Three Problems in Search of a Measure

Jonathan L. King

§P. PREFACE. What do the following three problems have in common?

Poncelet's Theorem. The lefthand figure below shows a pair of ellipses **C** and **E** which happen to have a *circuminscribed polygon*; a polygon which is simultaneously inscribed in the outer ellipse and circumscribed about the inner ellipse.

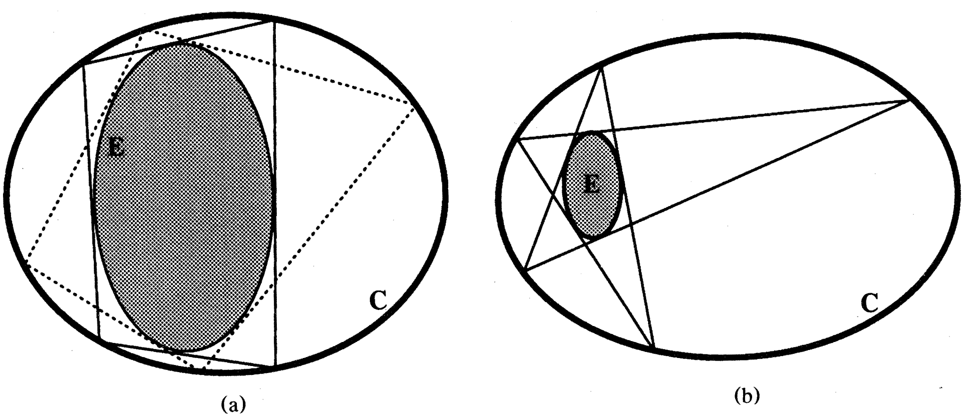

Figure P.1. On the left are two non-degenerate ellipses, with a quadrilateral (solid line) inscribed in **C** and circumscribed about **E**. The dotted-line shows another circuminscribed quadrilateral, which can be thought of as the first quadrilateral "rotated" about **E**. The right hand figure shows a pair of ellipses with a (self-intersecting) circuminscribed pentagon.

Jean-Victor Poncelet's famous CLOSURE THEOREM, published in his *Traité des propriétés projectives* of 1822, asserts that if there exists one circuminscribed n-gon then *any* point on the boundary of **C** is the vertex of some circuminscribed n-gon. Indeed, if we allow the n-gon to continuously change its shape, a circuminscribed n-gon can be continuously rotated around **E**.

How might this theorem be generalized to the case where the **CE** pair has no circuminscribed polygon?

Tarski's Plank Problem. Consider a circular table with diameter 9 feet. At your disposal are many planks, each 1 foot wide and longer than 9 feet. What is the minimum number of planks required to cover the surface of the table? Nine parallel planks certainly cover the tabletop. But can you take fewer and criss-cross them to cover the tabletop? If countably many planks of different widths are permitted, is there a cover using less total width than the parallel one?

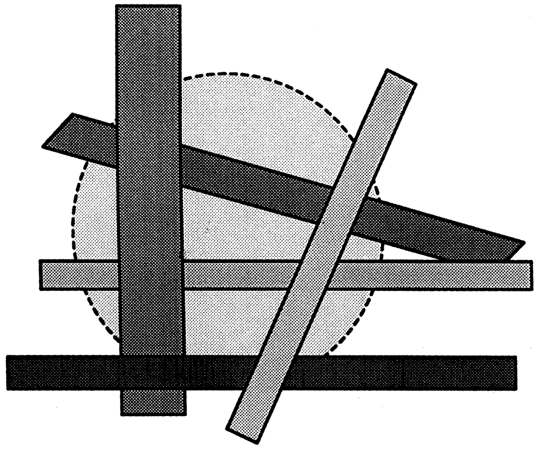

Figure P.2. Thin planks partially covering a round tabletop, in criss-cross fashion. The planks can be thought of as being infinitely long.

Gelfand's Question. In the table below, row n has the leftmost (high-order) digit of the numbers $2^n, 3^n, \ldots, 9^n$, when written in base ten. The "7" in the third row, column $\mathbf{9}^n$, is the 7 from $\underline{7}29 = 9^3$.

TABLE P.3 The leftmost digits of powers.

n:	$\mathbf{2}^n$	$\mathbf{3}^n$	$\mathbf{4}^n$	$\mathbf{5}^n$	$\mathbf{6}^n$	$\mathbf{7}^n$	$\mathbf{8}^n$	$\mathbf{9}^n$
1:	2	3	4	5	6	7	8	9
2:	4	9	1	2	3	4	6	8
3:	8	2	6	1	2	3	5	7
4:	1	8	2	6	1	2	4	6
5:	3	2	1	3	7	1	3	5
6:	6	7	4	1	4	1	2	5
7:	1	2	1	7	2	8	2	4
⋮				...				
99:	6	1	4	1	1	4	2	2
⋮				...				

