The Picard Theorems via Geometry

Dana P. Williams

August 25, 2015

Contents

1 The Picard Theorems

This discussion is taken from the first few chapters of Steven Krantz's Carus monograph "Complex Analysis: The Geometric Viewpoint" [Kra90]. Krantz's point is that by applying simple geometric techniques, the nasty analytic proofs of the Picard Theorems — which I've never been through — have elegant geometric analogues. Unfortunately, I don't have the time or expertise to put the proper motivation here. (See the full exposition in [Kra90] for a start.)

As a computation tool, we introduce the operators on continuously differentiable functions from $\mathbf{R}^2$ to itself:[1]

$$\frac{\partial}{\partial z} = \frac{1}{2}\Big(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\Big) \quad \text{and} \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\Big(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\Big).$$

[1]We are not requiring these functions to be holomorphic. In particular, we'll want to apply these to real-valued functions such as harmonic functions.

Of course, here, we view $\mathbf{R}^2$ as the complex plane. The Cauchy-Riemann equations tell us that if D is a domian, then $\frac{\partial}{\partial \bar{z}} f = 0$ on D if and only if f is holomorphic on D. On the other hand, if $f \in H(D)$, then $\frac{\partial}{\partial z} f = f'$. These operators are linear and obey the usual product rule. With a little bit of work, you can show that they act on polynomials in z and $\bar{z}$ exactly as you'd expect from the notation. But the chain rule is surprisingly complicated.

Proposition 1.1 ([Kra90, Proposition 1.2.1]). *Suppose that f and g are continuously differentiable such that $f \circ g$ is well-defined on a domain D. Then for all $z \in D$,*

$$\frac{\partial}{\partial z}(f \circ g)(z) = \frac{\partial f}{\partial z}\big(g(z)\big)\frac{\partial g}{\partial z}(z) + \frac{\partial f}{\partial \bar{z}}\big(g(z)\big)\frac{\partial \bar{g}}{\partial z}(z)$$

and

$$\frac{\partial}{\partial \bar{z}}(f \circ g)(z) = \frac{\partial f}{\partial z}\big(g(z)\big)\frac{\partial g}{\partial \bar{z}}(z) + \frac{\partial f}{\partial \bar{z}}\big(g(z)\big)\frac{\partial \bar{g}}{\partial \bar{z}}(z).$$

Proof. As in [Kra90], we simply use $\frac{\partial}{\partial z}(f \circ g) = \frac{1}{2}\big(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\big)(f \circ g)$ and then simplify using $\frac{\partial}{\partial x} = \frac{\partial}{\partial z} + \frac{\partial}{\partial \bar{z}}$ and $\frac{\partial}{\partial y} = i\big(\frac{\partial}{\partial z} - \frac{\partial}{\partial \bar{z}}\big)$. □

Lemma 1.2. *If either f or g is holomorphic, then*

$$\frac{\partial}{\partial z}(f \circ g)(z) = \frac{\partial f}{\partial z}\big(g(z)\big)\frac{\partial g}{\partial z}(z).$$

As an example of the utility of these operators, we note that the Laplace operator satisfies

$$\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = 4\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}} = 4\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z}.$$

Remark 1.3 (Harmonic Functions). Recall that a twice continuously differentiable function $u : D \to \mathbf{C}$ is called *harmonic* if $\Delta(u) = 0$.

Lemma 1.4. *Suppose that h is holomorphic on D and $f \circ h$ is defined on D. Then*

$$\Delta(f(h(z)) = \Delta(f \circ h)(z) = |h'(z)|^2 \Delta(f)\big(h(z)\big).$$

Proof. Well, since h is holomorphic we can apply Lemma 1.2 to conclude that

$$\begin{aligned}\Delta(f \circ h)(z) &= 4\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z}(f \circ h)(z) \\ &= 4\frac{\partial}{\partial \bar{z}}\Big(\frac{\partial f}{\partial z}(h(z))h'(z)\Big)\end{aligned}$$

which, by the product rule, is

$$= 4\Big(\frac{\partial}{\partial \bar{z}}\Big(\frac{\partial f}{\partial z}\circ h\Big)(z)h'(z) + \frac{\partial f}{\partial z}(h(z))\frac{\partial h'}{\partial \bar{z}}(z)\Big)$$

which, since $h' \in H(D)$, is

$$= 4\Big(\frac{\partial}{\partial \bar{z}}\Big(\frac{\partial f}{\partial z}\circ h\Big)(z)h'(z) + 0\Big)$$

which, using the chain rule formula for $\frac{\partial}{\partial \bar{z}}$, is

$$\begin{aligned} &= 4\Big(\Big(\frac{\partial}{\partial \bar{z}}\frac{\partial}{\partial z}(h(z))\frac{\partial h}{\partial \bar{z}}(z) + \frac{\partial}{\partial \bar{z}}\frac{\partial f}{\partial z}(h(z))\frac{\partial \bar{h}}{\partial \bar{z}}(z)\Big)h'(z)\Big) \\ &= 4\Big(0 + \frac{\partial}{\partial \bar{z}}\frac{\partial f}{\partial z}(h(z))\overline{h'(z)}\Big)h'(z)\Big) \\ &= \Delta(f)(h(z))|h'(z)|^2 \end{aligned}$$

□

1.1 Metrics and Curvature

We will refer to a nonzero real-valued function on a domain D as a *metric* on D. (The motivation for this terminology is carefully explained in [Kra90].) Although we don't really need to know, the odd terminology comes from the following. Let $\gamma : [a, b] \to D$ be a path in D. (Here as is usual in the subject, as path is denotes a piecewise smooth function $\gamma : [a, b] \to D$.) If σ is a metric on D, then we define

$$L_\sigma(\gamma) := \int_a^b \sigma(\gamma(t))|\gamma'(t)|\,dt.$$

Thus if σ is the *Euclidean metric* $\sigma \equiv 1$, then $L_\sigma(\gamma) = L(\gamma)$ is just the usual Euclidean length of γ. To get a bona fide metric from our "metric" σ, we define

$$d_\sigma(p, q) = \inf\{\, L_\sigma(\gamma) : \gamma \text{ is a path in } D \text{ from } p \text{ to } q \,\}.$$

It is an amusing exercise to see that in the case of the Euclidean metric, $d(p, q) = |p - q|$. In general, computing d_σ can be difficult as is the question of whether this is a minimal path γ from p to q such that $d_\sigma(p, q) = L_\sigma(\gamma)$. Here we'll just settle for the observation that d_σ is a metric on D.

If σ is a metric on D' and $f : D \to D'$ is continuously differentiable, then we define the pull-back metric on D to be $f^*\sigma$ where

$$f^*\sigma(z) = \sigma(f(z))\Big|\frac{\partial f}{\partial z}(z)\Big|.$$

If $f : D \to D'$ is holomorphic, so that in particular $\frac{\partial f}{\partial z}(z) = f'(z)$, then a straightforward calculation shows that

$$L_{f^*\sigma}(\gamma) = L_\sigma(f \circ \gamma).$$

Therefore,

$$d_\sigma(f(p), f(q)) \le d_{f^*\sigma}(p, q).$$

Remark 1.5 (Be Carefull). Actually, $f^*\sigma$ is only a metric (that is, nonzero) on the subset of D where $\frac{\partial f}{\partial z}(z) \neq 0$ and $\sigma(f(z)) \neq 0$.

We worry about zeros of our metrics because the key tool here is the *curvature* of one of our metrics.

Definition 1.6. Let ρ be a metric on D. Then the curvature of ρ on D is given by

$$\kappa_{(D,\rho)}(z) := \frac{-\Delta\big(\ln(\rho)\big)(z)}{\rho(z)^2}.$$

Remark 1.7. It is not so easy to see the geometric significance of "curvature" as defined above. Even Krantz punts on trying to motivate it. But as we shall see, it is indeed a useful concept — especially for our purposes here.

