -\varepsilon + A(1 - e^{-2t}) \geq 0.$$

A positive interior maximum of $(v - t) - u$ would similarly give a positive Laplacian, again impossible. Hence $v - t \leq u$. Exponentiation proves the density ratio. □

The condition $A > \varepsilon$ is substantive: it lets the lower barrier absorb the residual without making its logarithmic correction undefined. The two one-sided errors need not be assigned the same constant. Suppose more specifically that $-\delta_- \leq u - v \leq \delta_+$ on the boundary, where both $\delta_\pm \geq 0$. The proof works with the separate shifts

$$t_+ = \max\{\delta_+, \tfrac{1}{2}\log(1 + \varepsilon/A)\}, \qquad t_- = \max\{\delta_-, -\tfrac{1}{2}\log(1 - \varepsilon/A)\},$$

and yields $-t_- \leq u - v \leq t_+$. Indeed, the upper barrier needs only $A(e^{2t_+} - 1) \geq \varepsilon$, while the lower needs $A(1 - e^{-2t_-}) \geq \varepsilon$. Solving these two inequalities gives exactly the displayed

logarithms. The symmetric theorem used the larger of them to keep a single constant. When $\varepsilon = 0$, they both reduce to the boundary errors; with zero boundary errors this proves uniqueness once more. There is no division by a vanishing residual in this endpoint case.

For a small ratio ε/A, the correction is $\varepsilon/(2A) + O((\varepsilon/A)^2)$, but the displayed logarithm is the exact bound. No asymptotic approximation is needed to use it. The boundary term is also indispensable. An exact solution of the PDE with the wrong boundary values has zero residual and can still differ substantially from the desired solution.

If two positive densities on the same domain obey a ratio bound $e^{-t} \le \lambda_1/\lambda_2 \le e^t$, integration along every admissible path and then taking infima gives

$$e^{-t} d_2(x, y) \le d_1(x, y) \le e^t d_2(x, y).$$

Thus the residual theorem controls intrinsic distances as well as pointwise metric coefficients. It controls the metrics defined on that same domain; it does not assert that their geodesics coincide or that extending the domain preserves the bound.

Investigation 16.2 (A specified residual tolerance)**.** Suppose the hypotheses hold with $\varepsilon/A = 1/100$ and zero boundary error. State an exact density-ratio bound and a simple rational upper bound.

Solution. We have $t = \frac{1}{2}\log(100/99)$, so $\sqrt{99/100} \le e^u/e^v \le \sqrt{100/99}$. The upper bound is less than $101/100$, because squaring and multiplying positive denominators reduces this to $1{,}000{,}000 < 1{,}009{,}899$. These constants concern a residual bound over all of Ω. A reported maximum on a finite mesh establishes the hypothesis only if an independent continuous error estimate covers the unsampled regions.

16.5 An integral residual estimate

The maximum comparison used a uniform residual. If an integral norm is the available information, a different conclusion is justified. Assume now that Ω is bounded with smooth boundary and that u, v are C^2 up to the closure, have the same boundary values, and are both at least m. Let

$$r = \Delta v - \kappa e^{2v}, \qquad w = u - v, \qquad c_0 = 2\kappa e^{2m} > 0.$$

Subtract the equations, multiply by w, and integrate by parts; the boundary term vanishes because $w = 0$ there. The result is

$$\int_\Omega |\nabla w|^2 + \kappa \int_\Omega w(e^{2u} - e^{2v}) = \int_\Omega rw. \tag{16.4}$$

The sign on the right matters: the convention is that r is the residual of the approximation v, and w is exact minus approximate. The fundamental theorem of calculus gives

$$w(e^{2u} - e^{2v}) = 2w^2 \int_0^1 e^{2(v+tw)}\, dt \ge 2e^{2m} w^2.$$

Write $R = \|r\|_2$ and $x = \|w\|_2$. Cauchy–Schwarz in (16.4) yields $\|\nabla w\|_2^2 + c_0 x^2 \le Rx$. If $x > 0$, dropping the gradient term and dividing gives $x \le R/c_0$; the same bound is automatic if $x = 0$. Completing a scalar square also gives $Rx - c_0 x^2 \le R^2/(4c_0)$. Hence

$$\|u - v\|_2 \le \frac{\|r\|_2}{c_0}, \qquad \|\nabla(u - v)\|_2 \le \frac{\|r\|_2}{2\sqrt{c_0}}. \tag{16.5}$$

These control an area norm and a gradient energy. They do not give the pointwise density ratio of Theorem 16.1. In dimension two, an area-norm estimate alone cannot be promoted to a uniform estimate by omitting a regularity argument. Nor can an uncontrolled boundary discrepancy be dropped from the integration by parts. The stated common boundary data and lower bound are part of the theorem proved by this calculation.

16.6 Geodesic separation and the meaning of instability

In geodesic polar coordinates on the curvature -1 plane, $ds^2 = dR^2 + \sinh^2 R\, d\theta^2$. Two nearby radial geodesics with angular separation $\delta\theta$ have infinitesimal transverse separation $J(R) = \sinh R\, \delta\theta$. It obeys $J'' - J = 0$, $J(0) = 0$, $J'(0) = \delta\theta$. More generally, a normal Jacobi field along a unit-speed geodesic in a constant-curvature surface obeys $J'' + KJ = 0$; this follows by varying the geodesic equation and inserting the curvature tensor. In the present model the explicit polar metric independently verifies the stated solution and its initial conditions.

For curvature $-\kappa$, the solution is $J(s) = \delta\theta \sinh(\sqrt{\kappa}s)/\sqrt{\kappa}$. The length scale is $1/\sqrt{\kappa}$. Exponential separation is therefore a statement about the geodesic flow and its initial-direction sensitivity. It does not contradict Schwarz contraction, which concerns a holomorphic map between fixed metric spaces. One compares nearby geodesics evolving under a flow; the other compares images under a specified analytic map. The variables and operations are different even though the same metric appears in both.

For a finite travel length $s \leq S$, the separation is bounded by $|\delta\theta| \sinh(\sqrt{\kappa}S)/\sqrt{\kappa}$ at the linearized level. Turning that derivative bound into a bound for finite angular separation requires integration over a specified variation of geodesics. On a geodesic circle the exact arc length between angles differing by θ is $|\theta| \sinh(\sqrt{\kappa}S)/\sqrt{\kappa}$ as long as the indicated arc is chosen. This is an upper bound for the geodesic distance of the endpoints, not generally an equality, since the shortest path may cut across the circle.

16.7 What the field interpretation contributes

The curvature equation is a nonlinear elliptic equation, its Dirichlet functional is strictly convex in the conformal factor, and its metric distance can be bounded from a uniform residual. These are exact mathematical statements under specified regularity and boundary assumptions. In a two-dimensional scalar-field model one may interpret the gradient term as spatial variation cost and the exponential term as a local potential. The coefficient κ then sets a geometric scale. This interpretation does not by itself identify a microscopic material, a quantum theory or a gravitational field equation in a different dimension.

The benefit for computation is similarly precise. A numerical solver could propose v; a separate enclosure of its residual and boundary error could then feed Theorem 16.1. The theorem specifies the acceptance information needed from that later computation. No numerical solver or symbolic certification run is required to prove the comparison, and none is claimed by the illustrations. This keeps analytical stability, discretization error and physical-model uncertainty as distinct questions.

Chapter 17

Finite interpolation and constructive function theory

An analytic function is constrained by its values much more strongly than an arbitrary continuous function. A finite table can already contradict the existence of a bounded analytic interpolant, even when every pair of rows satisfies Schwarz–Pick. Conversely, the correct finite condition constructs a function on the whole disk. This chapter develops that condition with an elementary recursive proof. The recursion will also explain why singular data require a different treatment from strictly positive data.

The result is the classical finite Nevanlinna–Pick theorem. The reduction method belongs to the classical Schur–Nevanlinna tradition; see Marshall [6]. Shalit's author notes [7] give a complementary operator-theoretic approach. The argument below uses only disk automorphisms, Schwarz's lemma and elementary Hermitian matrix algebra. It includes the endpoint cases rather than adding a strict-positivity assumption that would remove them.

17.1 The interpolation problem and its conventions

Write $\mathcal{S}$ for the Schur class: the functions holomorphic on $\mathbb{D}$ that satisfy $|f(z)| \leq 1$ for every $z \in \mathbb{D}$. Allowing the closed unit disk as a range is convenient during a recursion. By maximum modulus, a member of $\mathcal{S}$ either maps into the open disk or is a constant of modulus one. Thus this convention introduces a precisely understood exceptional family, not an ambiguity about boundary values.

For $n \geq 1$, prescribe distinct nodes $z_1, \ldots, z_n \in \mathbb{D}$ and complex targets $w_1, \ldots, w_n$. The question is whether

$$f(z_j) = w_j \quad (1 \leq j \leq n)$$

has a solution $f \in \mathcal{S}$. Define the Hermitian Pick matrix

$$P_{ij} = \frac{1 - w_i \overline{w_j}}{1 - z_i \overline{z_j}}. \tag{17.1}$$

Every denominator is nonzero because $|z_i \overline{z_j}| < 1$. Positive semidefinite means $\xi^* P \xi \geq 0$ for every $\xi \in \mathbb{C}^n$; positive definite means strict positivity for every nonzero ξ. A singular positive semidefinite matrix is permitted.

Repeated nodes are settled before constructing this matrix. If the same node has two different prescribed values, there is no function at all. If all prescriptions at that node agree, replace them by one row. Repeated nodes with derivative conditions form a different, confluent interpolation problem; no derivative condition is silently inferred here.

The diagonal condition $P_{ii} \geq 0$ already forces $|w_i| \leq 1$. There is therefore no need to assume target admissibility separately in the theorem. Its matrix criterion tests that condition too. The empty data set, when it appears as a terminal recursion, has every member of $\mathcal{S}$ as a solution.

17.2 Two elementary matrix facts

Lemma 17.1 (A zero diagonal entry)**.** *If a Hermitian matrix H is positive semidefinite and $H_{ii} = 0$, then $H_{ij} = 0$ for every j.*

Proof. For a fixed j, restrict the quadratic form to vectors supported in positions i, j. Its value at $(t, 1)$ is $2\,\mathrm{Re}(\bar{t}H_{ij}) + H_{jj}$. If $H_{ij} \neq 0$, choose $t = -RH_{ij}$ with $R > 0$. The expression becomes $H_{jj} - 2R|H_{ij}|^2$, which is negative for sufficiently large R. This contradicts positive semidefiniteness. □

Lemma 17.2 (Positive pivot and Schur complement)**.** *Let*

$$H = \begin{pmatrix} p & r^* \\ r & A \end{pmatrix}, \qquad p > 0.$$

Then $H \geq 0$ if and only if $A - rr^*/p \geq 0$. Moreover $\operatorname{rank} H = 1 + \operatorname{rank}(A - rr^*/p)$*, and H is positive definite if and only if its Schur complement is positive definite.*

Proof. For a scalar t and column x, complete the square:

$$\begin{pmatrix} t \\ x \end{pmatrix}^* H \begin{pmatrix} t \\ x \end{pmatrix} = p \left| t + \frac{r^* x}{p} \right|^2 + x^* \left(A - \frac{rr^*}{p} \right) x.$$

The invertible change of variables $(t, x) \mapsto (t + r^*x/p, x)$ makes the quadratic form block diagonal. It proves all three assertions. □

These facts explain why an algorithm must distinguish a zero pivot from a small positive pivot. The former imposes exact row identities. Dividing by it is not an approximation to a legitimate step. In symbolic data the distinction is exact; in uncertain data it is a question about the model and its uncertainty, addressed later by a quantitative margin.

17.3 Reduction of a function and reduction of its data

Recall $m_a(z) = (z - a)/(1 - \overline{a}z)$ and $m_a^{-1}(u) = (a + u)/(1 + \overline{a}u)$. Fix a node $a \in \mathbb{D}$ and an interior target $b \in \mathbb{D}$. Every $f \in \mathcal{S}$ with $f(a) = b$ has the form

$$f(z) = \frac{b + m_a(z) g(z)}{1 + \overline{b}\, m_a(z) g(z)}, \qquad g \in \mathcal{S}. \tag{17.2}$$

Indeed, apply Schwarz's lemma to $m_b \circ f \circ m_a^{-1}$, which vanishes at zero. After division by its variable, the removable singularity is filled by the derivative at zero, and the resulting function has modulus at most one. In the original coordinate it is

$$g(z) = \frac{m_b(f(z))}{m_a(z)}.$$

Conversely, $|m_a(z) g(z)| < 1$ for every $z \in \mathbb{D}$, even if g is a unimodular constant. The denominator in (17.2) has modulus at least $1 - |b|\,|m_a(z) g(z)| > 0$. The formula is holomorphic, disk-valued and takes the value b at a. Reduction and augmentation are inverses.

Use the first datum as the pivot: $a = z_1$, $b = w_1$, assuming $|b| < 1$. For $i \geq 2$ put

$$B_i = m_a(z_i), \qquad u_i = m_b(w_i), \qquad v_i = u_i/B_i. \tag{17.3}$$

The nodes are distinct, so $B_i \neq 0$. Once $|w_i| \leq 1$ is known, $1 - \overline{b}w_i \neq 0$ as well. The reduced interpolation problem has the same remaining nodes $z_2, \ldots, z_n$ and targets $v_2, \ldots, v_n$. It is equivalent to the original function problem by (17.2).