Will a "9" ever occur in the **2**-column? Will the row "23456789" appear again? If so, will the set of n such that $\text{row}_n =$ "23456789" have a *frequency*? —and if it does, will it be rational or irrational? Will a row of all-the-same-digit occur? Will the decimal expansion for an 8-digit prime every appear?

Philosophy. At first glance the three problems seem to have little in common; Poncelet's theorem is a question about conic sections, the Plank problem about geometric set-inclusion, and Gelfand's question is number theoretic. Yet it turns out that two of these three are secretly isomorphic.

More significantly, they have a less precise but deeper commonality in that each of the three problems has, sitting somewhere inside it, a natural measure—a measure which, on some collection of sets, is preserved under some family of motions. This basic tool of an **invariant measure** is the theme of this article. It will turn out that the associated invariant measures make the hidden isomorphism conspicuous.

Anatomy. The next three sections construct in turn natural invariant measures for Poncelet's Theorem, the Plank Problem and Gelfand's Question, and can be read essentially independently. All three measures are finite measures which can be interpreted as lengths, areas or volumes. Consequently, the reader need only have an intuitive feel for the size of a set; no theorems of formal measure theory† are used. The hope is that the article can be read by motivated undergraduates who are comfortable with integration and elementary topology.

The APPENDIX contains a brief history of each problem or a pointer to such. It alludes to the connection with ergodic theory, and ends with an open problem.

Idiosyncrasy. Use "$a := b$" to mean "a is defined to be b". We use the usual symbol, "B^c", for the complement of a set B; let "$A \overset{c}{-} B$" denote the set-difference $A \cap B^c$. Symbol $\sqcup_1^\infty B_k$ indicates that the sets $\{B_k\}_k$ in the union happen to be disjoint.

If λ is a measure such as "length" on a space Y, then a "λ-nullset" $B \subset Y$ has zero length, $\lambda(B) = 0$. An example is a set B consisting of finitely many points. The pair (Y, λ) is called a measure space. A map φ: $X \to Y$ between two measure spaces (X, μ) and (Y, λ) is *measure-preserving* if

$$\mu(\varphi^{-1}B) = \lambda(B), \text{for all subsets } B \text{ of } Y.$$

A measure-preserving map R: $X \to X$ from a space to itself is called a *transformation*, which we write in full as (R: X, μ). Finally, for a point $z \in X$ let $R^n(z)$ denote the n-fold composition $R(R(\overset{n}{\ldots}R(z)\ldots))$. The sequence of points $z, R(z), R^2(z), \ldots$ is the *orbit* of z under R.

§1 PONCELET'S THEOREM. From any point outside of **E**, a "righthand tangent to **E**" can be consistently chosen. In particular, **E** gives rise to a "righthand homeomorphism," R, mapping **C** to itself.

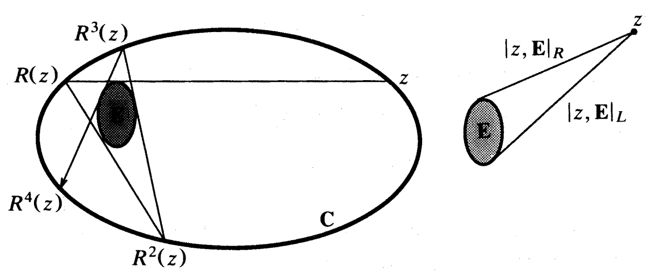

Figure 1.1. Two ellipses, with **E** properly inside of **C**. An observer standing at z peering inward at **E** sees a "righthand" and a "lefthand" tangent to **E**. Call the righthand tangent map R: $C \to C$. NOTATION: From a point z outside **E**, let $|z, \mathbf{E}|_L$ and $|z, \mathbf{E}|_R$ denote the distance from z to the lefthand and righthand tangent points on **E**. In the case **E** is a circle, ie. $|z, \mathbf{E}|_L = |z, \mathbf{E}|_R$, agree to write $|z, \mathbf{E}|$ for the common value.

†Originally, this article contained a fourth problem, invented by David Feldman, concerning billiards in the cusp between the curves $y = 1/x$ and $y = -1/x$. Because its solution does use elementary measure theory —and on an infinite measure space— as well as the idea of recurrence in a dynamical system, the discussion of that problem will appear separately, in [KING].

Viewed this way, Poncelet's theorem becomes a statement about the dynamics of the map R and explains why it is called a CLOSURE THEOREM; if the orbit of one point z "closes up" in n steps, that $R^n(z) = z$, then the orbit of any point closes up and also in n steps.