Example 1.8. If $\rho(z) = 1$ for all z, then $\kappa_\rho(z) = 0$ for all z. That is, Euclidean space is not curved.

Lemma 1.9. *Let $D = B_r(0)$ be the ball of radius $r > 0$ centered at the origin. Let*

$$\rho_r(z) = \frac{r}{r^2 - |z|^2}.$$

Then ρ_r has constant curvature -4. In particular, the Poincaré metric

$$\rho(z) = \frac{1}{1 - |z|^2}$$

on the unit disk has constant curvature -4.

Proof. We compute as follows:

$$\begin{aligned} -\Delta(\ln(p_r))(z) &= \Delta\big(\ln(r^2 - z\bar{z})\big) \\ &= 4\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}}\big(\ln(r^2 - z\bar{z})\big) \end{aligned}$$

$$
\begin{aligned}
&= 4\frac{\partial}{\partial z}\Big(\frac{-z}{r^2 - z\bar{z}}\Big) \\
&= 4\Big(\frac{(-1)(r^2 - z\bar{z}) - (-z)(-\bar{z})}{(r^2 - z\bar{z})^2}\Big) \\
&= \frac{-4r^2}{(r^2 - z\bar{z})^2}.
\end{aligned}
$$

The result follows (together with the observation that the Poincaré metric is ρ_1). $\square$

It is not hard to check that if σ is the metric obtained from ρ by multiplication by α, then $\kappa_\sigma(z) = \alpha^2 \kappa_\rho(z)$. Hence, if $A > 0$, then we can get a metric of constant curvature $-A$ on $D = B_r(0)$ via

$$
\rho_r^A = \frac{2r}{\sqrt{A}(r^2 - |z|^2)}.
$$

1.2 The Schwartz Lemma

The Schwartz Lemma that arises in most undergraduate courses says that if $f : U \to U$ is holomorphic (where $U = \{ z : |z| < 1 \}$) and $f(0) = 0$, then $|f(z)| \leq |z|$ for $z \in U$ and $|f'(0)| \leq 1$.

The genius of Ahlfors and the enlightening exposition of Krantz tells us the following is a generalization of the Schwartz Lemma. Furthermore, it is the key to the proofs of the Picard theorems.

Theorem 1.10 ([Kra90, Theorem 2.1.4])**.** *Suppose that D is domain equipped with a metric σ whose curvature is bounded above by $-B$ for some $B > 0$. If $f : B_r(0) \to D$ is holomorphic, then*

$$
f^*\sigma(z) \leq \frac{\sqrt{A}}{\sqrt{B}} \rho_r^A(z) \quad \textit{for all } z \in B_r(0).
$$

Before proceeding with the proof and seeing why we would care about this result, I'd like to take a moment to see what it is a generalization of the Schwartz Lemma. We specialize to the case $r = 1$ and $f : U \to U$ is a holomorphic function from the unit disk $U = \{ z : |z| < 1 \}$ to itself. We also let σ be the Poincare metric. Then the conclusion of Theorem 1.10 is that

$$
f^*\rho(z) \leq \rho(z) \quad \text{for all } z \in U.
$$

Then plugging 0 into

$$\frac{|f'(z)|}{1-|f(z)|^2} \leq \frac{1}{1-|z|^2},$$

gives us $|f'(0)| \leq 1$ even without the classical assumption that $f(0) = 0$. Furthermore, as observed above, f must also be d_ρ-distance reducing. As I'll show in Section 1.7 on page 18,

$$d_\rho(0,z) = \frac{1}{2}\ln\Big(\frac{1+|z|}{1-|z|}\Big). \tag{1.1}$$

Then, using $f(0) = 0$, we have $d_\rho(0, f(z)) \leq d_\rho(0,z)$. Since $x \mapsto \ln(\frac{1+x}{1-x})$ is increasing, we get $|f(z)| \leq |z|$.[2]

For the proof of our genearlized Schwartz Lemma (Theorem 1.10), we do need the observation in the next lemma.[3] Note that it does require some nasty chain rule calculations.

Lemma 1.11. *Suppose that $f : D \to D'$ is holomorphic and that ρ is a metric on D'. Suppose that $\rho \circ f$ and f' never vanish on D. Then*

$$\kappa_{(D,f^*\rho)}(z) = \kappa_{(D',\rho)}(f(z)).$$

Proof. We simply calculate using $\ln(f^*\rho(z)) = \ln\big(\rho(f(z)\big) + \ln(|f'(z)|)$,

$$\kappa_{f^*\rho}(z) = \frac{-\Delta(\ln(\rho(f(z)))) - \Delta(\ln(|f'(z)|))}{\big(\rho(f(z))|f'(z)|\big)^2}$$

which, since $\ln(|f'(z)|)$ is harmonic, is

$$= \frac{-\Delta(\ln(\rho(f(z))))}{\rho(f(z))^2|f'(z)|^2}$$

which, by Lemma 1.4, is

$$= \frac{-\Delta(\ln(\rho))(f(z))}{\rho(f(z))^2}$$
$$= \kappa_\rho(f(z)).$$

□

[2]Note that if there is a $z_0 \in U$ such that $|f(z_0)| = |z_0|$, then the Maximum Modulus Theorem implies that f is a rotation: $f(z) = e^{i\theta}z$. (This is also usually part of the Schwartz Lemma in undergraduate courses.) It's also true that f is a rotation if $|f'(0)| = 1$. But I don't see how to show that yet.

[3]If you believe the curvature really does measure the geometry, and your recall that holomorphic maps are conformal on domains where their derivative doesn't vanish, then the result is not very surprising.

Proof of Theorem 1.10. Let $0 < r' < r$ and let

$$\nu(z) = \frac{f^*\sigma(z)}{\rho^A_{r'}(z)}.$$

Then ν is continuous and $\nu(z) > 0$ for $z \in B_{r'}(0)$. Since $|\rho^A_{r'}(z)| \to \infty$ as $|z| \nearrow r'$, it follows that $\nu(z) \to 0$ as $|z| \nearrow r'$. Hence ν attains a maximum value, $M_{r'}$ at a point $\tau \in B_{r'}(0)$. If we can show that $M_{r'} \leq \frac{\sqrt{A}}{\sqrt{B}}$, then letting $r' \nearrow r$, we can conclude that $\frac{f^*\sigma}{\rho^A_r} \leq \frac{\sqrt{A}}{\sqrt{B}}$ and we're done.

If $f^*\sigma(\tau) = 0$, then $\nu \equiv 0$ and we have nothing to show. So we can assume that $f^*\sigma(\tau) > 0$. Then $\kappa_{f^*\sigma}$ is defined near τ. In particular, the hypotheses of Lemma 1.11 are met and so our hypotheses imply that

$$\kappa_{f^*\sigma}(\tau) = \kappa_\sigma(f(\tau)) \leq -B.$$

On the other hand, $z \mapsto \ln \nu(z)$ has a maximum at τ. By the second derivative test,

$$\begin{aligned} 0 &\geq \Delta(\ln(\nu))(\tau) \\ &= \Delta(\ln(f^*\sigma))(\tau) - \Delta(\ln(\rho^A_{r'}))(\tau) \\ &= -\kappa_{f^*\sigma}(\tau) \cdot (f^*\sigma(\tau))^2 + \kappa_{\rho^A_{r'}}(\tau) \cdot (\rho^A_{r'}(\tau))^2 \\ &\geq B \cdot (f^*\sigma(\tau))^2 - A \cdot (\rho^A_{r'}(\tau))^2. \end{aligned}$$

Hence

$$f^*\sigma(\tau) \leq \frac{\sqrt{A}}{\sqrt{B}} \rho^A_{r'}(\tau).$$

Thus $M_{r'} \leq \frac{\sqrt{A}}{\sqrt{B}}$ as required. □

1.3 The Little Picard Theorem

To see what our Theorem 1.10 on page 5 has to do with entire functions, we have the following.