The corresponding matrix equivalence is worth deriving. Define $u_1 = 0$ and

$$\widetilde{P}_{ij} = \frac{1 - u_i\overline{u_j}}{1 - z_i\overline{z_j}}, \qquad d_i = \frac{1}{1 - \overline{b}w_i}.$$

The automorphism deficit identity gives the diagonal congruence

$$\widetilde{P} = (1 - |b|^2)\,\mathrm{diag}(d_i)\,P\,\mathrm{diag}(d_i)^*. \tag{17.4}$$

Its pivot is $\widetilde{P}_{11} = 1/(1 - |a|^2) > 0$. For $i, j \geq 2$, the Schur complement at that pivot is

$$\begin{aligned} S_{ij} &= \frac{1 - u_i\overline{u_j}}{1 - z_i\overline{z_j}} - \frac{1 - |a|^2}{(1 - z_i\overline{a})(1 - a\overline{z_j})} \\ &= B_i\overline{B_j}\frac{1 - v_i\overline{v_j}}{1 - z_i\overline{z_j}}. \end{aligned} \tag{17.5}$$

For the second equality, subtract the two fractions that do not contain $u_i\overline{u_j}$. Their difference is $B_i\overline{B_j}/(1 - z_i\overline{z_j})$; then use $u_i = B_iv_i$. If P^{red} denotes the reduced Pick matrix, this says $S = \mathrm{diag}(B_i)P^{\mathrm{red}}\,\mathrm{diag}(B_i)^*$. All diagonal factors are invertible. Consequently

$$P \geq 0 \iff P^{\mathrm{red}} \geq 0, \qquad \operatorname{rank} P = 1 + \operatorname{rank} P^{\mathrm{red}}. \tag{17.6}$$

The equivalence is asserted at an interior pivot with admissible remaining targets. In the theorem below those hypotheses come either from a solution or from the nonnegative diagonal of P.

17.4 The finite theorem, including singular cases

Theorem 17.3 (Finite Nevanlinna–Pick interpolation)**.** *For distinct $z_1, \ldots, z_n \in \mathbb{D}$ and arbitrary complex targets, a solution $f \in \mathcal{S}$ exists if and only if the matrix* (17.1) *is positive semidefinite. When a solution exists, a rational solution of degree at most $n - 1$ can be constructed. If P is singular, the solution is unique. If P is positive definite, there are infinitely many solutions.*

Proof. We prove existence and the matrix criterion simultaneously by induction on n. For one datum, $P \geq 0$ is exactly $|w_1| \leq 1$, and the constant function w_1 solves the problem.

Suppose $n > 1$. If a solution has a target $|w_j| = 1$, maximum modulus forces $f \equiv w_j$. Hence all targets agree and $P = 0$. Conversely, if $P \geq 0$ and $|w_j| = 1$, its jth diagonal entry is zero. Lemma 17.1 forces $1 - w_j\overline{w_i} = 0$ for every i, so all targets equal w_j. The constant is again a solution. Targets of modulus greater than one are excluded in either direction.

It remains to consider interior targets. Reduce at the first datum. If a solution exists, its reduced function belongs to $\mathcal{S}$ and interpolates the reduced data. Induction gives $P^{\mathrm{red}} \geq 0$, and (17.4)–(17.5) give $P \geq 0$. In the opposite direction, $P \geq 0$ gives $P^{\mathrm{red}} \geq 0$. Its diagonal entries ensure that the reduced targets have modulus at most one. By induction a reduced solution g exists; augmentation constructs f and interpolates every original row. This proves both necessity and sufficiency, without assuming an external kernel-positivity theorem.

For a rational construction, stop at a one-point problem and take its constant solution, or stop earlier at a unimodular target. If $g = N/D$ has numerator and denominator of degree at most d,

then (17.2), after multiplying by $(1-\overline{a}z)D$, has numerator and denominator of degree at most $d+1$. There are at most $n-1$ augmentations. Any common polynomial factor may be cancelled; the holomorphic range argument already excludes uncancelled poles in $\mathbb{D}$.

For uniqueness, a boundary target has already forced a constant. At an interior pivot, reduction is a bijection on solution sets. If P is singular, the rank identity makes P^{red} singular. Continue until a one-point singular matrix is reached; its target has modulus one and has a unique solution. Reversing the bijections proves uniqueness at the original level.

Finally, if $P > 0$, every reduced matrix is positive definite. The last one-point problem has an interior target b at a node a. For each constant $t \in \mathbb{D}$, the functions $m_b^{-1}(t\, m_a(z))$ solve this last problem and are distinct. Every earlier augmentation is injective, so they yield infinitely many distinct solutions of the original problem. □

If all prescribed targets are interior, every resulting solution maps into the open disk: a unimodular constant cannot take even one interior value. Singular data therefore do not force a boundary-valued final answer. A boundary constant may occur only in an intermediate reduced problem and become a nonconstant disk map after augmentation. The example $f(z) = z^2$ below exhibits exactly this mechanism.

The proof also describes all solutions, not merely one. At each interior pivot retain a free reduced Schur function, impose the remaining reduced values, and augment. At a boundary pivot retain only the forced constant. After all data have been removed, the residual Schur function is arbitrary. This representation separates the information fixed by the data from the remaining analytic freedom.

17.5 What singular data force on the whole function

The uniqueness conclusion can be sharpened without appealing to an operator-theoretic interpolation theorem. A finite Blaschke product is a function

$$B(z) = \lambda \prod_{j=1}^{d} m_{\alpha_j}(z), \qquad |\lambda| = 1, \quad |\alpha_j| < 1,$$

where zeros are counted with multiplicity and the empty product is allowed. It is rational, holomorphic near the closed disk, has modulus one on the circle, and has exactly d zeros inside. Each assertion follows directly from the factors and their nonvanishing denominators.

Proposition 17.4 (Rank and the forced rational function)**.** *If the Pick matrix is positive semidefinite and singular of rank r, its unique Schur interpolant is a finite Blaschke product of degree r.*

Proof. At an interior pivot the rank decreases by exactly one. The reduced matrix stays singular, since both its size and rank decrease by one. Eventually a boundary target is encountered: otherwise a singular one-by-one matrix with a strictly interior target would be reached, which is impossible. A boundary target forces every remaining target to equal it, so the remaining matrix is zero and the reduced solution is a unimodular constant. The number of interior reductions was therefore exactly r.

It remains to show that each reversed augmentation increases the Blaschke degree by one. Suppose the reduced function g is a finite Blaschke product of degree d. Then $q = m_a g$ has degree $d+1$ and modulus one on the circle. For an interior target b, the augmented function is $f = (b+q)/(1+\overline{b}q)$. Its denominator has modulus at least $1-|b| > 0$ on the closed disk, since $|q| \le 1$ there. It is holomorphic on a neighborhood of that disk. On the circle $|f| = 1$, by the

automorphism deficit identity. Rouché's theorem applied to $q + b$ and q gives exactly $d + 1$ zeros inside, counting multiplicity, because $|b| < |q| = 1$ on the circle.

Let those zeros be $\alpha_1, \ldots, \alpha_{d+1}$ and divide f by $\prod_j m_{\alpha_j}$. The apparent singularities at the zeros are removable with nonzero limiting values, and the quotient is holomorphic and zero-free near the closed disk. Both it and its reciprocal have boundary modulus one. The maximum principle bounds their moduli by one throughout, so the quotient has constant modulus one and is a unimodular constant. This proves the degree assertion for one augmentation, and induction proves the proposition. □

Singularity thus identifies an exact analytic form, including its boundary modulus and degree. It is not merely a label attached to a nearly zero numerical eigenvalue. If all original targets are interior, the rank is positive and the final product is nonconstant, even though its recursion ended at a boundary constant.

There is also a useful direct formula from a null vector. Choose $0 \neq c \in \ker P$. For any Schur interpolant f and any additional node z different from the old nodes, the enlarged Pick matrix is positive semidefinite by the proved necessity direction. Its off-diagonal column is $q_i = (1 - w_i\overline{f(z)})/(1 - z_i\overline{z})$. Positivity forces $c^*q = 0$: otherwise the quadratic form on the span of $(c, 0)$ and the last coordinate could be made negative by choosing the scalar multiplying c, exactly as in Lemma 17.1. Taking complex conjugates gives

$$A(z) - f(z)B(z) = 0, \qquad A(z) = \sum_i \frac{c_i}{1 - \overline{z_i}z}, \quad B(z) = \sum_i \frac{c_i\overline{w_i}}{1 - \overline{z_i}z}.$$

The rational function B cannot vanish identically. If it did, so would A. The first n Taylor coefficients of A would then give $\sum_i c_i\overline{z_i}^k = 0$ for $0 \leq k < n$. The corresponding Vandermonde determinant is $\prod_{i<j}(\overline{z_j} - \overline{z_i}) \neq 0$, contradicting $c \neq 0$. Therefore $f = A/B$ wherever $B \neq 0$. At a zero of B the already proved existence of a holomorphic interpolant shows that the quotient's singularity is removable. This formula supplies another route to uniqueness, while the recursive proof supplies the exact degree and the admissible factorization.

17.6 One value and a prescribed derivative

Derivative data were excluded from the finite value matrix. The simplest derivative problem can nevertheless be solved directly by the same reduction, making its altered constraints visible. Fix $a, b \in \mathbb{D}$ and require $f(a) = b$, $f'(a) = c$. Expanding at a gives

$$m_a(z) = \frac{z - a}{1 - |a|^2} + O((z - a)^2), \qquad m_b(f(z)) = \frac{c(z - a)}{1 - |b|^2} + O((z - a)^2).$$

Consequently the reduced Schur function must have

$$g(a) = \frac{1 - |a|^2}{1 - |b|^2}c.$$

This target lies in the closed disk if and only if $|c| \leq (1 - |b|^2)/(1 - |a|^2)$. Necessity follows from the Schur bound on g. For sufficiency choose g equal to this constant and augment; the Taylor expansions verify the derivative as well as the value. Equality forces g to be a unimodular constant, so the solution is the unique corresponding disk automorphism. For strict inequality there are infinitely many choices of g with the prescribed value, and injectivity of augmentation preserves their distinctness.

Investigation 17.5 (The first two Taylor coefficients)**.** For a Schur function with $f(0) = 0$ and $f'(0) = d$, determine exactly which values of $f''(0)$ are possible.

Solution. Write $f(z) = zg(z)$ with $g \in \mathcal{S}$. Then $g(0) = d$ and $f''(0) = 2g'(0)$. If $|d| < 1$, the derivative calculation gives $|f''(0)| \le 2(1 - |d|^2)$. Every such value is attained by

$$f(z) = z\frac{d + tz}{1 + \overline{d}\,tz}, \qquad |t| \le 1,$$

because its second derivative at zero is $2(1 - |d|^2)t$ and its denominator is nonzero on $\mathbb{D}$. If $|d| = 1$, maximum modulus forces $g \equiv d$, hence $f(z) = dz$ and $f''(0) = 0$. If $|d| > 1$, there is no solution. These cases exhaust the data, including the boundary case where the fractional construction would no longer have an interior pivot.

17.7 Worked three-point investigations

Example 17.6 (Pairwise tests can all pass)**.** Take nodes $0, 1/2, -1/2$ with targets $0, 1/2, 1/2$. Every pair satisfies the Schwarz–Pick distance condition. The pairs involving zero have equal source and target distance, while the two nonzero nodes have the same target. Nevertheless no interpolant exists. Equality for the pair $0, 1/2$ forces $f(z) = z$, which fails at $-1/2$.

The Pick matrix displays the same obstruction:

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 3/5 \\ 1 & 3/5 & 1 \end{pmatrix}, \qquad (-2, 1, 1)P(-2, 1, 1)^{\mathsf{T}} = -4/5.$$

Its two-by-two principal tests do not replace positivity of the whole matrix. The example is small enough to show the distinction directly, without relying on numerical eigenvalues.

Example 17.7 (An exact symmetric family)**.** Let $0 < r < 1$ and prescribe

$$f(0) = 0, \qquad f(r) = f(-r) = b.$$

Reduction at zero asks for a Schur function g with $g(r) = b/r$ and $g(-r) = -b/r$. If $b = 0$, the zero function works. Otherwise rotate the range so b/r is nonnegative real and set $k = |b|/r$. The diagonal conditions first give $k \le 1$. The two-point criterion for g is

$$\frac{2k}{1 + k^2} \le \frac{2r}{1 + r^2}.$$

On $[0, 1]$ the function $t \mapsto 2t/(1 + t^2)$ is strictly increasing except for its zero derivative at the endpoint; alternatively its difference at r, k has the sign of $(r - k)(1 - rk)$. The condition is therefore $k \le r$, or $|b| \le r^2$.

For every such b, an explicit solution is $f(z) = (b/r^2)z^2$. When $|b| = r^2$, the reduced two-point problem has equality in Schwarz–Pick, and the solution is unique. For $r = 1/2$ and $b = 1/4$, it is z^2. Here

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 5/4 & 3/4 \\ 1 & 3/4 & 5/4 \end{pmatrix} = \left(1 + z_i\overline{z_j}\right)_{i,j=1}^{3}.$$

This is the Gram matrix of the vectors $(1, z_i)$, so it is positive semidefinite of rank two. The vector $(-2, 1, 1)$ is now a null vector, not a negative direction. The singular criterion distinguishes a forced analytic solution from inconsistent data.

17.8 Arithmetic work, conditioning and data perturbations

The recursive construction updates $n-1$ targets at its first stage, $n-2$ at the next, and so on. Each update in (17.3) uses a bounded number of complex arithmetic operations. Thus the complete scalar recursion uses $O(n^2)$ field operations, with at most n stages. It can store the active target vector and the saved pivot parameters in $O(n)$ complex numbers. Evaluating the resulting nested fractional formula at one point uses $O(n)$ field operations. Expanding a numerator and denominator is a separate output choice and is unnecessary for evaluating the function.

These are arithmetic-operation counts. For Gaussian-rational data, the updates are exact rational operations, but their bit cost depends on the sizes of intermediate numerators and denominators. No bound independent of those sizes follows from $O(n^2)$. For floating-point data, the divisions by $m_a(z_i)$ and by $1-\bar{b}w_i$ can magnify errors when nodes coalesce or a pivot approaches the unit circle. Reordering nodes changes the arithmetic path, while leaving the theorem unchanged. A small numerical pivot is not evidence that an exact singular case has occurred.