In closing up, if the orbit of z "winds around **E**" p times ($p = 1$ and $n = 4$ in Figure P.1(a); in P.1(b), $p = 2$ and $n = 5$) then every point's orbit would have to wind p times before closing up, since R preserves order along **C**. This suggests that —after a suitable change of coordinates—R is simply a rigid rotation of a circle by rational rotation number p/n. Let $\mathbb{K} := [0, 1)$ be the half-open interval topologized as a circle, and let $\oplus$ and $\ominus$ denote addition and subtraction modulo 1. For a *rotation number* $\alpha \in \mathbb{R}$, let

$$\rho_\alpha : \mathbb{K} \to \mathbb{K} : x \mapsto x \oplus \alpha$$

be the corresponding rigid rotation. A "change of coordinates" would then be a homeomorphism φ: $\mathbf{C} \to \mathbb{K}$ with a commutative diagram

$$\begin{array}{ccc} \mathbf{C} & \xrightarrow{R} & \mathbf{C} \\ \downarrow \varphi & & \downarrow \varphi \\ \mathbb{K} & \xrightarrow{\rho_\alpha} & \mathbb{K} \end{array} \tag{1.2}$$

Such a φ satisfying $\varphi \circ R = \rho_\alpha \circ \varphi$ is called a *topological conjugacy* from R to ρ_α. Poncelet's theorem would thus follow from

Lemma 1.3. *For ellipses* **C**, **E** *and mapping R as in Figure* 1.1, *there is a rotation number $\alpha \in [0, 1)$ and topological conjugacy φ carrying R to ρ_α.*

Indeed, this lemma asserts something even if the R-orbit of z *fails* to close up—the case when α is irrational.

The appearance of a measure. Arclength measure λ is *invariant* under the rotation ρ_α, that is,

$$\lambda\big(\rho_\alpha^{-1}(B)\big) = \lambda(B)$$

for any set B included in $\mathbb{K}$. We normalize λ to a probability measure, $\lambda(\mathbb{K}) = 1$. A conjugacy φ would lift λ to an R-invariant measure μ on **C**,

$$\mu(A) := \lambda(\varphi(A)), \tag{1.4}$$

which we will call *good*: finite, non-atomic (any individual point has zero μ-length) and giving positive length to open intervals.

It is the converse which will help us out:

Any R-invariant good μ gives rise to a topological conjugacy.

In order to define this conjugacy, for points $z, y \in \mathbf{C}$ let $[z, y)$ denote the half-open interval on **C** going counterclockwise from z to y. Next, normalize μ so that $\mu(\mathbf{C}) = \lambda(\mathbb{K}) = 1$; now, for any three points z, y, x on **C**

$$\mu([z, y)) \oplus \mu([y, x)) = \mu([z, x)).$$

Fix any particular point $z_0 \in \mathbf{C}$. Define φ and α by

$$\varphi(y) := \mu([z_0, y)) \quad \text{and } \alpha := \mu([z_0, Rz_0)). \tag{1.5}$$

Thus φ sends z_0 to 0 in $\mathbb{K}$ and is well-defined because of the normalization. Non-atomicity implies continuity of φ and the "positive length" condition insures that φ is invertible. Invariance yields that for any x,

$$\begin{aligned} \mu([x, Rx)) &= \mu([Rz_0, Rx)) \ominus \mu([Rz_0, x)) \\ &= \mu([z_0, x)) \ominus \mu([Rz_0, x)) \\ &= \mu([z_0, Rz_0)) = \alpha. \end{aligned}$$

In other words, the rotation number α did not truly depend on the arbitrary point z_0, but only on the R-invariant measure μ. With this, it is easy to verify that φ carries R to ρ_α, as in (1.2).

Our previous lemma can be restated now like this.

Lemma 1.3′. *For ellipses* **C** *and* **E**, *with R as in* Figure 1.1: *There exists an R-invariant good measure μ.*

When **C** and **E** are concentric circles, arclength measure along **C** is R-invariant. However, the case where they are non-concentric circles seems not as apparent, and the general elliptical case is less evident still. Happily, we can at least assume that the outer ellipse **C** is a circle—since any linear map of the plane carries ellipses to ellipses and tangent chords to tangent chords and so carries Figure 1.1 to another just like it. So before even starting to construct μ, one can linearly compress the figure along the major axis of **C** to arrange that now **C** is a circle.