Theorem 1.12. *Let D be a domain that admits a metric σ with negative curvature κ_σ bounded away from 0. (That is, $\kappa_\sigma(z) \leq -B < 0$ for all $z \in D$.) Then any entire function f with $f(\mathbf{C}) \subset D$ must be constant.*

Proof. Let σ be a metric on D with $\kappa_\sigma(z) \leq -B < 0$ for all $z \in D$. Fix $z \in \mathbf{C}$. Fix $r > |z|$ and consider $f : B_r(0) \to D$. Then for any $A > 0$, Theorem 1.10 implies that

$$f^*\sigma(z) \leq \frac{\sqrt{A}}{\sqrt{B}}\rho_r^A(z) \quad \text{for all } z \in B_r(0).$$

But this holds for all $r > |z|$ and $\lim_{r\to\infty} \rho_r^A(z) = 0$, so $f^*\sigma(z) \leq 0$. But this just means $f^*\sigma(z) = 0$. But this only holds if $f'(z) = 0$. But z was arbitrary. Hence f is constant. □

Of course if f where a bounded entire function, then its range would be included in a disk $D = B_R(0)$. Since the latter admits metrics of constant negative curvature, we see that the classic Liouville Theorem is a consequence of Theorem 1.12.

On the other hand, the exponential function $z \mapsto to = e^z$ has range contained in $\mathbf{C} \setminus \{0\}$. So we can conclude that neither $D = \mathbf{C} \setminus \{0\}$ nor $D = \mathbf{C}$ admit a metric with negative curvature bounded away from 0.

We now come to the crux of the matter.

Lemma 1.13. *Let D be a domain whose complement in $\mathbf{C}$ contains at least two points. Then D admits a metric σ such that*

$$\kappa_\sigma(z) \leq -B < 0 \quad \textit{for all } z \in D.$$

Remark 1.14. Of course, given such a μ, we can multiply μ by a positive scalar so that we can take B as large as we like.

The proof is very unsatisfying. It consists of writing down a metric and checking that its curvature is negative and bounded away from 0. However, once we do that, we obtain our goal.

Theorem 1.15 (The Little Picard Theorem)**.** *Let f be an entire function whose range omits at least two points. Then f is constant.*

Ok, time to pay the piper.

Proof of Lemma 1.13. Let z_1 and z_2 be distinct points in the complement of D. Let $\mathbf{C}_{0,1} := \mathbf{C} \setminus \{0, 1\}$ and define $f : D \to \mathbf{C}_{0,1}$ by

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

In view of Lemma 1.11, it will suffice to produce a metric μ on $\mathbf{C}_{0,1}$ such that $\kappa_\mu \leq -B < 0$ and pull-back via f.

Krantz gives some motivation for our choice of μ in a short remark on page 80 of [Kra90], but here we're just going to write it down and calculate. Let

$$\mu(z) := \Big(\frac{(1+|z|^{\frac{1}{3}})^{\frac{1}{2}}}{|z|^{\frac{5}{6}}}\Big) \cdot \Big(\frac{(1+|z-1|^{\frac{1}{3}})^{\frac{1}{2}}}{|z-1|^{\frac{5}{6}}}\Big).$$

Recall that $u(z) = \ln(|z|)$ is harmonic. Hence $\Delta(\alpha u(z)) = \Delta(\ln(|z|^{\alpha})) = 0$ for $\alpha > 0$. Hence we can calculate

$$\begin{aligned}\Delta\Big(\ln\Big(\frac{(1+|z|^{\frac{1}{3}})^{\frac{1}{2}}}{|z|^{\frac{5}{6}}}\Big)\Big) &= \frac{1}{2}\Delta\big(\ln(1+|z|^{\frac{1}{3}})\big)\\ &= 2\frac{\partial}{\partial z}\frac{\partial}{\partial \bar{z}}\big(\ln(1+|z|^{\frac{1}{3}})\big)\end{aligned}$$

which, after some work, is

$$= \frac{1}{18}\frac{1}{|z|^{\frac{5}{3}}(1+|z|^{\frac{1}{3}})^2}.$$

Similarly,

$$\Delta\Big(\ln\Big(\frac{(1+|z-1|^{\frac{1}{3}})^{\frac{1}{2}}}{|z-1|^{\frac{5}{6}}}\Big)\Big) = \frac{1}{18}\frac{1}{|z-1|^{\frac{5}{3}}(1+|z-1|^{\frac{1}{3}})^2}$$

Consequently,

$$\begin{aligned}\kappa_\mu(z) &= \frac{-\Delta\big(\ln(\mu)\big)(z)}{\kappa_\mu(z)^2}\\ &= -\frac{1}{18}\Big(\frac{|z-1|^{\frac{5}{3}}}{(1+|z|^{\frac{1}{3}})^3(1+|z-1|^{\frac{1}{3}})}\Big) + \Big(\frac{|z|^{\frac{5}{3}}}{(1+|z|^{\frac{1}{3}})(1+|z-1|^{\frac{1}{3}})^3}\Big).\end{aligned}$$

Now observe that

(a) $\kappa_\mu(z) < 0$ for all $z \in \mathbf{C}_{0,1}$.

(b) $\lim_{z\to 0} \kappa_\mu(z) = -\frac{1}{36}$.

(c) $\lim_{z\to 1} \kappa_\mu(z) = -\frac{1}{36}$ and

(d) $\lim_{z\to\infty} \kappa_\mu(z) = -\infty$.

In view of (b) and (c), we can view κ_μ as a continuous function on $\mathbf{C}$. In view of (a) and (d), the function must attain its minimum. Hence the result. □

1.4 Normal Families

While our goal is The Great Picard Theorem, this path takes us through normal families. I'll follow [Kra90] and say that a sequence $\{f_k\}$ of complex-valued functions on a domain D *converges normally* to a function f on D if $f_k \to f$ uniformly on all compact subsets of D. The sequence $\{f_k\}$ is *compactly divergent* on D if f_k diverges to infinity uniformly on compact subsets of D.

Definition 1.16. Let D be a domain. A family $\mathscr{F}$ of complex-valued functions on D is called a *normal family* on D if every sequence if $\mathscr{F}$ has a subsequence which is either normally convergent or compactly divergent.

Remark 1.17. This isn't quite the same as defined in, for example, [Rud87, Chap. 14] — the option for compact divergence is not usually included. The current definition will make it easier to work on the Riemann sphere, and will obviously be necessary for our generalizations of Montel's Theorem: compare Theorem 1.18, Remark 1.19 and Theorem 1.28 on page 15.

Normal families usually arise via Montel's Theorem:

Theorem 1.18 (Montel — [Rud87, Theorem 14.6])**.** *Suppose that D is a domain and $\mathscr{F} \subset H(D)$. If $\mathscr{F}$ is uniformly bounded on compact subsets of D, then $\mathscr{F}$ is a normal family.*

Remark 1.19. Note that in this case, no subsequence of $\mathscr{F}$ can be compactly divergent. Hence every sequence in $\mathscr{F}$ has a normally convergent subsequence. The proof just amounts to a clever argument showing that $\mathscr{F}$ is equicontinuous. Since $\mathscr{F}$ is obviously pointwise bounded, we can apply Arzela-Ascoli. For the details, see Section 1.6 on page 16.

Now we want to introduce the Riemann sphere, $\mathbf{C}^+$, into the mix. Rather than think of $\mathbf{C}^+$ as S^2, in these notes $\mathbf{C}^+$ is meant to denote the one-point compactification of $\mathbf{C}$.[4] Recall that a function on a domain D is called *meromorphic* if it is holomorphic on D with the possible exception of poles (necessarily isolated). Clearly we can view a meromorphic function on D as a function from D to $\mathbf{C}^+$.

[4]Thus $\mathbf{C}^+$ is the compact Hausforff space $\mathbf{C} \cup \{\infty\}$ where $\mathbf{C}$ is open and U is an open neighborhood of ∞ if and only if $\mathbf{C}^+ \setminus U$ is compact in $\mathbf{C}$.