There is a direct stability statement for strictly feasible data. Suppose $|z_i| \le r < 1$, $|w_i| \le 1$ and $P \ge \delta I$ with $\delta > 0$. Change targets to $\widehat{w}_i = w_i + e_i$, where $|e_i| \le \eta$. Then

$$|\widehat{P}_{ij} - P_{ij}| \le \frac{2\eta + \eta^2}{1 - r^2} =: M.$$

Indeed, the numerator changes by $-e_i\overline{w_j} - w_i\overline{e_j} - e_i\overline{e_j}$, and the source denominator has modulus at least $1-r^2$. For every ξ, Cauchy–Schwarz gives

$$|\xi^*(\widehat{P} - P)\xi| \le M\Big(\sum_i |\xi_i|\Big)^2 \le nM\,\|\xi\|_2^2.$$

Consequently $\delta > nM$ guarantees $\widehat{P} > 0$ and hence interpolation of every target table in that uncertainty set. Its nonnegative diagonal also guarantees admissibility of the perturbed targets. This sufficient margin need not be sharp. At singular data there is no positive δ to use; Example 17.7 shows why arbitrarily small outward changes can destroy feasibility.

17.9 The exact range of an unprescribed value

Interpolation is also a way to describe uncertainty. Suppose $P > 0$ and choose a point $\zeta \in \mathbb{D}$ different from every interpolation node. What values can $f(\zeta)$ take among all Schur interpolants? The reduction proof answers this question with a Euclidean disk, without selecting one preferred interpolant.

Perform all n interior reductions. Let a_j be the node removed at stage j and $b_j \in \mathbb{D}$ its target at that stage. The last reduced problem has no data and an arbitrary Schur function h. At the fixed point ζ, augmentation is represented by the matrix

$$M_j(\zeta) = \begin{pmatrix} m_{a_j}(\zeta) & b_j \\ \overline{b_j} m_{a_j}(\zeta) & 1 \end{pmatrix}.$$

With the order determined by successive substitutions, put

$$M_1(\zeta)\cdots M_n(\zeta) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}.$$

Then all permitted values are

$$f(\zeta) = \frac{At + B}{Ct + D}, \qquad |t| \le 1. \tag{17.7}$$

Indeed, any reduced function gives $t = h(\zeta)$ in the closed disk. Conversely, every such t is supplied by the constant reduced function $h \equiv t$, and all augmentations are admissible. Thus this is an exact range description, not merely an outer bound.

Each determinant is $m_{a_j}(\zeta)(1 - |b_j|^2) \neq 0$, so $AD - BC \neq 0$. Each individual denominator is nonzero even for $|t| = 1$, because $|m_{a_j}(\zeta)| < 1$. Substituting backwards shows that the final denominator $Ct + D$ has no zero on the closed disk. Consequently $|D| > |C|$: if $C \neq 0$, its only possible zero is $-D/C$ and must have modulus greater than one; if $C = 0$, invertibility gives $D \neq 0$. Write $k = |D|^2 - |C|^2 > 0$.

Solving (17.7) for t shows that the output set is the disk

$$\left| w - \frac{B\overline{D} - A\overline{C}}{k} \right| \leq \frac{|AD - BC|}{k}. \tag{17.8}$$

For completeness, $|t| \leq 1$ becomes $|Dw - B|^2 \leq |A - Cw|^2$. Expanding and completing the square gives (17.8); the constant term simplifies using

$$|B\overline{D} - A\overline{C}|^2 - k(|B|^2 - |A|^2) = |AD - BC|^2.$$

The point where the inverse fractional formula could fail does not create an extra solution: if $A - Cw = 0$, the inequality also forces $Dw - B = 0$, contradicting the nonzero determinant. This checks the exceptional denominator explicitly.

As a simple example, prescribe $f(0) = f(r) = 0$ for $0 < r < 1$. Two reductions give $f(z) = z\, m_r(z) h(z)$, so the exact range at ζ is the centered disk of radius $|\zeta m_r(\zeta)|$. At $\zeta = -r$, this radius is $2r^2/(1 + r^2)$. Every point of that disk is attained by choosing a constant h with the corresponding modulus and phase. The uncertainty has a precise shape even though the function is not unique.

For singular positive semidefinite data there is only one interpolant, so the range at an unprescribed point is a singleton. At a prescribed node it is a singleton in either case. These cases are handled directly; the derivation above deliberately assumed invertible matrices at ζ. This distinction is useful when finite interpolation is used for design envelopes: a range of possible responses and a unique forced response are qualitatively different outputs.

17.10 Solved exercises

Investigation 17.8 (One boundary target)**.** At distinct interior nodes prescribe $w_1 = 1$ and $w_2 = 1 - \varepsilon$, where $0 < \varepsilon < 1$. Explain why no Schur interpolant exists even though both targets lie in the closed disk.

Solution. The first value forces the constant function 1 by maximum modulus. In matrix language, $P_{11} = 0$, but $P_{12} = \varepsilon/(1 - z_1\overline{z_2}) \neq 0$, contradicting Lemma 17.1. The contradiction involves the relation between two rows, not the admissibility of either row separately.

Investigation 17.9 (A unique rational interpolant)**.** Let $a, b \in \mathbb{D}$ and $|\lambda| = 1$. Prescribe the values of $F(z) = m_b^{-1}(\lambda m_a(z))$ at two distinct nodes. Show directly that F is the unique disk-valued interpolant of those values.

Solution. Both normalizing automorphisms preserve pseudohyperbolic distance, as does the rotation. The prescribed pair therefore has equality in Schwarz–Pick. Any interpolant is an automorphism and, after normalization at the first pair of source and target points, is a rotation. The second datum fixes that rotation uniquely. The displayed F already satisfies both data, so it is that unique function. Equivalently, its two-point Pick matrix is positive semidefinite and singular.

Investigation 17.10 (Why a finite table is not a frequency grid proof)**.** Explain precisely what is and is not implied by a positive Pick matrix for finitely many measured complex response values.

Solution. For exact nodes and targets it proves the existence of a Schur function with those values, and the recursion constructs one on the entire disk. It does not prove that an unknown device has that function as its response, that the measurements are exact, or that a chosen circuit topology realizes the interpolant. A measurement uncertainty set can be treated with a separate robustness estimate such as the preceding margin. The analytic existence theorem and the identification of a physical model are different mathematical questions.

Chapter 18

Passive systems, matching and analytic sensitivity

The right half-plane is a natural parameter domain for many linear systems, while the disk is a natural range for reflection coefficients. A Cayley transformation connects the two. This connection makes hyperbolic contraction a statement about the sensitivity of a response function and turns interpolation into a finite feasibility condition for positive-real models.

We use scalar, linear, time-invariant models and distinguish a complex Laplace variable s from a real angular frequency ω. The latter appears through boundary limits $s \to i\omega$ only when those limits exist. Reference impedances are real and positive. The transmission-line conventions are those of Haus and Melcher [8]; the inequalities and examples below are derived explicitly from the stated models.

18.1 Positive real part and the reflection coordinate

Put $\mathbb{H}_R = \{s : \operatorname{Re} s > 0\}$. In this chapter a strictly positive-real analytic model is a function Z holomorphic on $\mathbb{H}_R$ with $\operatorname{Re} Z(s) > 0$ there. When it represents a real time-domain scalar system, we additionally require $Z(\overline{s}) = \overline{Z(s)}$. This symmetry is independent of the positivity inequality. We do not assume that an arbitrary function in this class has already been realized by a finite network.

Let $Z_0 > 0$ be a reference impedance. The reflection coordinate and its inverse are

$$\Gamma = \frac{Z - Z_0}{Z + Z_0}, \qquad Z = Z_0 \frac{1 + \Gamma}{1 - \Gamma}. \tag{18.1}$$

For $\operatorname{Re} Z > 0$, the first denominator is nonzero and

$$1 - |\Gamma|^2 = \frac{4 Z_0 \operatorname{Re} Z}{|Z + Z_0|^2} > 0. \tag{18.2}$$

Conversely, if $|\Gamma| < 1$, then $1 - \Gamma \neq 0$ and

$$\operatorname{Re}\left(Z_0 \frac{1 + \Gamma}{1 - \Gamma} \right) = Z_0 \frac{1 - |\Gamma|^2}{|1 - \Gamma|^2} > 0.$$

Thus the two coordinate systems express exactly the same strict range condition. These formulas also show why a lossless boundary value, with zero real part, belongs on the unit circle rather than inside it.

For a lossless reference line use voltage-wave phasors V_+, V_- and set $V = V_+ + V_-$, $I = (V_+ - V_-)/Z_0$. If $V_+ \neq 0$ and $V = ZI$, elimination gives $V_-/V_+ = \Gamma$. With peak-amplitude

phasors,

$$P_+ = \frac{|V_+|^2}{2Z_0}, \qquad P_- = \frac{|V_-|^2}{2Z_0}, \qquad \frac{1}{2}\operatorname{Re}(V\bar{I}) = P_+ - P_-.$$

To verify the last equation, expand $(V_+ + V_-)(\overline{V_+} - \overline{V_-})$; its cross terms are purely imaginary. The load law also gives average absorbed power $\operatorname{Re} Z\,|I|^2/2$. Therefore $1 - |\Gamma|^2$ is the absorbed fraction in this ideal steady-state model. For root-mean-square phasors the factors $1/2$ disappear from both powers, leaving the fraction unchanged.

18.2 Hyperbolic sensitivity on the half-plane

The curvature -1 density on $\mathbb{H}_R$ is $|\mathrm{d}s|/\operatorname{Re} s$. For a positive scale σ with the same units as s, the map $\chi_\sigma(s) = (s-\sigma)/(s+\sigma)$ carries $\mathbb{H}_R$ to $\mathbb{D}$. Indeed,

$$1 - |\chi_\sigma(s)|^2 = \frac{4\sigma \operatorname{Re} s}{|s+\sigma|^2}, \qquad |\chi_\sigma'(s)| = \frac{2\sigma}{|s+\sigma|^2}.$$

Pulling back $2|\mathrm{d}z|/(1-|z|^2)$ gives the asserted density. The scale σ disappears from the metric, as it should.

Proposition 18.1 (Response sensitivity)**.** *For a strictly positive-real analytic Z and its reflection Γ,*

$$\frac{|Z'(s)|}{\operatorname{Re} Z(s)} = \frac{2|\Gamma'(s)|}{1-|\Gamma(s)|^2} \le \frac{1}{\operatorname{Re} s} \qquad (s \in \mathbb{H}_R). \tag{18.3}$$

For any $s, t \in \mathbb{H}_R$,

$$\left|\frac{Z(s) - Z(t)}{Z(s) + \overline{Z(t)}}\right| \le \left|\frac{s-t}{s+\bar{t}}\right|. \tag{18.4}$$

Proof. Schwarz–Pick applied after the source Cayley transformation gives the inequality in (18.3). Differentiate (18.1) to obtain $\Gamma' = 2Z_0 Z'/(Z+Z_0)^2$ and substitute (18.2); this gives its equality. For the two-point form, the pseudohyperbolic distance transported from the disk to the right half-plane is $|(u-v)/(u+\overline{v})|$. Substitution of the Cayley formulas cancels the positive scale factors. Applying the two-point Schwarz–Pick inequality then proves (18.4). □

If s is measured in inverse seconds and Z in ohms, Z' has units ohm-seconds. Each side of (18.3) has units of time. The two-point ratios are dimensionless; numerator and denominator in each ratio have the same units. Setting $s = i\omega$ directly in the derivative bound would replace its right side by infinity. It yields no finite universal bound on boundary-frequency derivatives. A boundary estimate needs additional hypotheses, such as continuation through an interval or a quantitative loss margin.

At a fixed operating point, an impedance perturbation has an equally useful geometric meaning. If $\operatorname{Re} Z > 0$ and $|\Delta Z| \le \varepsilon \operatorname{Re} Z$ with $0 \le \varepsilon < 1$, then $\operatorname{Re}(Z + \Delta Z) > 0$ and

$$\left|\frac{\Delta Z}{2\operatorname{Re} Z + \Delta Z}\right| \le \frac{\varepsilon}{2-\varepsilon}.$$

This follows from the reverse triangle inequality in the denominator. The hyperbolic distance between the two impedances is consequently at most $2\operatorname{artanh}(\varepsilon/(2-\varepsilon))$. The reference impedance does not appear: changing the reflection coordinate does not change this intrinsic distance.

18.3 Finite matching constraints in impedance coordinates

Suppose distinct complex sampling parameters $s_i \in \mathbb{H}_R$ are assigned impedances Z_i with positive real part. The existence of a strictly positive-real analytic interpolant is equivalent to

$$K = \left(\frac{Z_i + \overline{Z_j}}{s_i + \overline{s_j}} \right)_{i,j=1}^{n} \geq 0. \tag{18.5}$$

To prove this, set $z_i = \chi_\sigma(s_i)$ and $w_i = (Z_i - Z_0)/(Z_i + Z_0)$. Direct subtraction gives

$$P_{ij} = \frac{Z_0}{\sigma} \frac{s_i + \sigma}{Z_i + Z_0} K_{ij} \frac{\overline{s_j} + \sigma}{\overline{Z_j} + Z_0}.$$

This is an invertible diagonal congruence times a positive scalar. Theorem 17.3 constructs the Schur interpolant. Because at least one target is interior, it is disk-valued; the inverse range Cayley transform then constructs Z with strictly positive real part. Necessity follows by the same transformations.

If data occur in conjugate pairs with conjugate target values, the interpolant can also be chosen to satisfy the real-system symmetry. Given one disk interpolant f, its reflected function $f^{\#}(z) = \overline{f(\overline{z})}$ interpolates the same table. The average $(f + f^{\#})/2$ is holomorphic, remains in the disk by convexity, and is invariant under reflection. Real positive source and range scales preserve the symmetry when transforming back. A rational interpolant remains rational under this operation, although the degree bound for this symmetrized construction may increase.

The criterion concerns existence in the analytic positive-real class. It does not impose a specified number of inductors or capacitors, a prescribed circuit graph, an asymptotic order or a separate realization theorem. These additional design constraints cannot be read from $K \geq 0$ alone. Likewise, samples on the imaginary axis are boundary interpolation data, not the interior data of (18.5).