1.4.1 The Spherical Metric

When working with Meromorphic functions, it seems to be very useful to consider the *spherical metric* on $\mathbf{C}$ is given by

$$\tau(z) = \frac{2}{1+|z|^2}.$$

Remark 1.20. Although we don't need it, I'll show in Section 1.7 on page 18 that

$$d_\tau(z,w) = 2\arctan\Big(\frac{|z-w|}{|1-\bar{z}w|}\Big). \tag{1.2}$$

(I found it much harder to verify this than Krantz seems to suggest in the text.) I'll also verify that $d_\tau(p,q)$ is the "great circle distance" from P' and Q' on the Riemann sphere where P' and Q' are the images of p and q on the Riemann sphere, respectively.

I'd like to extend d_τ to $\mathbf{C}^+$. We say that a path from $z \in \mathbf{C}$ to $\infty \in \mathbf{C}^+$ is a peicewise smooth function $\gamma : [0,1) \to \mathbf{C}$ such that $\gamma(0) = z$ and $\lim_{t\nearrow 1} |\gamma(t)| = \infty$.[5] A path from ∞ to z is defined similarly. For example, if $z \in \mathbf{C} \setminus \{0\}$, then let $\gamma(t) = \frac{z}{1-t}$. Then $\gamma : [0,1) \to \mathbf{C}$ is a path from z to ∞ and

$$d_\tau(z,\infty) \le L_\tau(\gamma) = \int_0^1 \frac{2|z|}{(1-t)^2+|z|^2}\,dt \le \frac{2}{|z|}.$$

Hence we can define $d_\tau(\infty,z)$ and $d_\tau(z,\infty)$ exactly as above. In particular, for convenience down the road, we can allow paths that "pass through ∞" in the obvious sense.[6] It is fairly clear that d_τ is finite on $\mathbf{C}^+$ and that it is a metric on $\mathbf{C}^+$.[7]

Notice that $d_\tau(z,w) \le 2|z-w|$. Hence the map $z \mapsto z$ from $\mathbf{C} \subset \mathbf{C}^+$ to the metric space $(\mathbf{C}^+, d_\tau)$ is continuous. But if $z_n \to \infty$ in $\mathbf{C}^+$, then by the above $d_\tau(z_n,\infty) \le \frac{2}{|z_n|}$ and it follows that the identity map from $\mathbf{C}^+$ to $(\mathbf{C}^+, d_\tau)$ is a continuous bijection and hence a homeomorphism.

Definition 1.21. A family $\mathscr{F}$ of meromorphic functions on a domain $D \subset \mathbf{C}$ is called a normal family if every sequence of elements from $\mathscr{F}$ has a subseqence which is normally convergent in $(\mathbf{C}^+, d_\tau)$.

[5] Of course, we can use any half closed interval.

[6] Note that we need only consider curves that pass through ∞ or any other point at most once: we are interested in distance minimizing curves and can clearly eliminate any loop.

[7] Using (1.2), we have

$$d_\tau(z,\infty) = \lim_{w\to\infty} d_\tau(z,w) = 2\arctan(|z|^{-1}).$$

Comparing our two definitions of normal family — Definition 1.16 and Definition 1.21 — should cause annoyance. Nevertheless, we have the following which is incredibly useful and nontrivial. Krantz does not truly justify it in the book — in my humble opinion.

Proposition 1.22. *Let $\mathscr{F}$ be a family of holomorphic functions on a domain D. Then $\mathscr{F}$ is a normal family as in Definition 1.16 if and only if $\mathscr{F}$ is a normal family considered as meromorphic functions as in Definition 1.21.*

To prove Proposition 1.22, we'll need some observations. The first is a classic application of the Argument Principal.

Lemma 1.23 (Hurwitz's Theorem)**.** *Suppose that f_n is a sequence of holomorphic functions converging normally to f on a domain D. If each f_n is zero free, then either f is zero free or identically zero.*

The second is a strange artifact of d_τ.[8]

Lemma 1.24. *With the usual conventions regarding $1/\infty$ and $1/0$, we have*

$$d_\tau(z,w) = d_\tau\Big(\frac{1}{z}, \frac{1}{w}\Big)$$

for all $z, w \in \mathbf{C}^+$.

Proof. Let γ be a path from z to w. We can assume that γ passes through 0 or ∞ at most once. Then $w(t) = 1/\gamma(t)$ is a path and

$$\begin{aligned} L_\tau(w) = \int_0^1 \frac{2}{1+|w(t)|^2}|w'(t)|\,dt &= \int_0^1 \frac{2|\gamma(t)|^2}{1+|\gamma(t)|^2}\Big|\frac{-1}{\gamma(t)^2}\gamma'(t)\Big|\,dt \\ &= \int_0^1 \frac{2}{1+|\gamma(t)|^2}|\gamma'(t)|\,dt \\ &= L_\tau(\gamma). \end{aligned}$$

The result follows. □

Lemma 1.25. *Let $\{f_n\} \subset H(D)$ and let $f \in H(D)$. Then $f_n \to f$ normally in $\mathbf{C}$ with the usual metric if and only if $f_n \to f$ normally in $(\mathbf{C}, d_\tau)$.*

[8] Of course, Lemma 1.24 follows from (1.2), but it is nice to have a proof without that overhead.

Proof. Suppose that $f_n \to f$ normally with respect to d_τ. If the assertion were false, then there is a compact set $K \subset D$ such that $f_n \not\to f$ uniformly on K. Thus there is an $\epsilon > 0$ such that for all n there is a $m_n \geq n$ and a $z_n \in K$ such that

$$|f_{m_n}(z_n) - f(z_n)| \geq \epsilon. \tag{1.3}$$

We can certainly arrange that $m_n \leq m_{n+1}$. Since K is compact, we can also assume that $z_n \to z_0$ in K. But the uniform convergence of $\{f_{m_n}\}$ with respect to d_τ implies that $f_{m_n}(z_n) \to f(z_0)$ with respect to d_τ. Hence $f_{m_n}(z_n) \to f(z_0)$ in $\mathbf{C}$ which contradicts (1.3).

The other direction is immediate since $d_\tau(z, w) \leq 2|z - w|$. □

Lemma 1.26. *A sequence $\{f_n\}$ diverges compactly on D if and only if $f_n \to f$ normally in $(\mathbf{C}^+, d_\tau)$ where $f(z) = \infty$ for all $z \in D$.*

Proof. Suppose that $\{f_n\}$ diverges compactly. Let $K \subset D$ be compact. If $f_n \not\to f$ uniformly on K, then there is an $\epsilon > 0$ such that for all n there is a $m_n \geq n$, and $z_n \in K$ such that $d_\tau(f_{m_n}(z_n), \infty) \geq \epsilon$. But we can assume that $z_n \to z \in K$ and that $m_n \leq m_{n+1}$. Hence $f_{m_n}(z_n) \to infty$ in $\mathbf{C}$. Hence $f_{m_n}(z_n) \to \infty$ in $(\mathbf{C}^+, d_\tau)$. This is contradiction.

Now suppose that $f_n \to f$ normally and $K \subset D$ is compact. If $\{f_n\}$ does not diverge to infinity uniformly on K, then there is $M > 0$ such that for all n there is a $m_n \geq n$ and $z_n \in K$ such that $|f_{m_n}(z_n)| \leq M$. But we can assume that $z_n \to z \in K$ and that $m_n \leq m_{n+1}$. Hence $f_{m_n}(z_n) \to f(z) = \infty$ in $(\mathbf{C}^+, d_\tau)$. But this means $f_{m_n}(z_n) \to \infty$ in $\mathbf{C}$. This is a contradiction. □

Proof of Proposition 1.22. Suppose that $\mathscr{F}$ is a normal family of holomorphic functions. We want to see that $\mathscr{F}$ is a normal family of meromorphic functions. Let $\{f_n\}$ be a sequence in $\mathscr{F}$. Then $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ that either converges normally to a function $f \in H(D)$ or diverges compactly. In the first case, Lemma 1.25 implies that $f_{n_k} \to f$ normally in $(\mathbf{C}^+, d_\tau)$. Hence we need to see that if $\{f_{n_k}\}$ diverges compactly, then $f_{n_k} \to f$ normally in $\mathbf{C}^+$ where f is the function which is identically infinite. This is Lemma 1.26.