Example 18.2 (An impossible two-point specification)**.** Choose a time scale $\tau > 0$ and an impedance scale $R_0 > 0$. Request $Z(1/\tau) = R_0$ and $Z(2/\tau) = 4R_0$. The target half-plane distance ratio is $3/5$, whereas the source ratio is $1/3$. Inequality (18.4) forbids the specification. Both requested impedances are positive real numbers; pointwise positivity alone does not make the whole table analytically feasible.

In contrast, $Z(1/\tau) = R_0$ and $Z(2/\tau) = 2R_0$ have equality in the distance condition. The model $Z(s) = R_0\tau s$ supplies an interpolant, with the units of an ideal inductance $L = R_0\tau$. The two-point equality makes this half-plane interpolant unique.

18.4 Propagation along an ideal lossless line

Suppose a uniform lossless line of length $\ell \geq 0$ has constant propagation speed $v_p > 0$ and real characteristic impedance Z_0. Choose the Laplace convention in which a time delay t_0 has transfer factor e^{-st_0}. A wave sent toward the load and reflected back has round-trip delay $\tau = 2\ell/v_p$, so its input reflection is

$$\Gamma_{\mathrm{in}}(s) = e^{-\tau s}\Gamma_L(s). \tag{18.6}$$

This formula follows by multiplying the forward and return propagation factors by the load reflection. It is the ideal distributed-line model; the exponential need not be replaced by a finite rational function.

For $\operatorname{Re} s > 0$ and $\ell > 0$, $|e^{-\tau s}| = e^{-\tau \operatorname{Re} s} < 1$. A Schur-valued load therefore remains Schur-valued, and the inverse Cayley transform gives an input impedance with positive real part. At a fixed such s, changing the load reflection acts through a rotation followed by a radial contraction

of factor $e^{-\tau \operatorname{Re} s}$. Its hyperbolic Lipschitz factor is at most that number, by the disk contraction estimate proved in Proposition 19.1 below.

On the imaginary axis, whenever the load boundary value exists,

$$\Gamma_{\mathrm{in}}(i\omega) = e^{-i\tau\omega}\Gamma_L(i\omega), \qquad |\Gamma_{\mathrm{in}}(i\omega)| = |\Gamma_L(i\omega)|.$$

For each fixed frequency this is a disk rotation and hence a hyperbolic isometry of the load-reflection plane. The frequency dependence of the rotation does not make it one fixed isometry when frequencies themselves are varied. The modulus identity proves an engineering limitation of this specific operation: adding length of the same ideal lossless line alone cannot improve the reflected power fraction at a chosen frequency. It can change the phase and the impedance observed at the input.

For instance, at a frequency where $\tau\omega = \pi$, the reflection changes sign. If $\zeta = Z_L/Z_0$ denotes normalized load impedance, then

$$\frac{Z_{\mathrm{in}}}{Z_0} = \frac{1-\Gamma_L}{1+\Gamma_L} = \frac{1}{\zeta}.$$

A normalized resistance 2 therefore becomes $1/2$, while the reflection modulus remains $1/3$. A quarter-wavelength line has exactly this round-trip phase. More elaborate matching networks can change the effective load by combining branches or altering characteristic impedance; they are not the single operation analyzed here.

The strict attenuation for $\operatorname{Re} s > 0$ is also not a claim that the ideal line dissipates energy. A positive real part of the Laplace variable includes an exponential time weight, and a delay changes that weighted representation. The boundary-frequency calculation shows the lossless power relation directly. This distinction prevents a geometric contraction in an analytic parameter domain from being misread as heat production in a physical component.

18.5 Interior matching zeros and exact analytic constraints

Suppose a Schur reflection Γ has zeros at distinct points $s_1, \dots, s_k \in \mathbb{H}_R$. In disk coordinates $a_j = \chi_\sigma(s_j)$, successive reductions at zero target values give

$$\Gamma(\chi_\sigma^{-1}(z)) = \prod_{j=1}^{k} m_{a_j}(z)\, h(z), \qquad h \in \mathcal{S}.$$

To justify each step, divide a Schur function vanishing at a_j by m_{a_j}. Schwarz's lemma after normalization proves that the quotient is Schur, and the zero is removable. At every other distinct zero the factor being divided out is nonzero, so the remaining zero conditions persist. Repeating the argument proves the product formula. Prescribed higher multiplicities are treated by repeating a factor; the order of vanishing decreases by one at each division.

Taking absolute values and transporting back to the half-plane yields

$$|\Gamma(s)| \le \prod_{j=1}^{k} \left| \frac{s - s_j}{s + \overline{s_j}} \right|. \tag{18.7}$$

This inequality constrains an additional interior response value. For example, two exact matches at $s_1 = 1/\tau_0$ and $s_2 = 2/\tau_0$ imply, at $s = 3/\tau_0$,

$$|\Gamma(3/\tau_0)| \le \frac{1}{2}\frac{1}{5} = \frac{1}{10}.$$

A request for modulus $1/4$ there is inconsistent with the two matches in this analytic class. Choosing a unimodular constant h attains the upper bound at every nonzero point in the product,

so the inequality is sharp as a Schur-class statement. Real positive nodes admit a real choice of phase and hence a real-system symmetry when desired.

The matching points in this argument lie in the open right half-plane. They are complex Laplace parameters, not ordinary steady-state frequencies. Boundary matching is a different question. One elementary statement is available with an explicit continuation hypothesis: if Γ extends holomorphically through an open segment of the imaginary axis and vanishes at every point of that segment, then it is identically zero. The segment has an accumulation point inside the extended domain, so the identity theorem applies, first locally and then on the connected half-plane. In that case $Z \equiv Z_0$.

Without continuation, this particular identity-theorem argument cannot be used: the zeros would lie on the boundary of the original domain. And a small nonzero reflection on a frequency band is not the same as an identically zero reflection there. Quantitative broadband matching limits require additional assumptions on growth, network structure or boundary integrals. The present finite interpolation and zero-factor arguments state their own exact constraints without asserting such a broader synthesis theorem.

18.6 The ideal series-RC family

Take $R, C > 0$ in the lumped-element model and set

$$Z(s) = R + \frac{1}{Cs}, \qquad \Gamma(s) = \frac{1 + (R - Z_0)Cs}{1 + (R + Z_0)Cs}. \tag{18.8}$$

For $\operatorname{Re} s > 0$, $\operatorname{Re}(1/s) = \operatorname{Re} s/|s|^2 > 0$, so this is a strictly positive-real model. The only pole of the displayed reflection is at $-1/[C(R + Z_0)]$, in the left half-plane, and

$$\Gamma'(s) = \frac{-2Z_0C}{[1 + (R + Z_0)Cs]^2}.$$

The impedance pole at zero and the reflection's finite limiting value there are not contradictory: the Cayley coordinate can remain finite while the impedance tends to infinity.

For real $\omega \neq 0$, put $A = (R - Z_0)^2$, $B = (R + Z_0)^2$. Then

$$Q(\omega) := |\Gamma(i\omega)|^2 = \frac{1 + AC^2\omega^2}{1 + BC^2\omega^2}, \qquad 1 - Q(\omega) = \frac{4RZ_0C^2\omega^2}{1 + BC^2\omega^2}. \tag{18.9}$$

The limiting extension is $Q(0) = 1$. The original impedance at zero is not assigned a finite value. Differentiation and subtraction give

$$Q'(\omega) = \frac{-8RZ_0C^2\omega}{(1 + BC^2\omega^2)^2} < 0 \quad (\omega > 0), \qquad Q(\omega) - \frac{A}{B} = \frac{B - A}{B(1 + BC^2\omega^2)} > 0.$$

Thus A/B is the sharp lower limit on the positive-frequency axis and is not attained at a finite frequency.

Proposition 18.3 (Exact matching threshold)**.** *Fix $0 < q < 1$. A finite frequency satisfies $|\Gamma(i\omega)| \leq q$ if and only if $q^2B > A$ and*

$$\omega^2 \geq \frac{1 - q^2}{C^2(q^2B - A)}. \tag{18.10}$$

If $q^2B \leq A$, no finite frequency satisfies the bound.

Proof. The denominator in (18.9) is positive. Multiplying $Q \le q^2$ by it gives $1-q^2 \le C^2\omega^2(q^2B-A)$. Its left side is positive. A nonpositive coefficient on the right is therefore impossible; for a positive coefficient division gives exactly (18.10). □

For $R = Z_0$, introduce the dimensionless frequency $u = RC\omega$. Then $Q = 1/(1+4u^2)$, and $q = 1/2$ is equivalent to $|u| \ge \sqrt{3}/2$. The quantity RC has units of time, and $1/(RC)$ sets the angular-frequency scale. Dividing by 2π converts an angular frequency in radians per second to cycles per second. The unlimited-frequency limit is a property of the ideal model; parasitic elements and distributed effects can change the model before that limit is approached in a device.

18.7 The RC law in time and its stored energy

The analytic positivity of the RC impedance has a direct time-domain account under the same ideal component assumptions. Let $q(t)$ be capacitor charge, $i(t) = q'(t)$ the real current, and $v(t)$ the terminal voltage. For an ideal series resistor and capacitor the constitutive equations are

$$v(t) = Ri(t) + \frac{q(t)}{C}, \qquad R, C > 0.$$

Assume q is continuously differentiable on the finite time interval under consideration. Its stored electrostatic energy is $\mathcal{E}(t) = q(t)^2/(2C)$, and differentiation gives the exact power balance

$$v(t)i(t) = Ri(t)^2 + \mathcal{E}'(t). \tag{18.11}$$

Integrating between t_0 and t_1 yields

$$\int_{t_0}^{t_1} vi\,dt = R\int_{t_0}^{t_1} i^2\,dt + \mathcal{E}(t_1) - \mathcal{E}(t_0) \ge -\mathcal{E}(t_0).$$

With zero initial charge, the net supplied energy is nonnegative. With nonzero initial charge, the circuit can return previously stored energy; a negative terminal-power interval is then compatible with passivity. The inequality specifies the storage allowance that would be lost by the inaccurate assertion that passive systems must always have nonnegative instantaneous terminal power.

For a concrete discharge, close the capacitor through the resistor with no applied terminal voltage, $v = 0$. The equation becomes $q' = -q/(RC)$, so $q(t) = q_0e^{-t/(RC)}$ for $t \ge 0$ and $\mathcal{E}(t) = q_0^2e^{-2t/(RC)}/(2C)$. Integration gives

$$R\int_0^T i(t)^2\,dt = \frac{q_0^2}{2C}(1 - e^{-2T/(RC)}).$$

The dissipated energy equals the loss of stored energy exactly. The units also agree: charge squared divided by capacitance is energy, and RC is a time. These are model identities, without a claim that an actual component has been measured over an unlimited time or frequency range.

Now assume Laplace transforms exist in a right half-plane and write $Q(s), I(s), V(s)$ for the transforms starting at time zero. Integration by parts gives $I(s) = sQ(s) - q(0)$, including the initial-state term. Thus

$$V(s) = \left(R + \frac{1}{Cs}\right) I(s) + \frac{q(0)}{Cs}.$$

The impedance $Z = R + 1/(Cs)$ is the zero-initial-state input-output multiplier. When the initial charge is nonzero, it remains the same multiplier but there is also an independent response term. Dividing the full V by I without controlling that term need not recover the stated impedance. Transfer-function conventions and stored energy must therefore be specified together when translating between the analytic and time-domain pictures.

For a periodic steady-state solution, $\mathcal{E}$ has equal values at the endpoints of a period. The averaged storage derivative is zero, so the average terminal power equals the resistor's average dissipation. For a nonzero sinusoidal frequency and peak current phasor I this is $R|I|^2/2$, in agreement with $\operatorname{Re} Z(i\omega) = R$. This derivation uses the stipulated periodic state; it does not discard an arbitrary transient by algebra.

18.8 Series and parallel composition within the analytic class

Some preservation rules can be proved without a general network realization theorem. If Z_1, Z_2 are strictly positive-real analytic functions on $\mathbb{H}_R$, then their series sum has positive real part simply because real parts add. Neither impedance vanishes there. For any complex number z with positive real part,

$$\operatorname{Re} \frac{1}{z} = \frac{\operatorname{Re} z}{|z|^2} > 0.$$

Hence each admittance $Y_j = 1/Z_j$ is holomorphic and strictly positive real, their sum is nonzero with positive real part, and the parallel impedance $(Y_1 + Y_2)^{-1}$ has the same properties. These arguments prove that all the denominators introduced by the compositions are nonzero in the open right half-plane. They also preserve $Z(\overline{s}) = \overline{Z(s)}$ when that symmetry holds initially.

The ideal elements R, Ls and $1/(Cs)$, with positive real parameters, each satisfy strict positivity on this open domain. Finite series-parallel combinations therefore belong to the class by induction. The induction is only a sufficient construction: it does not show that every positive-real function can be realized by such a combination or by a finite circuit at all.

Boundary behavior remains separate. An ideal inductor and an ideal capacitor have purely imaginary impedance at a nonzero real frequency, despite strict positivity when $\operatorname{Re} s > 0$. In their parallel combination the boundary admittances can cancel at a resonant frequency, creating an infinite boundary impedance. This does not contradict the nonvanishing sum proved in the open half-plane. For example,

$$\left(\frac{1}{Ls} + Cs\right)^{-1} = \frac{Ls}{1 + LCs^2}$$

has poles at $s = \pm i/\sqrt{LC}$ and none in $\mathbb{H}_R$. The example distinguishes an interior analytic guarantee from a claim of bounded response all the way to the boundary.