For the converse, it will suffice to see that if $\{f_n\} \subset \mathscr{F}$ converges normally to f in $(\mathbf{C}^+, d_\tau)$ then either f is the constant function $f(z) = \infty$ and f_n is compactly divergent, or f is holomorphic and $f_n \to f$ uniformly on K with respect to the usual metric. In view of Lemma 1.25 on the preceding page, we only have to worry about the case where f is not everywhere finite valued.

Suppose that $f(z_0) = \infty$. Then $f(z) \neq 0$ in an open (connected) neighborhood U of z_0. By taking $n \geq N$, there is no harm in assuming each f_n is zero free in U.

It follows from Lemma 1.24, then $1/f_n \to 1/f$ normally on U with respect to d_τ. Arguing as above, $1/f_n \to 1/f$ normally on U with respect to the usual metric. Since the f_n are holomorphic, $1/f_n$ is zero free. Since $1/f$ has a zero, we see that $1/f$ is identically zero in U. Hence the set where f takes the value ∞ is open. It is also clearly closed. The result follows from Lemma 1.26 on the previous page. □

If f is a meromorphic function on D, then we define

$$f^\sharp(z) = \begin{cases} \dfrac{2|f'(z)|}{1+|f(z)|^2} & \text{if } z \text{ is not a pole, and} \\ \lim_{w\to z} f^\sharp(w) & \text{otherwise.} \end{cases}$$

Using the fact that if f has a pole at p, then $f(z) = g(z)/(z-p)^m$ for some m and some holomorphic g with $g(p) \neq 0$, it follows that if f has a pole at p, then $f^\sharp(p)$ equals 0 if $m \geq 2$ and $2/\operatorname{Res}(f;p)$ otherwise. In particular, $f^\sharp$ is always finite-valued for any meromorphic function f.

Theorem 1.27 (Marty)**.** *Let $\mathscr{F}$ be a family of meromorphic functions on a domain D. Then $\mathscr{F}$ is a normal family of meromorphic functions if and only if for each compact subset $K \subset D$ there is a constant M_K such that*

$$f^\sharp(z) \leq M_K \quad \textit{for all } z \in K \textit{ and } f \in \mathscr{F}.$$

Proof. First, suppose that M_K exist as in the statement of the result. Fix $K \subset D$. By covering K with closed balls, we can assume that there is a compact set $K \subset K'$ and $\delta > 0$ so that $B_\delta(z) \subset K'$ for all $z \in K$. In particular, if $z, w \in K$ and $|z-w| < \delta$, then the line segment $[z,w] \subset K'$. Then $f \circ \gamma$ is a path from $f(z)$ to $f(w)$, and

$$\begin{aligned} d_\tau(f(z), f(w)) \leq L_\tau(f\circ\gamma) &= \int_0^1 \frac{2}{1+|f(\gamma(t))|^2} \big|f'(\gamma(t))\gamma'(t)\big|\,dt \\ &= \int_0^1 \frac{|f'(\gamma(t))|}{1+|f(\gamma(t))|^2} |\gamma'(t)| \\ &\leq M_{K'} \int_0^1 |\gamma'(t)| \\ &= M_{K'} L(\gamma) = M_{K'}|z-w|. \end{aligned}$$

Since $M_{K'}$ does not depend on f, we seem that $\mathscr{F}$ is an equicontinuous family of functions from K with the usual metric to the compact metric space $(\mathbf{C}^+, d_\tau)$. Hence a variation of the usual Arzela-Ascoli Theorem applies — see Theorem 1.32 on page 17.

Now for the converse, assume $\mathscr{F}$ is a normal family of meromorphic functions on D. Suppose to the contrary, that there is a compact set $K \subset D$ such that $\{ f^\sharp : f \in \mathscr{F} \}$ is unbounded. Let $\{f_n\} \subset \mathscr{F}$ be such that $\|f_n^\sharp\|_{\infty,K} \nearrow \infty$. By assumption on $\mathscr{F}$, we can pass to a subsequence, relabel, and assume that there is a function f such that $f_n \to f$ normally on D. I claim that each point $z \in K$ has a neighborhood U_z such that $f_n^\sharp \to f^\sharp$ normally on U_z. If $f(z) \neq \infty$, then let U_z be such that $f \in H(U_z)$. Then using basic complex analysis and the proof of Proposition 1.22, we see that $f_n' \to f'$ normally on U_x. The claim follows easily from this. If $f(z) = \infty$, then we can choose U_z so that for large n, f_n is nonzero. Then $1/f_n$ is homomorphic and $1/f_n \to 1/f$ normally (using Lemma 1.24). Thus we can assume that $1/f \in H(U_z)$. Hence by the above, $\big(\frac{1}{f_n}\big)^\sharp \to \big(\frac{1}{f}\big)^\sharp$. But $(\frac{1}{g})^\sharp = g^\sharp$. This proves the claim, and completes the proof. □

Here is our big result on normal families — all due to Paul Montel.

Theorem 1.28 (Montel)**.** *Let D be a domain in $\mathbf{C}$ and $\mathscr{F}$ a family of meromorphic functions on D whose ranges omit three distinct points P, Q and R in $\mathbf{C}^+$. Then $\mathscr{F}$ is a normal family of meromorphic functions.*

Proof. Applying a linear fractional transformation, we can assume $P = 0$, $Q = 1$ and $R = \infty$. Thus, we assume the functions in $\mathscr{F}$ are holomorphic and take values in $\mathbf{C}_{0,1} := \mathbf{C} \setminus \{0, 1\}$.

It will suffice to see that $\mathscr{F}$ is normal on any disk $B_r(z_0)$.[9] There is no real harm in assuming $z_0 = 0$. (Just translate D by $-z_0$.) Let μ be a metric on $\mathbf{C}_{0,1}$ such that $\kappa_\mu \leq -4$ for all $z \in \mathbf{C}_{0,1}$ (multiply the metric from Lemma 1.13 by an appropriate constant). Then Theorem 1.10 implies (letting $A = B = 4$) implies that

$$f^*\mu(z) \leq \rho_r^4(z) \quad \text{for all } z \in B_r(0).$$

But if $\tau(z) = 2/(1 + |z|^2)$, then $\tau(z)/\mu(z)$ tends to 0 if z tends to either 0, 1 or ∞. Thus there is a constant $M > 0$ such that

$$\tau(z) \leq M \cdot \mu(z) \quad \text{for all } z \in \mathbf{C}_{0,1}.$$

Thus

$$f^\sharp \equiv f^*\tau \leq M \cdot f^*\mu \leq M \cdot \rho_r^4 \quad \text{for all } z \in B_r(0).$$

[9] If K is compact in D, then we can cover K by open balls $B_r(z_i)$ such that $B_{2r}(z_i) \subset D$. Then K is the union of the compact sets $K_i = K \cap \overline{B_r(z_i)}$. If $\{f_n\} \subset \mathscr{F}$, then pass to repeated subsequences that converge uniformly on the $K_i \subset B_{2r}(z_i)$.

The constant M is independent of $\mathscr{F}$ and since ρ_r^4 is continuous on $B_r(0)$, we have shown that $f^\sharp$ is bounded on each compact set of $B_r(0)$ independent of f. Now we apply Marty's Theorem 1.27 to conclude that $\mathscr{F}$ is normal. □

In the above proof, by virtue of Proposition 1.22, I've been sloppy with my use of normal families for holomorphic functions vs. meromorphic functions.