18.9 An exact tolerance calculation

Suppose independently

$$0 < R_- \le R \le R_+, \qquad 0 < C_- \le C \le C_+,$$

with fixed $Z_0 > 0$, and ask for $|\Gamma(i\omega)| \le q$ for every component choice and every $|\omega| \ge \omega_0 > 0$. Define

$$F(R) = q^2(R + Z_0)^2 - (R - Z_0)^2.$$

The desired statement is equivalent to

$$C_-^2 \omega_0^2 \min\{F(R_-), F(R_+)\} \ge 1 - q^2. \tag{18.12}$$

Here is the full argument. The pointwise condition is $C^2\omega^2 F(R) \ge 1 - q^2$. If it holds uniformly, it holds at $C = C_-$, $|\omega| = \omega_0$ and both resistance endpoints, proving necessity. Conversely, (18.12) makes both endpoint values of F positive. The coefficient of R^2 in F is $q^2 - 1 < 0$, so F is

concave. A concave function on an interval is at least the smaller of its endpoint values: write an interior point as a convex combination and use the defining concavity inequality. Hence $F(R)$ remains positive and has that lower bound everywhere. Increasing $C^2\omega^2$ from $C_-^2\omega_0^2$ can only increase the product, which proves sufficiency.

For a concrete normalized design take $q = 1/2$, $R_- = 9Z_0/10$, $R_+ = 11Z_0/10$, and $C_- = 9C_0/10$, $C_+ = 11C_0/10$. The two endpoint factors are

$$F(R_-) = \frac{357}{400}Z_0^2, \qquad F(R_+) = \frac{437}{400}Z_0^2.$$

Writing $u_0 = Z_0C_0\omega_0$, condition (18.12) becomes

$$\frac{28917}{40000}u_0^2 \geq \frac{3}{4}, \qquad \text{or} \qquad u_0^2 \geq \frac{10000}{9639}.$$

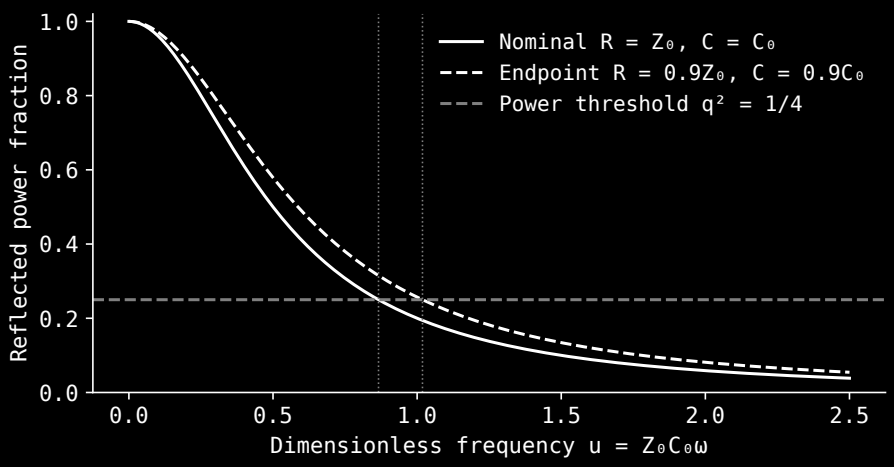


Figure 18.1: The nominal RC power curve and the component endpoint controlling the stated tolerance threshold. The dotted frequencies are the exact nominal and robust thresholds. The endpoint curve is not asserted to be the worst response at every frequency.

This is a uniform statement over a continuous component rectangle and an unbounded model frequency band. It was obtained from inequalities and concavity, rather than a finite parameter sweep. Correlations between component errors can make the rectangular uncertainty set conservative; the assertion remains exactly about the rectangle that was specified.

18.10 Solved exercises

Investigation 18.4 (The geometry of a matching specification)**.** In normalized impedance coordinates $Z/Z_0 = u + iv$, show that $|\Gamma| \leq q < 1$ is a Euclidean disk strictly contained in $u > 0$.

Solution. The inequality is $(u-1)^2 + v^2 \leq q^2((u+1)^2 + v^2)$. Divide by $1 - q^2$ and complete the square to obtain

$$\left(u - \frac{1+q^2}{1-q^2}\right)^2 + v^2 \leq \left(\frac{2q}{1-q^2}\right)^2.$$

Its leftmost real coordinate is $(1-q)/(1+q) > 0$. This is a geometric way to read a reflection tolerance, including its allowed reactive part.

Investigation 18.5 (A change of reference impedance)**.** Let $Z_0, Z_1 > 0$ and form the two reflection coordinates of the same impedance. Express Γ_1 in terms of Γ_0 and explain what happens to their hyperbolic distances.

Solution. Substitute $Z = Z_0(1+\Gamma_0)/(1-\Gamma_0)$ into the second Cayley formula. With $a = (Z_1 - Z_0)/(Z_1 + Z_0) \in (-1, 1)$, the result is $\Gamma_1 = (\Gamma_0 - a)/(1 - a\Gamma_0) = m_a(\Gamma_0)$. This is a disk automorphism, so every hyperbolic distance is unchanged. Euclidean radii and the numerical reflection at a given point can change; intrinsic separation does not.

Investigation 18.6 (What a passive model does not determine)**.** Does $\operatorname{Re} Z(s) > 0$ on $\mathbb{H}_R$ imply that Z is bounded there, or that the ideal law is valid at arbitrarily high physical frequencies?

Solution. No. The ideal inductor $Z(s) = Ls$, $L > 0$, is positive real but unbounded. The ideal capacitor $1/(Cs)$ is unbounded near zero. These examples satisfy the analytic range condition while having very different growth. Physical validity is an assumption about a device and a frequency regime; it is not a consequence of holomorphy. In particular, the positive-real condition is not by itself a claim of bounded-input bounded-output stability for the impedance regarded as an arbitrary transfer function.

Chapter 19

Controlled iteration and computational accuracy

Schwarz–Pick gives a nonexpansion for every disk map, but a numerical iteration needs more than nonexpansion. To obtain a uniform rate, one must identify a genuine contraction factor. A useful hypothesis is that the entire image stays in a smaller disk. It gives convergence, a stopping rule and a perturbation estimate in a single geometric framework. The results below concern an exact mathematical iteration and explicitly bounded evaluation errors. They do not infer convergence from a plotted sequence or assign a timing to an unexecuted algorithm.

19.1 An interior image produces a uniform factor

Proposition 19.1 (A global contraction)**.** *Suppose $f : \mathbb{D} \to \mathbb{D}$ is holomorphic and $|f(z)| \le r < 1$ for all z, with $0 < r < 1$. Then*

$$\frac{(1-|z|^2)|f'(z)|}{1-|f(z)|^2} \le r, \qquad d_{\mathbb{D}}(f(z), f(w)) \le r\, d_{\mathbb{D}}(z, w).$$

Proof. If f/r takes a value of modulus one, maximum modulus makes it constant and both conclusions hold. Otherwise apply Schwarz–Pick to f/r:

$$\frac{(1-|z|^2)|f'(z)|}{1-|f(z)|^2} \le \frac{r^2-|f(z)|^2}{r(1-|f(z)|^2)} \le r.$$

The last inequality is equivalent, after multiplication by a positive denominator, to $(1-r^2)|f(z)|^2 \ge 0$. For every piecewise smooth path from z to w, the hyperbolic length of its image is at most r times its length. Taking the infimum over such paths proves the distance statement, including paths whose images have critical points. □

The factor r cannot be improved uniformly over this class: $f(z) = rz$ has hyperbolic derivative r at zero. A particular function can have a smaller optimal factor. Also the image need not be centered at zero. If $f(\mathbb{D})$ lies in the closed hyperbolic ball of radius $R > 0$ about a, conjugate by m_a. The new image lies in the Euclidean disk of radius $r = \tanh(R/2)$, and the same global factor follows. The center and radius are geometric information about the whole image, not about a finite set of observed iterates.

19.2 Existence, uniqueness and usable stopping rules

Theorem 19.2 (Iteration with error bounds)**.** *Under Proposition 19.1, f has a unique fixed point $p \in \mathbb{D}$. For $z_{n+1} = f(z_n)$ and any $z_0 \in \mathbb{D}$,*

$$d_{\mathbb{D}}(z_n, p) \le r^n d_{\mathbb{D}}(z_0, p), \tag{19.1}$$

$$d_{\mathbb{D}}(z_n, p) \le \frac{r^n}{1-r} d_{\mathbb{D}}(z_1, z_0), \tag{19.2}$$

$$d_{\mathbb{D}}(z_n, p) \le \frac{d_{\mathbb{D}}(z_n, z_{n+1})}{1-r}. \tag{19.3}$$

Proof. Repeated contraction gives $d_{\mathbb{D}}(z_{k+1}, z_k) \le r^k d_{\mathbb{D}}(z_1, z_0)$. For $m > n$, the triangle inequality bounds $d_{\mathbb{D}}(z_m, z_n)$ by the corresponding finite geometric sum. Its tail tends to zero as $n \to \infty$, so the sequence is Cauchy. Completeness of the disk metric gives a limit p in the disk. Continuity implies $f(p) = p$. If p' were another fixed point, $d_{\mathbb{D}}(p, p') \le r d_{\mathbb{D}}(p, p')$ would force $p = p'$.

Letting $m \to \infty$ in the tail estimate proves (19.2). Applying contraction n times to z_0, p proves (19.1). Finally,

$$d_{\mathbb{D}}(z_n, p) \le d_{\mathbb{D}}(z_n, z_{n+1}) + r\, d_{\mathbb{D}}(z_n, p),$$

which gives (19.3). □

The first estimate involves the unknown fixed point. The second uses only the initial step and the known factor. The third is an a posteriori stopping rule: a hyperbolic residual at most $(1-r)\varepsilon$ guarantees hyperbolic error at most ε. A small Euclidean residual without a known location or factor is not the same statement.

To compare metrics, the density bound $2/(1-|z|^2) \ge 2$ implies

$$2|z - w| \le d_{\mathbb{D}}(z, w). \tag{19.4}$$

If also $|z|, |w| \le R < 1$, their Euclidean segment stays in $R\overline{\mathbb{D}}$, so integration along that segment gives

$$d_{\mathbb{D}}(z, w) \le \frac{2|z - w|}{1 - R^2}. \tag{19.5}$$

After the first exact iterate, both z_n and p lie in $r\overline{\mathbb{D}}$. These comparisons translate the hyperbolic bounds into ordinary absolute-error guarantees with explicit constants. They also show why an estimate deteriorates if it is used too close to the boundary.

For a prescribed tolerance, (19.2) requires

$$r^n d_{\mathbb{D}}(z_1, z_0) \le (1-r)\varepsilon.$$

If the initial step is nonzero and the right side is smaller than that step, a sufficient integer is the ceiling of $\log(d_{\mathbb{D}}(z_1, z_0)/((1-r)\varepsilon))/\log(1/r)$. If the initial step vanishes, the initial point is already fixed; no logarithm of zero is taken. For fixed $r < 1$ and fixed initial data this is $O(\log(1/\varepsilon))$ iterations. When r varies toward one, the constant must be retained: $\log(1/r)$ tends to zero.

19.3 Inexact updates and stability of the answer

Let computed states $y_n \in \mathbb{D}$ satisfy

$$d_{\mathbb{D}}(y_{n+1}, f(y_n)) \le \delta_n.$$

The requirement that the states remain in the disk is part of the model; the inequality cannot be interpreted after a routine returns an exterior point. If $E_n = d_{\mathbb{D}}(y_n, p)$, the triangle inequality gives $E_{n+1} \le rE_n + \delta_n$. Induction yields

$$E_n \le r^n E_0 + \sum_{j=0}^{n-1} r^{n-1-j}\delta_j. \tag{19.6}$$

For $\delta_j \le \delta$, this becomes

$$E_n \le r^n E_0 + \delta\frac{1-r^n}{1-r}.$$

Thus a fixed per-step error gives a limiting error allowance $\delta/(1-r)$, not necessarily convergence to the exact fixed point. To guarantee a final tolerance ε, one may allocate half to the decaying initial term and require $\delta \le (1-r)\varepsilon/2$ for the other half.

If $\delta_j \le \delta\sigma^j$ with $0 < \sigma < 1$, the sum in (19.6) is

$$\delta\frac{r^n - \sigma^n}{r-\sigma} \quad (r \ne \sigma), \qquad \delta n r^{n-1} \quad (r = \sigma).$$

The quotient is nonnegative for either ordering of r and σ. Both expressions tend to zero. This proves convergence under a decaying error budget and identifies the extra factor n when the two rates coincide.

Perturbing the map rather than its evaluations gives another useful bound. Suppose f, g are contractions with factor at most $r < 1$, have fixed points p_f, p_g, and $d_{\mathbb{D}}(f(z), g(z)) \le \eta$ for every z. Then

$$d_{\mathbb{D}}(p_f, p_g) \le d_{\mathbb{D}}(f(p_f), f(p_g)) + d_{\mathbb{D}}(f(p_g), g(p_g)) \le r d_{\mathbb{D}}(p_f, p_g) + \eta,$$

so $d_{\mathbb{D}}(p_f, p_g) \le \eta/(1-r)$. This is a model sensitivity statement. It does not say how to establish the uniform bound η from finitely many samples; that is a separate analytic task.

19.4 Turning Euclidean evaluation bounds into admissible updates

An implementation usually returns a complex coordinate with an absolute error estimate. The metric error in (19.6) must then be derived, including the requirement that the returned coordinate lies in the disk. Suppose the exact map obeys $|f(z)| \le r < 1$ throughout $\mathbb{D}$ and the routine supplies y_{n+1} with

$$|y_{n+1} - f(y_n)| \le e_n.$$

Choose a fixed $\overline{e} < 1 - r$ and assume $e_n \le \overline{e}$ at every step. Starting with any $y_0 \in \mathbb{D}$, induction gives $|y_{n+1}| \le r + \overline{e} =: R < 1$. Thus every next call is made inside the domain, and both endpoints of the evaluation-error segment belong to $R\overline{\mathbb{D}}$. The metric comparison already proved gives

$$d_{\mathbb{D}}(y_{n+1}, f(y_n)) \le \frac{2e_n}{1-R^2}. \tag{19.7}$$

This derives an admissible hyperbolic budget from a Euclidean one. The margin $\overline{e} < 1 - r$ cannot be replaced by a small-looking decimal independent of r: near the boundary the same absolute error can leave the domain altogether.

For a desired eventual Euclidean error $\epsilon > 0$, one convenient explicit choice is

$$R = \frac{1+r}{2}, \qquad e_n \le e_* := \min\left\{\frac{1-r}{2}, \frac{(1-r)(1-R^2)\epsilon}{2}\right\}.$$

These bounds ensure disk membership, and the limiting hyperbolic evaluation allowance is at most ϵ. Since Euclidean distance is at most half hyperbolic distance, the eventual Euclidean allowance from evaluation errors is at most $\epsilon/2$. The initial term $r^n E_0/2$ can be made at most $\epsilon/2$ by sufficiently many iterations. This is a sufficient budget, with its constants displayed, not an optimal allocation of computational precision.

There is also a stopping test involving only a certified approximate step. Let $y \in \mathbb{D}$, let v approximate $f(y)$ with $|v - f(y)| \le e$, and suppose $|y|, |v|, |f(y)| \le R < 1$. The triangle inequality and the segment comparison give

$$d_{\mathbb{D}}(y, f(y)) \le \frac{2(|y - v| + e)}{1 - R^2}, \qquad |y - p| \le \frac{|y - v| + e}{(1 - r)(1 - R^2)}.$$

The location bound on y is an additional requirement for this particular Euclidean formula. It is automatic after one approximate step under the preceding budget, but need not hold for an arbitrary initial point. A routine can therefore turn a rigorous local error enclosure into a rigorous global stopping decision, provided it also knows the analytic contraction factor and the location margin.

19.4.1 Vanishing errors need no prescribed geometric rate

The explicit geometric error schedule is useful but stronger than necessary. If $\delta_j \to 0$, the sum in (19.6) still tends to zero. To prove this, fix an arbitrary $\epsilon > 0$ and choose J so that $\delta_j \le (1 - r)\epsilon/2$ for $j \ge J$. The tail contribution for $n > J$ is at most $\epsilon/2$ by the geometric-series bound. The finitely many earlier terms have sum

$$r^{n-J} \sum_{j=0}^{J-1} r^{J-1-j} \delta_j,$$

which tends to zero; so does $r^n E_0$. For sufficiently large n their combined contribution is below $\epsilon/2$. Hence $E_n \to 0$. This proves convergence from the stated limit of the certified errors, without silently assuming summability of $\sum_j \delta_j$.

By contrast, a constant nonzero error budget need not lead to the exact answer even for the simplest map. Take $f \equiv 0$ and repeatedly return the fixed real coordinate $a \in (0, 1)$. Every update has hyperbolic error $2 \operatorname{artanh} a$, but every computed state after the first stays at a. The exact fixed point is zero. The earlier limiting error allowance correctly permits this behavior.

19.5 Analytic dependence of a uniformly attracting fixed point

Let t vary over a complex domain T, and suppose $F(t, z)$ is holomorphic jointly for $(t, z) \in T \times \mathbb{D}$. Assume a single $r < 1$ satisfies $|F(t, z)| \le r$ throughout. Denote the unique fixed point of $F_t(z) = F(t, z)$ by $p(t)$. The following proof explains why one may differentiate the fixed-point equation; differentiability is not an extra conclusion drawn from convergence at a few parameter values.

Start $z_0(t) = 0$ and define $z_{n+1}(t) = F(t, z_n(t))$. Each iterate is holomorphic by composition. Its first displacement satisfies $d_{\mathbb{D}}(z_1(t), 0) \le 2 \operatorname{artanh} r$, uniformly in t. The a priori theorem therefore gives the common bound

$$|z_n(t) - p(t)| \le \frac{r^n \operatorname{artanh} r}{1 - r}.$$

Thus convergence is uniform on T, in particular locally uniform. The Cauchy integral formula on parameter disks compactly contained in T passes to the limit and shows that p is holomorphic,

with derivatives obtained by the corresponding Cauchy integrals. This is the precise analytic step behind differentiating the limiting answer.

Differentiate $p(t) = F(t, p(t))$ to obtain

$$p'(t) = \frac{\partial_t F(t,p(t))}{1 - \partial_z F(t,p(t))}, \qquad |p'(t)| \le \frac{|\partial_t F(t,p(t))|}{1-r}. \tag{19.8}$$

At a fixed point the source and target metric densities coincide, so Proposition 19.1 gives $|\partial_z F(t,p(t))| \le r$. The denominator in (19.8) is consequently nonzero and has modulus at least $1-r$. A uniform bound is needed only locally in the parameter for a local conclusion; the global hypothesis above keeps the statement and its constants simple.

For example, $F(t,z) = (1+z^2+t)/4$ has image in the disk of radius $3/4$ for $|t| < 1$. At $t = 0$, $p = 2-\sqrt{3}$ and $\partial_z F = p/2$, $\partial_t F = 1/4$, so $p'(0) = 1/(2\sqrt{3})$. The estimate using $r = 3/4$ gives $|p'(0)| \le 1$. The exact derivative is smaller, as expected of a uniform estimate over a whole family. This argument controls dependence on an analytic parameter; it does not assign accuracy or execution cost to an unspecified evaluator of F.

19.6 A nonaffine worked iteration

Consider

$$f(z) = \frac{1+z^2}{4}.$$

For $|z| < 1$, $|f(z)| < (1+1)/4 = 1/2$, so $r = 1/2$ is available. The fixed-point equation is $z^2 - 4z + 1 = 0$. Its roots are $2-\sqrt{3}$ and $2+\sqrt{3}$, and only $p = 2-\sqrt{3}$ lies in the disk. Starting at zero gives

$$z_1 = \frac{1}{4}, \qquad z_2 = \frac{17}{64}, \qquad z_3 = \frac{4385}{16384}.$$

The global theorem proves convergence from every starting point in the disk, not just from this real sequence.

At z_2, the Euclidean residual is $|z_3 - z_2| = 33/16384$. Both points lie in the disk of radius $1/2$. Equations (19.3)–(19.5) therefore give

$$|z_2 - p| \le \frac{1}{(1-r)(1-r^2)}|z_3 - z_2| = \frac{11}{2048}.$$

This is valid without using the explicit square-root formula for p. For this particular map one can do better. On $|z| \le 1/2$, $|f'(z)| = |z|/2 \le 1/4$, and the Euclidean segment between any two points remains in that disk. Hence f is a Euclidean contraction there with factor $1/4$, giving

$$|z_2 - p| \le \frac{4}{3}|z_3 - z_2| = \frac{11}{4096}.$$

The stronger special estimate does not invalidate the geometric one. It shows the value of identifying additional structure after a general argument has established existence and convergence.

19.7 An affine family and the cost of exact output

For $f(z) = az + b$ with $|a| + |b| \le r < 1$, the image is contained in $r\overline{\mathbb{D}}$. Since $|a| < 1$, the fixed point and iterates are

$$p = \frac{b}{1-a}, \qquad f^n(z) - p = a^n(z-p). \tag{19.9}$$

The first identity is substitution; the second follows by induction from $f(z) - p = a(z - p)$. Also $|p| \le |b|/(1 - |a|) < 1$. The constant case $a = 0$ reaches the fixed point in one step and needs no asymptotic rate discussion.

Take $a = 1/4$, $b = 1/2$, $z_0 = 0$. Then

$$p = \frac{2}{3}, \qquad z_n = \frac{2}{3}(1 - 4^{-n}), \qquad |p - z_n| = \frac{2}{3}\,4^{-n}.$$

For $n \ge 1$, reduction of this fraction gives denominator 2^{2n-1} and numerator $(4^n - 1)/3$. The latter is $1 + 4 + \cdots + 4^{n-1}$, an odd integer. Thus a numerator/denominator binary representation has $\Theta(n)$ bits. The exact answer itself grows with the number of steps, even though each abstract field update uses only a fixed number of operations.

Under a bit model in which addition, shifts and division by a fixed small integer cost linearly in the operand length, the closed expression can be formed in $O(n)$ bit operations: write the $2n + 1$-bit integer 4^n, subtract one, divide by three, and write the power-of-two denominator. The output has $\Omega(n)$ bits, so this final-value construction has the correct order for that output convention. This claim does not require or imply a measured implementation speed.

If instead all exact iterates through step n must be output, their combined size is $\sum_{k=1}^{n} \Theta(k) = \Theta(n^2)$. Writing the whole trajectory therefore has a different lower bound from returning only its last point. An iteration count, an arithmetic-operation count, the cost of large integers and the requested output format are distinct quantities. Confusing them can turn a valid geometric rate into an incorrect computational complexity claim.

For an absolute tolerance 2^{-m} with integer $m \ge 1$, taking $n = \lceil m/2 \rceil$ is sufficient. The final exact rational then has $\Theta(m)$ bits and is constructible in $O(m)$ bit operations in the stated model. This is a complete complexity statement for this explicit family, not for an arbitrary holomorphic map supplied by an unspecified evaluation oracle.

19.8 Solved exercises

Investigation 19.3 (A rigorous stopping rule with evaluation error)**.** Suppose $d_{\mathbb{D}}(y, f(y))$ is not directly known, but an approximate update v satisfies $d_{\mathbb{D}}(v, f(y)) \le \delta$. Derive a bound for the distance from y to the fixed point using $d_{\mathbb{D}}(y, v)$.

Solution. The residual is at most $d_{\mathbb{D}}(y, v) + \delta$. The same one-step argument used in (19.3) therefore gives

$$d_{\mathbb{D}}(y, p) \le \frac{d_{\mathbb{D}}(y, v) + \delta}{1 - r}.$$

Stopping at a small observed step without including δ can underestimate the true error. The estimate is valid for any $y, v \in \mathbb{D}$ satisfying the stated bound, whether or not they belong to a long run.

Investigation 19.4 (Why a boundary fixed point is not covered)**.** Consider $f(z) = (z + 1)/2$. Explain why the interior-image theorem cannot be applied with any $r < 1$, and compute its iterates.

Solution. Along the real radius $z \uparrow 1$, $f(z) \uparrow 1$, so $\sup_{\mathbb{D}} |f| = 1$. Its iterates are $f^n(z) = 1 + 2^{-n}(z - 1)$, tending to the boundary point 1. There is no fixed point in $\mathbb{D}$. Euclidean convergence to the boundary is compatible with completeness of the hyperbolic disk because the sequence is not hyperbolically Cauchy. The missing uniform interior-image hypothesis is essential.

Investigation 19.5 (Composing several controlled updates)**.** If f_j has hyperbolic Lipschitz factor $r_j < 1$ for $1 \le j \le k$, find a factor for $F = f_k \circ \cdots \circ f_1$ and state an error rule for repeating F.

Solution. Apply the Lipschitz inequalities successively to obtain $d_{\mathbb{D}}(F(z), F(w)) \le (\prod_j r_j) d_{\mathbb{D}}(z, w)$. With $r = \prod_j r_j$, the fixed-point proof and residual estimate apply unchanged. If one cycle is evaluated with a known hyperbolic error δ, the limiting error allowance is $\delta/(1-r)$. To derive δ from errors inside a cycle, an error introduced after stage j is multiplied by at most $\prod_{\ell=j+1}^{k} r_\ell$ before the cycle ends; sum these propagated contributions.

Chapter 20

Lorentz geometry, velocity space and rapidity

The hyperbolic disk has a second physical interpretation, distinct from the impedance coordinate: it parametrizes future-directed unit timelike vectors. The resulting geometry is the space of velocities in special relativity. Its points are velocities, not positions in physical space. The distinction matters both for units and for the meaning of distance.

We work in two spatial dimensions so that the velocity space is a disk. The same hyperboloid construction extends to three spatial dimensions, but complex holomorphic notation is specific to the two-dimensional spatial slice. The physical assumptions are flat Minkowski spacetime, inertial-frame special relativity, a fixed speed of light $c > 0$, and massive future-directed motion with speed less than c. Dray's geometric treatment [9] is a primary pedagogical reference. The coordinate changes and metric calculations are supplied here in full.

20.1 The hyperboloid and its disk coordinate

For $U = (T, X, Y)$ and $V = (T', X', Y')$, use the Lorentz bilinear form

$$L(U, V) = TT' - XX' - YY'.$$

The future unit hyperboloid is

$$\mathcal{H} = \{(T, X, Y) : L(U, U) = 1,\ T > 0\}.$$

If W is tangent to $\mathcal{H}$ at U, differentiation gives $L(U, W) = 0$. The form $-L(W, W)$ is positive definite on this tangent plane. For example, writing the tangent condition as $TW_T = XW_X + YW_Y$ and using $T^2 = 1 + X^2 + Y^2$ shows

$$-L(W, W) = W_X^2 + W_Y^2 - \frac{(XW_X + YW_Y)^2}{1 + X^2 + Y^2} > 0$$

for every nonzero tangent vector. Cauchy–Schwarz bounds the subtracted term by a strict factor smaller than one times $W_X^2 + W_Y^2$. This defines a Riemannian metric on the hyperboloid.

For $z = x + iy \in \mathbb{D}$, define

$$\Psi(z) = \left(\frac{1 + |z|^2}{1 - |z|^2}, \frac{2x}{1 - |z|^2}, \frac{2y}{1 - |z|^2} \right). \tag{20.1}$$

The identity $(1 + r^2)^2 - 4r^2 = (1 - r^2)^2$ shows that $\Psi(z) \in \mathcal{H}$. Conversely, for $U \in \mathcal{H}$,

$$z = \frac{X + iY}{T + 1} \tag{20.2}$$

has $|z|^2 = (T-1)/(T+1) < 1$, and substitution recovers U. The two maps are smooth inverses.