Corollary 1.29. *Let $\mathscr{F}$ be a family of holomorphic functions on a domain D whose ranges omit two values. Then $\mathscr{F}$ is a normal family.*

1.5 And Now Picard

Theorem 1.30. *Let f be holomorphic in the punctured disk $B_r'(0)$ and suppose that f has an essential singularity at 0. Then $f(B_r'(0))$ is either all $\mathbf{C}$ or $\mathbf{C} \setminus \{z_0\}$ for some z_0.*

Proof. Replacing f by $z \mapsto f(z/r)$, we can assume $D' := B_r'(0) = B_1(0)$. It will suffice to show that if $f(D') \subset \mathbf{C}_{0,1}$, then f has either a removable singularity at 0 or a pole at 0.

For $z \in D'$, let $f_n(z) = f(z/n)$ and let $\mathscr{F} = \{f_n\}$. Since $\mathscr{F}$ takes values in $\mathbf{C}_{0,1}$, $\mathscr{F}$ is normal. Thus $\{f_n\}$ has a subsequence $\{f_{n_k}\}$ which converges normally to $g \in H(D')$ or diverges compactly.

In the first case, we would have $\{f_{n_k}\}$ bounded on compact subsets of D'. In particular, there is a $M > 0$ such that the $|f_{n_k}|$ is bounded by M on $\{ z : |z| = \frac{1}{2} \}$. This means f is bounded by M on the circles $\{ z : |z| = \frac{1}{2n_k} \}$. By the maximum modulus principal, f is bounded on $\{ z : 0 < |z| < 1/2n_1 \}$. Then 0 is a removable singularity.

In the second case, a similar argument implies that $1/f$ has a removable singularity and a zero at 0. Thus, f has a pole at 0. This completes the proof. □

Corollary 1.31 (The Great Picard Theorem)**.** *If f has an essential singularity at z_0, then with possibly one exception, f attains every complex value infinitely often in every deleted neighborhood of z_0.*

1.6 A Generalized Arzela-Ascoli Theorem

No doubt there are numerous generalizations of Arzela-Ascoli out there. Here we'll settle for just enough for our purposes. I got the just of the argument from [Kna05, Theorem 10.48].

Theorem 1.32. *Suppose that (X,ρ) and (M,d) are compact metric spaces and that $\mathscr{F} \subset C(X,M)$ is equicontinuous. Then every sequence in $\mathscr{F}$ has a uniformly convergent subsequence.*

Proof. We assume that $\{f_n\}$ contains infinitely many distinct functions. (Otherwise, the assertion is trivial.) By Tychonoff, $C := \prod_{x\in X} C$ is compact in the product topology (a.k.a. the topology of pointwise convergence). Let $S \subset C$ be the image of $\{f_n\}$ in C. If S were closed in C, then $O_n := \{ f \in C : f \neq f_n \} \setminus S$ is an open subset of C and $\{O_n\}$ is an open cover of the compact space C with no finite subcover. This is a contradiction. Hence there exists $g \in \overline{S} \setminus S$.

I claim $\{g\} \cup \{f_n\}$ is equicontinuous. (So, in particular, g is continuous.) To see this, let $\epsilon > 0$ and $U_{x,\epsilon}$ a neighborhood such that

$$d\big(f(y), f(x)\big) \leq \epsilon \quad \text{if } y \in U_{x,\epsilon} \text{ and } f \in \mathscr{F}.$$

But $C(y) := \{ h \in C : d\big(h(y), f(x)) \leq \epsilon\big) \}$ is closed in C for any $y \in X$. Hence

$$F := \{ h \in C : d\big(h(y), f(x)\big) \leq \epsilon \text{ for all } y \in U_{x,\epsilon} \} = \bigcap_{y\in U_{x,\epsilon}} C(y)$$

is closed in C and $\mathscr{F} \subset F$. Therefore $g \in F$ and

$$d\big(g(y) - g(x)\big) \leq 2\epsilon \quad \text{if } y \in U_{x,\epsilon}.$$

The claim follows.

To complete the proof, it will suffice, given $\epsilon > 0$ and k, to find $N \geq k$ such that

$$d\big(f_N(x), g(x)\big) < \epsilon \quad \text{for all } x \in X.$$

But if $x \in X$, then there is a neighborhood U_x such that $y \in U_x$ implies that

$$d\big(f_n(y), f_n(x)\big) < \frac{\epsilon}{3} \quad \text{and} \quad d\big(g(y), g(x)\big) < \frac{\epsilon}{3}.$$

Since X is compact, there are $U_{x_1}, \dots, U_{x_n}$ which cover X. Then given $y \in X$, there is a x_j such that

$$d\big(f_n(y), f_n(x_j)\big) < \frac{\epsilon}{3} \quad \text{and} \quad d\big(g(y), g(x_j)\big) < \frac{\epsilon}{3}.$$

Since $g \in \overline{S}$, there is an $N \geq k$ such that

$$d\big(f_N(x_k), g(x_k)\big) < \epsilon 3 \quad \text{for } k = 1, 2, \dots, n.$$

But then for the appropriate j,

$$d\big(f_N(y), g(y)\big) \leq d\big(f_N(y), f_N(x_j)\big) + d\big(f_N(x_j), g(x_j)\big) + d\big(g(x_j), g(y)\big) < \epsilon. \qquad \square$$

Corollary 1.33. *Suppose that* (X,ρ) *is a second countable locally compact measure space, that* (C,d) *a compact measure space and that* $\mathscr{F} \subset C(X,M)$ *is equicontinuous. Then every sequence* $\{f_n\}$ *in* $\mathscr{F}$ *has a subsequence that converges uniformly on compact subsets of* X *to a function* $f \in C(X,M)$.

Proof. We can let $X = \bigcup K_n$ with K_n compact and $K_n \subset K_{n+1}^o$. Then every compact subset of X is contained in some K_n^o. So it will suffice to produce a function g and a subsequence converging uniformly to g on each K_n. We can use the previous theorem to find a subsequence $\{g_n^{(1)}\}$ such that $g_n^{(1)} \to g^{(1)}$ uniformly on K_1. Continuing inductively, we can construct a subsequence $\{g_n^{(k)}\}$ of $\{g_n^{(k-1)}\}$ such that $g_n^{(k)}$ converges uniformly to $g^{(k)}$ on K_k.

Clearly, $g^{(k)}(x) = g^{(k-1)}(z)$ if $z \in K_{k-1}$. Hence we get $g \in C(X,M)$ by defining $g(z) = g^{(k)}(z)$ if $z \in K_k$. Let $f_{n_k} = g_k^{(k)}$. Fix K_r. Then given $\epsilon > 0$, there is a $N \geq r$ such that $n \geq N$ implies

$$d\big(g_n^{(r)}(z), g(z)\big) < \epsilon \quad \text{for all } z \in K_r.$$

Then $k \geq N$ implies

$$d\big(f_{n_k}(z), g(z)\big) < \epsilon \quad \text{for all } z \in K_r. \qquad \square$$

1.7 Metric Distances

It is comforting to be able to give explict formulas for d_σ for standard choices of the metric σ. Krantz dances around this, and it my mind, doesn't give convincing arguments. As we will see in detail below, often the key step is to establish the formula for $d_\alpha(0,r)$ with $r > 0$. This is done by proving that the straight line segment $[0,r]$ achieves the minimal distance. For the Poincare metric, Krantz outlines this in §1.1 of his book, but the argument is hardly tight. But we can easily tighten it up.