To compute the metric, let $z = re^{i\theta}$ away from zero. Put $R = 2r/(1-r^2)$ and $T = (1+r^2)/(1-r^2)$. Then

$$T'(r) = \frac{4r}{(1-r^2)^2}, \qquad R'(r) = \frac{2(1+r^2)}{(1-r^2)^2}, \qquad (R')^2 - (T')^2 = \frac{4}{(1-r^2)^2}.$$

Since $\mathrm{d}X^2 + \mathrm{d}Y^2 = \mathrm{d}R^2 + R^2\mathrm{d}\theta^2$, the pullback is

$$\mathrm{d}X^2 + \mathrm{d}Y^2 - \mathrm{d}T^2 = \frac{4(\mathrm{d}r^2 + r^2\mathrm{d}\theta^2)}{(1-r^2)^2} = \frac{4(\mathrm{d}x^2 + \mathrm{d}y^2)}{(1-|z|^2)^2}. \tag{20.3}$$

The Cartesian expression extends smoothly across zero. It is precisely the curvature -1 disk metric used throughout this volume.

20.2 Distance as relative rapidity

For $z, w \in \mathbb{D}$, expansion of (20.1) gives

$$L(\Psi(z), \Psi(w)) = 1 + \frac{2|z-w|^2}{(1-|z|^2)(1-|w|^2)}. \tag{20.4}$$

If $\rho = |m_w(z)|$, the disk deficit identity gives

$$\frac{\rho^2}{1-\rho^2} = \frac{|z-w|^2}{(1-|z|^2)(1-|w|^2)}.$$

Because $d_{\mathbb{D}}(z,w) = 2\,\mathrm{artanh}\,\rho$ and $\cosh(2\,\mathrm{artanh}\,\rho) = 1 + 2\rho^2/(1-\rho^2)$, we obtain

$$L(\Psi(z), \Psi(w)) = \cosh d_{\mathbb{D}}(z,w). \tag{20.5}$$

This is a direct bridge from complex fractional transformations to a Lorentz-invariant quantity. A future-preserving linear Lorentz transformation preserves L, carries $\mathcal{H}$ to itself, and therefore preserves this distance. It also preserves the tangent metric by its defining bilinear identity.

For a physical velocity vector $v = (v_x, v_y)$ with $|v| < c$, write $\beta = v/c$ and $\gamma = (1-|\beta|^2)^{-1/2}$. Its unit four-velocity, divided by c, is

$$U = (\gamma, \gamma\beta_x, \gamma\beta_y).$$

The corresponding disk coordinate is

$$z = \frac{\gamma(\beta_x + i\beta_y)}{\gamma + 1} = \frac{\beta_x + i\beta_y}{1 + \sqrt{1-|\beta|^2}}, \qquad \beta_x + i\beta_y = \frac{2z}{1+|z|^2}. \tag{20.6}$$

Thus physical normalized velocity is the Klein coordinate of the same hyperboloid point. It is not the Poincaré coordinate z.

In one spatial direction, let signed rapidity be $\eta = \mathrm{artanh}\,\beta$. Then $U = (\cosh\eta, \sinh\eta, 0)$ and $z = \tanh(\eta/2)$ is real. Hence

$$d_{\mathbb{D}}(z_1, z_2) = |\eta_1 - \eta_2|.$$

If one instead inserts the real number β itself into the Poincaré metric, the resulting distance is twice the rapidity difference. That is a different identification of physical data with the disk. Formula (20.6) fixes the identification used here and removes the factor-of-two ambiguity.

For two velocities, their invariant relative rapidity η_{rel} is determined by

$$\cosh\eta_{\mathrm{rel}} = \gamma_1\gamma_2(1 - \beta_1 \cdot \beta_2), \qquad \frac{|v_{\mathrm{rel}}|}{c} = \tanh\eta_{\mathrm{rel}}.$$

The first formula follows from the Lorentz product of the two unit four-velocities. In the rest frame of either one, that product is the other one's Lorentz factor, which proves the second formula. It is a relative speed, not a vector obtained by subtracting laboratory velocities.

20.3 The geodesic between two four-velocities

The distance formula can be accompanied by a minimizing path and a direct length proof. This also identifies the equality case in the proper-acceleration budget below. Take different $U, V \in \mathcal{H}$ and put

$$\theta = \operatorname{arcosh} L(U,V) > 0, \qquad E = \frac{V - (\cosh\theta)U}{\sinh\theta}.$$

The inverse hyperbolic cosine is well-defined because the disk identity (20.4) gives $L(U,V) > 1$ for distinct points. Compute the Lorentz products before using these vectors geometrically:

$$L(U,E) = 0, \qquad L(E,E) = \frac{1 - 2\cosh^2\theta + \cosh^2\theta}{\sinh^2\theta} = -1.$$

Thus E is a unit tangent vector at U for the positive tangent metric $-L$. Define

$$\gamma(s) = \cosh s\, U + \sinh s\, E, \qquad 0 \le s \le \theta. \tag{20.7}$$

The preceding products give $L(\gamma,\gamma) = 1$ and $L(\gamma',\gamma') = -1$. Its time component stays positive: it starts positive and cannot cross zero on the locus $L = 1$. Consequently it is a path in $\mathcal{H}$, parametrized by unit hyperbolic speed. Its endpoints are U and V, so its length is θ.

To prove no path is shorter, let $W(t)$ be any piecewise C^1 path from U to V in $\mathcal{H}$. Write $R(t) = \operatorname{arcosh} L(U, W(t))$. At a time with $R > 0$, the vector

$$Q = U - L(U,W)W$$

is tangent at W, since $L(Q,W) = 0$, and has $L(Q,Q) = 1 - L(U,W)^2 = -\sinh^2 R$. Also $L(W,W') = 0$. Cauchy–Schwarz in the positive tangent metric gives

$$|\sinh R\, R'| = |L(U,W')| = |L(Q,W')| \le \sinh R\sqrt{-L(W',W')}.$$

Division by $\sinh R > 0$ bounds $|R'|$ by the path speed. The division cannot be made at $R = 0$, but no singular case needs to be ignored. For $0 < \epsilon < \theta$, choose the last time before the endpoint at which $R = \epsilon$. It exists by continuity and compactness of the parameter interval. Thereafter $R > \epsilon$, so integrating the derivative bound on that terminal portion gives $\theta - \epsilon \le \text{length}(W)$. Letting $\epsilon \downarrow 0$ proves the lower bound θ.

If $U = V$, the distance is zero and the constant path is minimizing; the formula with division by $\sinh\theta$ is not used. For distinct endpoints, equality in the tangent Cauchy–Schwarz inequality forces motion in the radial tangent direction wherever the path is moving. Equality in the integral also rules out radial backtracking. Thus a minimizing path traces the geodesic above, up to an increasing change of parameter and stationary intervals. This is the geometric reason that changing the direction of acceleration generally spends more velocity-space length than the endpoint rapidity requires.

The hyperbolic midpoint has a particularly simple invariant formula:

$$M = \frac{U+V}{\sqrt{L(U+V,U+V)}} = \frac{U+V}{2\cosh(\theta/2)}. \tag{20.8}$$

Indeed, $L(U+V,U+V) = 2 + 2\cosh\theta = 4\cosh^2(\theta/2)$, and substitution of $V = \cosh\theta U + \sinh\theta E$ shows $M = \gamma(\theta/2)$. Its future time component is positive. This formula averages unit timelike vectors and then normalizes their Lorentz length. Averaging the physical velocity coordinates without that normalization would answer a different question.

Investigation 20.1 (An invariant midpoint in one direction)**.** Find the velocity-space midpoint between rest and speed $4c/5$ in one spatial direction, and compare it with their arithmetic mean speed.

Solution. The endpoint rapidity is $\operatorname{artanh}(4/5) = \log 3$. Half rapidity gives speed $c\tanh(\frac{1}{2}\log 3) = c/2$. Equivalently, the two unit vectors are $(1,0,0)$ and $(5/3,4/3,0)$. The ratio of spatial to temporal coordinates in their positive sum is $(4/3)/(8/3) = 1/2$, unaffected by the normalization in (20.8). The arithmetic mean of the two initial speeds is $2c/5$, which differs from the hyperbolic midpoint speed. Both averaging rules are well-defined, but only the normalized Lorentz sum has the stated invariant geometric meaning.

20.4 Boosts and their fractional action

For a real parameter η_0, define the active boost in the X direction by

$$\begin{pmatrix} T' \\ X' \\ Y' \end{pmatrix} = \begin{pmatrix} \cosh\eta_0 & \sinh\eta_0 & 0 \\ \sinh\eta_0 & \cosh\eta_0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} T \\ X \\ Y \end{pmatrix}. \tag{20.9}$$

Expansion using $\cosh^2\eta_0 - \sinh^2\eta_0 = 1$ proves that it preserves L. Put $a = \tanh(\eta_0/2)$. Under (20.2), its disk action is

$$B_a(z) = \frac{z+a}{1+az}. \tag{20.10}$$

For a direct check, use $\cosh\eta_0 = (1+a^2)/(1-a^2)$ and $\sinh\eta_0 = 2a/(1-a^2)$ in (20.9), substitute (20.1), and form $(X'+iY')/(T'+1)$. Cancellation gives (20.10); its denominator is nonzero for $|a|<1$, $|z|<1$.

Dividing the boosted spatial coordinates by T' gives, with $\beta_0 = \tanh\eta_0$ and $\gamma_0 = \cosh\eta_0$,

$$\beta_x' = \frac{\beta_x + \beta_0}{1+\beta_0\beta_x}, \qquad \beta_y' = \frac{\beta_y}{\gamma_0(1+\beta_0\beta_x)}.$$

The denominator is positive because $|\beta_0\beta_x| < 1$. For a passive change to the oppositely moving coordinate frame, replace η_0 by $-\eta_0$; fixing the active convention avoids a sign ambiguity in the addition law.

Collinear boosts add rapidities. Their physical normalized velocities therefore combine as

$$\beta \oplus \beta_0 = \frac{\beta+\beta_0}{1+\beta\beta_0}. \tag{20.11}$$

The denominator is positive, and

$$1-(\beta\oplus\beta_0)^2 = \frac{(1-\beta^2)(1-\beta_0^2)}{(1+\beta\beta_0)^2} > 0.$$

Thus finite subluminal inputs remain subluminal. This follows both from the real fractional identity and from addition of finite rapidities.

Example 20.2 (An exact collinear composition)**.** Take $\beta = 3/5$ and $\beta_0 = 5/13$. Their Lorentz factors are $5/4$ and $13/12$, and their Poincaré coordinates are $1/3$ and $1/5$, respectively. Equation (20.11) gives $\beta\oplus\beta_0 = 4/5$. In disk coordinates,

$$\frac{1/3+1/5}{1+1/15} = \frac{1}{2}, \qquad \frac{2(1/2)}{1+(1/2)^2} = \frac{4}{5}.$$

The two calculations agree after the correct coordinate conversion. Their rapidities are $\log 2$, $\log(3/2)$ and $\log 3$.

20.5 Why noncollinear boosts are different

Allow $a \in \mathbb{D}$ and write $B_a(z) = (z + a)/(1 + \overline{a}z)$. Rotating the axes reduces this to (20.10), so it describes a boost in the direction of a. Multiplying the two fractional transformations gives

$$B_a \circ B_b = B_c \circ R_\kappa, \qquad c = \frac{a + b}{1 + \overline{a}b}, \quad \kappa = \frac{1 + a\overline{b}}{1 + \overline{a}b}, \quad R_\kappa(z) = \kappa z. \tag{20.12}$$

The denominator factors are nonzero. Also $c = B_a(b) \in \mathbb{D}$ and $|\kappa| = 1$ because its numerator and denominator are conjugates. To check the identity, the left side has numerator $(1 + a\overline{b})z + a + b$ and denominator $(\overline{a} + \overline{b})z + 1 + \overline{a}b$. Divide both by $1 + \overline{a}b$ and note $\overline{c}\,\kappa = (\overline{a} + \overline{b})/(1 + \overline{a}b)$.

For collinear parameters the rotation can disappear. For general directions it does not. Taking $a = 1/3$ and $b = i/3$ gives

$$c = \frac{15 + 12i}{41}, \qquad \kappa = \frac{40 - 9i}{41}.$$

The rotation is nontrivial, and $|\kappa| = 1$ follows from $40^2 + 9^2 = 41^2$. The final rest-state velocity is $2c/(1 + |c|^2) = 3/5 + (12/25)i$, agreeing with the transverse-velocity formula above. A scalar addition of rapidities cannot describe this two-direction composition. Equation (20.12) exhibits the rotation algebraically; it is the geometric origin of the rotation associated with composing noncollinear boosts.

20.6 Acceleration, energy and the boundary

Let a timelike worldline be parametrized by proper time τ and let $U(\tau)$ be its dimensionless unit four-velocity. Its physical four-acceleration is $A = c\,\mathrm{d}U/\mathrm{d}\tau$. Since $L(U, U) = 1$, differentiation gives $L(U, U') = 0$, so the magnitude of proper acceleration is

$$a_{\mathrm{prop}} = \sqrt{-L(A, A)} = c\sqrt{-L(U', U')} = c\,\frac{\mathrm{d}\ell_{\mathcal{H}}}{\mathrm{d}\tau}. \tag{20.13}$$

Here $\mathrm{d}\ell_{\mathcal{H}}/\mathrm{d}\tau$ means metric speed, which is nonnegative. Hyperbolic length is dimensionless, so the right side has units of length divided by time squared. Integrating and using that distance is at most path length gives

$$d_{\mathcal{H}}(U(\tau_0), U(\tau_1)) \le \frac{1}{c}\int_{\tau_0}^{\tau_1} a_{\mathrm{prop}}(\tau)\,\mathrm{d}\tau.$$

This is a geometric bound on the change of velocity from a proper acceleration budget. It does not identify proper time with laboratory coordinate time.

For collinear motion with constant signed proper acceleration $a > 0$ and initial rest, $\eta' = a/c$, so $\eta = a\tau/c$ and $v = c\tanh(a\tau/c)$. Since $\mathrm{d}t/\mathrm{d}\tau = \cosh\eta$ and $\mathrm{d}x/\mathrm{d}\tau = c\sinh\eta$, initial position and time zero give

$$t(\tau) = \frac{c}{a}\sinh(a\tau/c), \qquad x(\tau) = \frac{c^2}{a}\big(\cosh(a\tau/c) - 1\big).$$

The argument of each hyperbolic function is dimensionless, c/a is a time and c^2/a a length. These formulas describe an ideal prescribed worldline in flat spacetime, with no dynamical model of an accelerating engine or its energy supply.