Suppose that $0 < r < 1$ and that $\gamma(t) = x(t) + iy(t)$ is a path from 0 to r. Then

$$\begin{aligned} L_\rho(\gamma) &= \int_0^1 \frac{\sqrt{x'(t)^2 + y(t)^2}}{1 - x(t)^2 - y'(t)^2}\,dt \\ &\geq \int_0^1 \frac{|x'(t)|}{1 - x(t)^2}\,dt \\ &\geq \int_0^1 \frac{x'(t)}{1 - x(t)^2}\,dt \\ &= \frac{1}{2}\ln\Big(\frac{1+r}{1-r}\Big). \end{aligned}$$

But this valued in attained by the straight line path from 0 to r. But it is not hard to see that $d_\rho(0,z) = d_\rho(0, e^{i\theta} z)$ for any θ. Hence

$$d_\rho(0,z) = d_\rho(0,|z|) = \frac{1}{2}\ln\Bigl(\frac{1+|z|}{1-|z|}\Bigr).$$

Krantz suggests the same approach should work for the spherical metric. I think not! (At least I see no direct estimates such as the above, nor do I find any heuristic arguments convincing in this case.) But eventually, I realized there was a general principal at work.

Proposition 1.34. *Let σ be any metric on a convex neighborhood D of the origin. Then for any $r \in \mathbf{R}$,*

$$d_\sigma(0,r) = \int_0^r \sigma(t)\,dt.$$

Remark 1.35 (Pay-off). In other words, for $d_\sigma(0,r)$, the distance is attained by $L_\sigma([0,r])$. Thus $\rho(z) = 1/(1-|z|^2)$ is the Poincare metric on U, then

$$d_\rho(0,r) = \frac{1}{2}\ln\Bigl(\frac{1+r}{1-r}\Bigr)$$

as derived above. On the other hand, if $\tau(z) = 2/(1+|z|^2)$, then

$$d_\tau(0,r) = 2\arctan(r). \tag{1.4}$$

This result also gives a proof (of the easily proved result) that in the Euclidean metric, that the minimal distance is given by line segments.

Proof. To prove the Proposition, it suffices to see that the derivative of the function $r \mapsto d_\sigma(0,r)$ is $\sigma(r)$.

Fix $\epsilon > 0$. Observe that if $h \in \mathbf{R}$, then

$$L_\sigma([r, r+h]) = \int_r^{r+h} \sigma(t)\,dt.$$

Since σ is continuous, there is a $\delta > 0$ such that $0 < |h| < \delta$ implies that $r + h \in D$ and that

$$-\frac{\epsilon}{2} < \frac{1}{h} L_\sigma([r, r+h]) - \sigma(r) < \frac{\epsilon}{2}.$$

Suppose $0 < h < \delta$. Then we can find a contour γ_h from 0 to r such that

$$L_\sigma(\gamma_h) - \frac{h\epsilon}{2} < d_\sigma(0,r).$$

Since $\gamma_h + [r, r+h]$ is a path from 0 to $r+h$ in D,

$$\begin{aligned}\frac{1}{h}\big(d_\sigma(0,r+h) - d_\sigma(0,r)\big) - \sigma(r)&\\ &< \frac{1}{h}\Big(L_\sigma(\gamma_h) + L_\sigma([r,r+h] - L(\gamma_\sigma) + \frac{h\epsilon}{2}\Big) - \sigma(r)\\ &= \frac{1}{h}L_\sigma([r,r+h]) - \sigma(r) + \frac{\epsilon}{2}\\ &< \epsilon.\end{aligned}$$

On the other hand, we can find a contour γ_h' from 0 to $r+h$ such that

$$L_\sigma(\gamma_h') - \frac{h\epsilon}{2} < d_\sigma(0,r+h).$$

Then we have

$$\begin{aligned}\frac{1}{h}\big(d_\sigma(0,r+h) - d_\sigma(0,r)\big) - \sigma(r)&\\ &> \frac{1}{h}\Big(L_\sigma(\gamma_h') - \frac{h\epsilon}{2} - (L_\sigma(\gamma_h') + L_\sigma([r+h,r]))\Big) - \sigma(r)\\ &= -\frac{1}{h}L([r+h,r]) - \sigma(r) - \frac{\epsilon}{2}\\ &> -\epsilon.\end{aligned}$$

To summarize, if $0 < h < \delta$, then

$$\Big|\frac{1}{h}\big(d_\sigma(0,r+h) - d_\sigma(0,r)\big) - \sigma(r)\Big| < \epsilon. \tag{1.5}$$

We argue similarly if $-\delta < h < 0$. We pick a contour γ_h such that

$$L_\sigma(\gamma_h) + \frac{h\epsilon}{2} < d_\sigma(0,r+h).$$

Then

$$\frac{1}{h}\big(d_\sigma(0,r+h) - d_\sigma(0,r)\big) - \sigma(r) = \frac{1}{-h}\big(d_\sigma(0,r) - d_\sigma(0,r+h)\big)$$

$$\begin{aligned} &< \frac{1}{-h}\Big(L_\sigma(\gamma_h) + L_\sigma([r+h, r]) - L_\sigma(\gamma_h) - \frac{h\epsilon}{2}\Big) - \sigma(r) \\ &= -\frac{1}{h} L_\sigma([r+h, r]) - \sigma(r) + \frac{\epsilon}{2} \\ &< \epsilon. \end{aligned}$$

A similar computation gives $-\epsilon < \frac{1}{h}\big(d_\sigma(0, r+h) - d_\sigma(0, r)\big) - \sigma(r)$.

Hence (1.5) holds for all $0 < |h| < \delta$. The result follows. □

Now recall that a *Möbius transformation* is a linear fractional transformation of the form

$$h_a(z) = \frac{z-a}{1-\bar{a}z} \quad \text{for some } a \in U.$$

Since $|z| = 1$ implies $|h_a(z)| = 1$, it follows from the maximum modulus principal that h_a maps to unit ball to itself. Since $h_a^{-1} = h_{-a}$, h_a is a bijection of U onto itself. Also,

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

If ρ is the Poincare metric, then

$$\begin{aligned} h_a^*\rho(z) &= \rho\big(h_a(z)\big)|h_a'(z)| \\ &= \frac{1}{1-\left|\frac{z-a}{1-\bar{a}z}\right|^2} \cdot \frac{1-|a|^2}{|1-\bar{a}z|^2} \\ &= \frac{1-|a|^2}{|1-\bar{a}z|^2 + |z-a|^2} \\ &= \rho(z). \end{aligned}$$

Hence h_a is d_ρ-distance reducing on U. But so is $h_a^{-1} = h_{-a}$. So each Möbius transformation h_a is Poincare distance preserving. This is the key to the following.

Corollary 1.36 ([Kra90, Proposition 2.4.2])**.** *If ρ is the Poincare metric on U then*

$$d_\rho(p, q) = \frac{1}{2} \ln \left(\frac{1 + \left|\frac{p-q}{1-\bar{p}q}\right|}{1 - \left|\frac{p-q}{1-\bar{p}q}\right|} \right).$$

Proof. As we have already observed, $d_\rho(0, z) = d_\rho(0, e^{i\eta} z)$, and hence Remark 1.35 on page 19 gives

$$d_\rho(0, z) = \frac{1}{2} \ln\Big(\frac{1 + |z|}{1 - |z|}\Big) \tag{1.6}$$

as claimed in (1.1).