Eliminating proper time makes the distinction between proper and laboratory acceleration explicit. Since $\sinh(a\tau/c) = at/c$,

$$v(t) = \frac{at}{\sqrt{1 + (at/c)^2}}, \qquad \frac{dv}{dt} = \frac{a}{[1 + (at/c)^2]^{3/2}}.$$

The latter equals a initially and then decreases, although the proper acceleration remains constant. The position formula becomes

$$x(t) = \frac{c^2}{a}\left(\sqrt{1+(at/c)^2}-1\right), \qquad \left(x+\frac{c^2}{a}\right)^2 - c^2t^2 = \left(\frac{c^2}{a}\right)^2.$$

The hyperbola is in spacetime, while the earlier hyperboloid parametrizes velocities. Their related formulas do not identify these two spaces. At short times $|at/c| \ll 1$, the Taylor expansion gives $x(t) = at^2/2 + O(a^3t^4/c^2)$ and $v(t) = at + O(a^3t^3/c^2)$. These are local small-parameter statements, with a and c fixed; they do not extend a constant laboratory-acceleration approximation to arbitrarily large times.

For rest mass $m > 0$, the energy-momentum vector is $(E/c, p_x, p_y) = mcU$. Hence

$$E = mc^2\gamma, \qquad p = m\gamma v, \qquad E^2 - c^2|p|^2 = m^2c^4.$$

For the preceding constant-proper-acceleration trajectory, these identities give $p(t) = mat$ in its fixed spatial direction and $E(t) = mc^2\sqrt{1+(at/c)^2}$. Hence the laboratory force $dp/dt = ma$ is constant in this one-dimensional idealization, and direct differentiation verifies the power relation $dE/dt = (ma)v(t)$. In particular, constant force, constant proper acceleration and constant laboratory acceleration are not three interchangeable statements. The first two coincide for this fixed-mass collinear model, while the explicit derivative above distinguishes the third. None of these kinematic calculations supplies a model of fuel consumption or the total energy of an accelerating apparatus.

Approaching $|z| = 1$ means $|v| \to c$, $\gamma \to \infty$ and hyperbolic distance from rest tending to infinity. At fixed positive rest mass the energy therefore diverges. A photon is not obtained by simply adding a boundary point to this unit massive hyperboloid: its null momentum has a different normalization and no rest frame.

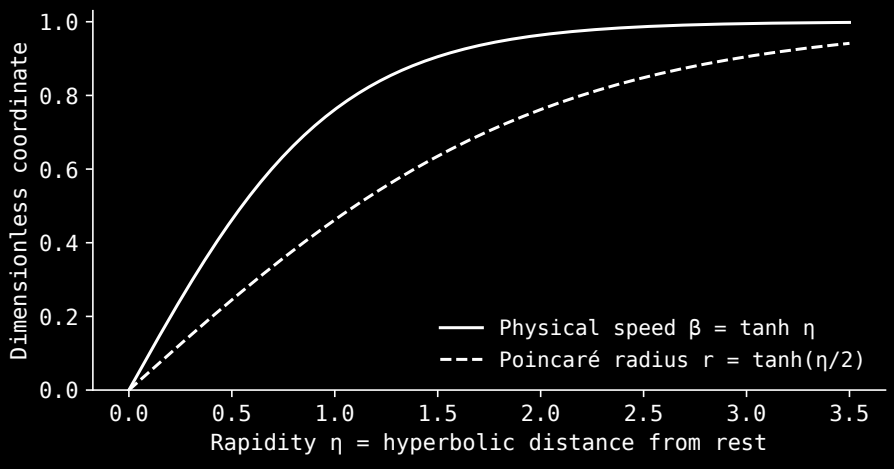


Figure 20.1: Physical normalized speed and Poincaré radius at the same rapidity. Their distinct formulas describe the same velocity-space point; the hyperbolic distance from rest equals rapidity.

20.7 Solved exercises

Investigation 20.3 (Equal speeds in perpendicular directions)**.** Two observers have laboratory velocities $(3c/5, 0)$ and $(0, 3c/5)$. Find their invariant relative Lorentz factor and relative speed.

Solution. Both laboratory Lorentz factors are $5/4$ and the normalized velocity dot product is zero. Thus $\gamma_{\mathrm{rel}} = 25/16$ and

$$|v_{\mathrm{rel}}|/c = \sqrt{1 - (16/25)^2} = \frac{\sqrt{369}}{25}.$$

This differs from the Euclidean speed difference $3\sqrt{2}/5$. The distinction follows from the Lorentz product, without choosing an explicit coordinate transformation to either observer's rest frame.

Investigation 20.4 (A proper acceleration budget)**.** A body starts at rest and has integrated proper acceleration at most cR over a proper-time interval, where $R \geq 0$. Bound its final speed and its Lorentz factor.

Solution. Equation (20.13) bounds the hyperbolic distance from rest by R. This distance is the final rapidity magnitude, so $|v| \leq c \tanh R$ and $\gamma \leq \cosh R$. Equality can be attained by motion with acceleration always in the same spatial direction and with the prescribed integrated magnitude. Changing direction can spend path length without producing the same endpoint distance.

Investigation 20.5 (Low-speed coordinates)**.** Find the first nonzero term in the relation between the Poincaré coordinate and a small normalized velocity, and interpret the factor two.

Solution. From (20.6), $z = \beta/(1 + \sqrt{1 - |\beta|^2}) = \beta/2 + O(|\beta|^3)$, where complex notation represents the planar vector. Near zero, the Poincaré metric is $2|\mathrm{d}z|$ to first order, so it becomes $|\mathrm{d}\beta|$ to first order. The coordinate factor $1/2$ and metric factor 2 cancel. This is consistent with rapidity being approximately $|v|/c$ at low speed; it is not a change in the physical value of c.

Chapter 21

Source correspondence and mathematical perspective

The principal source is Lyubich's Section 7 in the supplied manuscript [1]. The developments here expand that short research exposition and supply supporting analysis and applications. This chapter records how the topics correspond and explains the scope of the corrections. It is a reading guide, not a count of independently formalized theorems.

21.1 The seven source subsections

The source's first subsection establishes Schwarz contraction on hyperbolic surfaces. Chapter 2 proves the disk inequality, its equality case and distance form, then transports the result through covering maps. The classical uniformization and lifting theorems supply the global setting. Chapter 4 gives the curvature comparison argument for the differential inequality, including critical points. The two arguments share elementary complex analysis but use different contraction mechanisms.

The symmetric principle in source §7.2 is developed through circular neighborhoods of a real interval. Chapter 7 identifies these neighborhoods as exact hyperbolic tubes, proves their inclusion under maps preserving the interval, and gives quantitative derivative and angular estimates. A fixed non-Möbius map has a strict compact-set contraction factor. The family in Proposition 5.1 shows why that fact does not produce a common factor for every such map with only the same interior interval separation. The distinction is a quantifier issue rather than a small numerical change to a constant.

Source §7.3 concerns the hyperbolic density near the boundary. The exact punctured-disk metric, isolated-cusp asymptotic, positive spherical comparison and small-distance comparison near a compact removed set appear in Chapter 7. The spherical formula uses $d\log(e\pi/d)$, which remains positive over the whole spherical distance range. The local background-metric formula uses $d\log(1/d)$ only for a specified sufficiently small d. The literal expression at arbitrary distances can vanish or change sign and cannot be used as a global positive comparison.

Source §7.4 compares the metric with inverse Euclidean boundary distance. Theorem 3.4 and Corollary 3.5 give the full univalent-function prerequisites and the corrected curvature--1 bounds $1/(2d) \le \lambda_U \le 2/d$. The proof uses the included disk for the upper bound and the Koebe quarter disk for the lower bound. The direction of that latter inclusion is as important as the factor of two.

Source §7.5 turns qualitative contraction into a definite estimate. Theorem 6.2 gives the sharp function $B(r) = 2r\log(1/r)/(1-r^2)$ for an omitted point. Its surface form and the interior-depth estimate are proved in Propositions 7.13 and 7.14. Round annuli show why absolute upper and

lower modulus bounds alone do not give a strict inclusion gap. The two-sheeted covering clause has an additional modulus relation; Theorem 9.4 uses it to obtain a uniform expansion factor. The two statements must be treated separately.

Source §7.6 joins cusps, collars, conformal width and thick–thin geometry. Chapters 8 and 9 supply exact quotient coordinates, the embedded unit-cusp argument and the individual sharp collar proof. The subsequent geometric development relates area, injectivity radius and conformal comparisons. The simultaneous collar theorem and the Margulis input used for the global thin-component classification are named classical theorems with their scopes stated. A proof for the lifts of one simple geodesic is not silently promoted to the unequal-geodesic simultaneous theorem. The formulas use area $= -2\pi\chi$ in curvature -1 and keep the annular modulus distinct from its reciprocal width.

Source §7.7 concerns Carathéodory convergence. The chapters beginning at Chapter 11 define the pointed kernel using eventual containment of compact sets in the basepoint component and require the same kernel for every subsequence. They prove the normalized mapping theorem, inverse convergence, continuity of conformal radius and appropriate compactness statements. Chapter 12 supplies the planar Riemann mapping existence proof needed by those normalized coordinates: homotopy invariance yields logarithms, a square root gives a bounded injective map, and a normal-family extremizer is proved onto by a strict derivative improvement. The examples with escaping basepoints, strict annular bounds and uncontrolled target scale explain the hypotheses. Boundary attraction and the Denjoy–Wolff theorem extend the source's contraction theme into iteration.

21.2 Which conclusions are classical inputs

The local complex-analysis and univalent-function chains needed for the main disk estimates are developed explicitly, as is the Riemann mapping theorem for simply connected proper plane domains. General Riemann-surface uniformization, elementary covering-space lifting and the classification of finite-type hyperbolic ends are classical prerequisites. The simultaneous collar theorem and the Margulis lemma are invoked where their global content is needed. Gauss–Bonnet is stated with its boundary terms, and the surface area consequences are derived with the chosen curvature convention.

This division has mathematical consequences. A disk calculation can be checked entirely within the preceding local chapters. A theorem about an arbitrary hyperbolic surface additionally depends on the global existence and covering inputs. A later proof-assistant programme must represent both the theorem and those inputs if it is to check that full statement. The present publication supplies an expository mathematical argument with named classical dependencies; it does not claim that all of its analytical and geometrical content has already been checked by a proof assistant.

21.3 The applications and their boundaries

The harmonic chapter connects the metric to positive harmonic functions, reproducing kernels and two-dimensional energy. The curvature chapter proves residual bounds for a specified elliptic equation. Finite interpolation includes singular data. Passive-system calculations specify an ideal impedance model and distinguish interior constraints from boundary frequency values. Controlled iteration states its metric error criterion and arithmetic cost model. The Lorentz interpretation specifies velocity coordinates, units and flat spacetime. These assumptions delimit the applications of the proved mathematics; no measured physical validation is asserted.

Bibliography

[1] Mikhail Lyubich, *Conformal Geometry and Dynamics of Quadratic Polynomials*, Volumes I–II, supplied 702-page manuscript, PDF metadata 14 September 2023. Principal source: §7, pp. 133–145; metric normalization: p. 65. Author's book location: `https://www.math.stonybrook.edu/~mlyubich/book.pdf`. Page references in this volume concern the supplied file, whose identifier is recorded in the source bundle; a later online file may have different pagination.

[2] Lars V. Ahlfors, "An extension of Schwarz's lemma," *Transactions of the American Mathematical Society* 43 (1938), 359–364. `https://doi.org/10.1090/S0002-9947-1938-1501949-6`.

[3] Dmitry Beliaev, *Conformal Maps and Geometry*, author lecture notes, 8 June 2015, hosted in Christopher Bishop's MAT 627 course materials. Planar uniformization and univalent-function theory. `https://www.math.stonybrook.edu/~bishop/classes/math627.S25/Beliaev_notes.pdf`.

[4] Ara Basmajian, "Generalizing the hyperbolic collar lemma," *Bulletin of the American Mathematical Society* 27 (1992), 154–158. Author manuscript: `https://arxiv.org/abs/math/9207211`. The introduction records the sharp simultaneous collar theorem and its attribution.

[5] Samuel Ballas, *Lecture 7: The Margulis Lemma*, author lecture notes in hyperbolic geometry. `https://www.math.fsu.edu/~ballas/260p/lectures/lecture7/`.

[6] Donald E. Marshall, "An elementary proof of the Pick–Nevanlinna interpolation theorem," *Michigan Mathematical Journal* 21 (1974, printed issue year), 219–223. `https://doi.org/10.1307/mmj/1029001307`.

[7] Orr Shalit, *Topics in Operator Theory, Lecture 4: Pick interpolation via commutant lifting*, author lecture notes, 11 November 2018. Author's lecture page at Noncommutative Analysis.

[8] Hermann A. Haus and James R. Melcher, *Electromagnetic Fields and Energy*, 1989, Chapter 14, especially §14.6, reflection coefficients and steady-state transmission lines. Author textbook made available by MIT: `https://web.mit.edu/6.013_book/www/chapter14/14.6.html`.

[9] Tevian Dray, *The Geometry of Special Relativity*, author online text, chapters on the Poincaré disk and velocity addition. `https://sites.science.oregonstate.edu/physics/coursewikis/GSR/book/gsr/poincare.html`; `https://sites.science.oregonstate.edu/physics/coursewikis/GSR/book/content/vaddition.html`.

[10] PHI research collective, *Uniformization and Geometric Analysis*, PHI Technical Monographs, Volume V. Series background on uniformization, harmonic analysis and variational geometry.

[11] PHI research collective, *Extremal Length and Variational Geometry*, PHI Technical Monographs, Volume VI. Series background on extremal length, annular modulus and collar comparisons.