But by the above discussion

$$d_\rho(p, q) = d_\rho\big(h_p(p), h_p(q)\big) = d_\rho\Big(0, \frac{q - p}{1 - \bar{p}q}\Big)$$

Now the assertion follows from (1.6). □

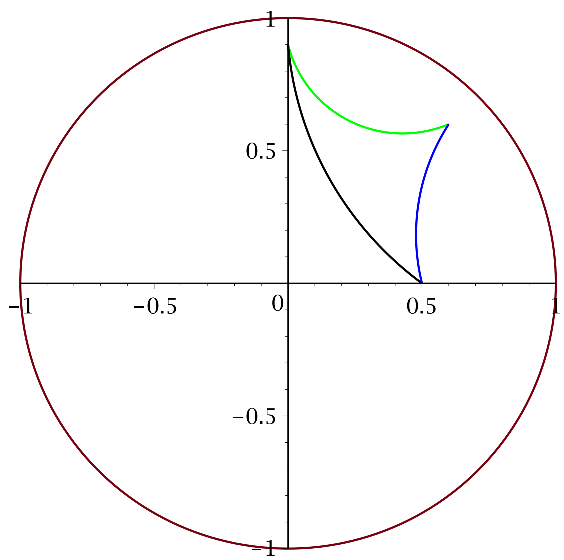


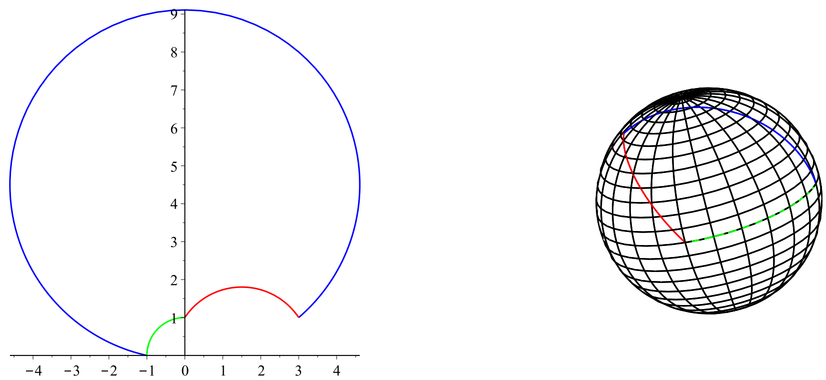
Figure 1: The Poincare Triangle $\Delta(0.5, 0.9i, 0.6 + 0.6i)$

Corollary 1.37. *If τ is the spherical metric on* $\mathbf{C}$ *then*

$$d_\tau(p, q) = 2 \arctan\Big(\frac{|p - q|}{|1 + \bar{p}q|}\Big).$$

Proof. Since $L_\tau(\gamma) = L_\tau(e^{i\theta} \cdot \gamma)$, d_τ is invariant under simple rotations. In particular,

$$d_\tau(0, p) = d_\tau(0, |p|) = 2 \arctan(|p|). \tag{1.7}$$



(a) In the Plane (b) On the Riemann Sphere

Figure 2: The Spherical Triangle $\Delta(-1, 3+i, i)$ in the Plane and on the Sphere

Although I am not prepared to show this here, every rotation of the Riemann sphere is given by a linear fractional transformation of the form

$$\varphi_{a,b}(z) = \frac{az+b}{-\bar{b}z+\bar{a}} \quad \text{where } |a|^2 + |b|^2 = 1.$$

A computation shows that[10]

$$\varphi_{a,b}^{*}\tau = \tau.$$

This gives $d_\tau\big(\varphi_{a,b}(p), \varphi_{a,b}(q)\big) \leq d_\tau(p,q)$. But $\varphi_{a,b}^{-1} = \varphi_{\bar{a},-b}$ is of the same type. Hence any such $\varphi_{a,b}$ is actually d_τ-distance preserving.

[10] First, check that

$$\varphi'_{a,b}(z) = \frac{1}{(-\bar{b}z+\bar{a})^2}.$$

Then

$$\begin{aligned}
\varphi_{a,b}^{*}\tau(z) &= \tau\big(\varphi_{a,b}(z)\big)|\varphi'_{a,b}(z)| \\
&= \frac{2}{1+\left|\frac{az+b}{-\bar{b}z+\bar{a}}\right|} \cdot \frac{1}{|-\bar{b}z+\bar{a}|^2} \\
&= \frac{2}{|-\bar{b}z+\bar{a}|^2 + |az+b|^2} \\
&= \frac{2}{1+|z|^2}.
\end{aligned}$$

Next notice that

$$\varphi(z) = \frac{z-p}{\bar{p}z+1}$$

is a rotation with $a = 1/\sqrt{1+|p|^2}$ and $b = -p/\sqrt{1+|p|^2}$.

Hence

$$d_\tau(p,q) = d_\tau\big(\varphi(p),\varphi(q)\big) = d_\tau\Big(0, \frac{q-p}{\bar{p}q+1}\Big) = d_\tau\Big(0, \Big|\frac{q-p}{\bar{p}q+1}\Big|\Big).$$

The result follows from (1.7). □

Remark 1.38. It is not obvious that there will be a minimal path from p to q in D for a given metric σ on D. As Krantz tells us, such paths always exist for many metrics and in particular, for the Poincare Metric. Given $p, q \in U$, let φ be a Möbius transformation taking p to 0. Then as we noted above, $d_\rho(p,q) = d_\rho(0,\varphi(q))$. Furthermore, the later value is attained by the line segment $[0,\varphi(q)]$ which is parameterized by $\gamma(t) = t\varphi(q)$ for $t \in [0,1]$. Hence the shortest path from p to q is given Γ_ρ parameterized by $t \mapsto \varphi^{-1}(t\varphi(q))$. That is,

$$\Gamma_\rho(t) = \frac{t\frac{q-p}{1-q\bar{p}} + p}{1 + t\bar{p}\frac{q-p}{1-q\bar{p}}}$$

I used this to make the "Poincare Triangle" in Figure 1 on page 22.

A similar analysis can be used to find minimal curves in the spherical metric. In this case the minimal distance from p to q is given by

$$\Gamma_\tau(t) = \frac{t\frac{q-p}{1+q\bar{p}} + p}{1 - t\bar{p}\frac{q-p}{1+q\bar{p}}}$$

This can used to draw the "Spherical Triangle" in Figure 2a on the previous page. If the shape is surprising, as we'll see in Remark 1.39, d_τ is the distance between the appropriate points on the Riemann sphere. (Thus the "size" of the interior of the unit circle and the exterior are the same!) I've drawn the same triangle on the Riemann Sphere in Figure 2b.

Remark 1.39 (Spherical Distance on the Riemann Sphere). Recall that the sterographic projection

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

is a map of $\mathbf{C}$ onto the sphere $S^2 := \{\, (x, y, z) \in \mathbf{R}^3 : x^2 + y^2 + z^2 = 1 \,\}$ minus the north pole $(1, 0, 0)$. I want to argue here that the spherical distance $d_\tau(p, q)$ is just the "great circle distance" from $S(p)$ to $S(q)$ on S^2. Since both distances are invariant under rotations of S^2, it suffices to check this for $d_\tau(0, r)$.[11]

But $S(0) = (0, 0, -1)$ and $S(r) = \big(\frac{2r}{1+r^2}, 0, \frac{r^2-1}{r^2+1}\big)$. However the great circle distance from $S(0)$ to $S(r)$ is just the central angle θ they determine. But using the dot product,

$$\cos(\theta) = \frac{1 - r^2}{1 + r^2}.$$

We want to compare this with (1.4), but we have to be careful. At the moment $\theta \in [0, \pi]$ and may not play nice with arctan. But

$$\cos\Big(\frac{\theta}{2}\Big) = \sqrt{\frac{1 + \cos\theta}{2}} = \sqrt{\frac{1 + \frac{1-r^2}{1+r^2}}{2}} = \sqrt{\frac{1}{1 + r^2}}.$$

But then

$$\tan\Big(\frac{\theta}{2}\Big) = \frac{\sqrt{1 - \frac{1}{1+r^2}}}{\sqrt{\frac{1}{1+r^2}}} = r$$

Now it follows that the distance, θ, is $2\arctan(r)$ as required.

[11] Ok, I never proved the maps $\varphi_{a,b}$ used above are rotations of S^2. Maybe later.

References

[Kna05] Anthony W. Knapp, *Basic real analysis*, Cornerstones, Birkhäuser Boston Inc., Boston, MA, 2005. Along with a companion volume *Advanced real analysis*.

[Kra90] Steven G. Krantz, *Complex analysis: the geometric viewpoint*, Carus Mathematical Monographs, vol. 23, Mathematical Association of America, Washington, DC, 1990.

[Rud87] Walter Rudin, *Real and complex analysis*, McGraw-Hill, New York, 1987.