Rolling the chord. The righthand map, R, generally stretches or shrinks sub-intervals $I \subset \mathbf{C}$, thus making the standard arclength non-invariant. A direct approach to making an R-invariant "length" μ, is to compensate for stretch/shrink by integrating against arclength an appropriately chosen "height function" h whose height varies so as to cancel out the distortion introduced by R. Then the "μ-length" of a set A would be

$$\mu(A) = \mu_h(A) := \int_A h(z)\, dz \qquad \text{where “}dz\text{” denotes arclength measure on } \mathbf{C}. \tag{1.6}$$

If h is a continuous function from **C** to the positive reals, then automatically μ will be non-atomic and give positive length to open intervals. The issue becomes: *What property does $h(\cdot)$ need to fulfill for the measure $\mu = \mu_h$ to be R-invariant near a point z?*

In Figure 1.7, the ratio $((z\text{-arc})/\Delta z) \to 1$ as the angle $\theta \searrow 0$, and similarly $((y\text{-arc})/\Delta y) \to 1$. Consequently, the "infinitesimal ratio" of the y-arc to the z-arc is

$$\lim_{\theta \searrow 0} \frac{y\text{-arc}}{z\text{-arc}} = \lim_{\theta \searrow 0} \frac{\Delta y}{\Delta z} = \frac{\ell_y}{\ell_z};$$

this last equality comes from the equality $\measuredangle Pyz = \measuredangle Pzy$ of the base angles of isosceles triangle yPz at the right of Figure 1.7.

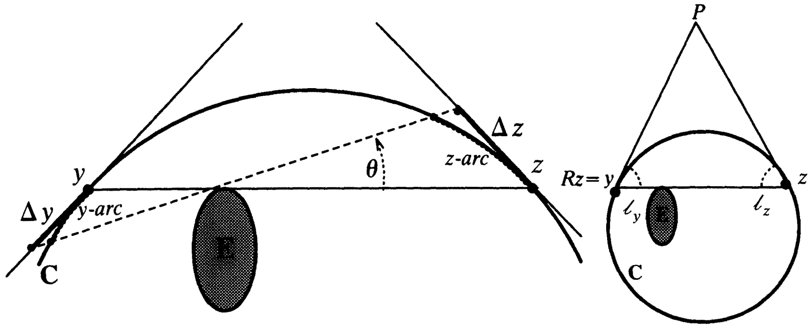

Figure 1.7. Circle **C** surrounds ellipse **E**. Let z and y be points at opposite ends of a chord tangent to **E**, and let ℓ_z and ℓ_y denote their distance to the common point of tangency. Roll the chord through a small angle θ. Its "z end" sweeps out an arc on **C** as does its "y end"; on the corresponding tangent lines, distances Δz and Δy are swept out.

Infinitesimally, the μ-length of the "z-arc" is simply its arclength times $h(z)$. So for μ to be R-invariant at z, the function h must satisfy that —infinitesimally— the product $h(z)$ Length(z-arc) equals $h(y)$ Length(y-arc). In consequence, the invariance condition required of $h(\cdot)$ is

$$h(z) \cdot |z, \mathbf{E}|_R = h(y) \cdot |y, \mathbf{E}|_L, \text{ where } y = Rz. \tag{1.8a}$$

This uses the previous displayed-equation as well as $\ell_z = |z, \mathbf{E}|_R$ and $\ell_y = |y, \mathbf{E}|_L$.

Proof of Lemma 1.3′: If our inside ellipse **E** happens to be a circle, then the lefthand and righthand tangential-distances equal a common function $z \mapsto |z, \mathbf{E}|$. Then

$$h(z) := 1/|z, \mathbf{E}|$$

satisfies (1.8a), and the corresponding μ is the desired R-invariant measure.

To handle the non-circular case, choose some linear map $\mathscr{A}$ which transforms **E** into a circle, as shown in Figure 1.9. Since a linear map preserves the ratio of lengths of parallel line-segments, we have that

$$\frac{|y, \mathbf{E}|_L}{|z, \mathbf{E}|_R} = \frac{|\mathscr{A}y, \mathscr{A}\mathbf{E}|_L}{|\mathscr{A}z, \mathscr{A}\mathbf{E}|_R}.$$

But since $\mathscr{A}\mathbf{E}$ is a circle, the subscripts on the second ratio are unnecessary and it can be written $|\mathscr{A}y, \mathscr{A}\mathbf{E}|/|\mathscr{A}z, \mathscr{A}\mathbf{E}|$. Just as before, then,

$$h(z) := \frac{1}{|\mathscr{A}z, \mathscr{A}\mathbf{E}|}, \quad \text{for all } z \in \mathbf{C}, \tag{1.8b}$$

satisfies (1.8a) and makes the measure μ_h —which we will call *Poncelet-measure*— invariant under R.

Exercise: *Poncelet-measure for a special case.* Suppose **C** and **E** are confocal ellipses, with foci at $(0, \pm F)$ and with semi-minor axis lengths of 1 and r

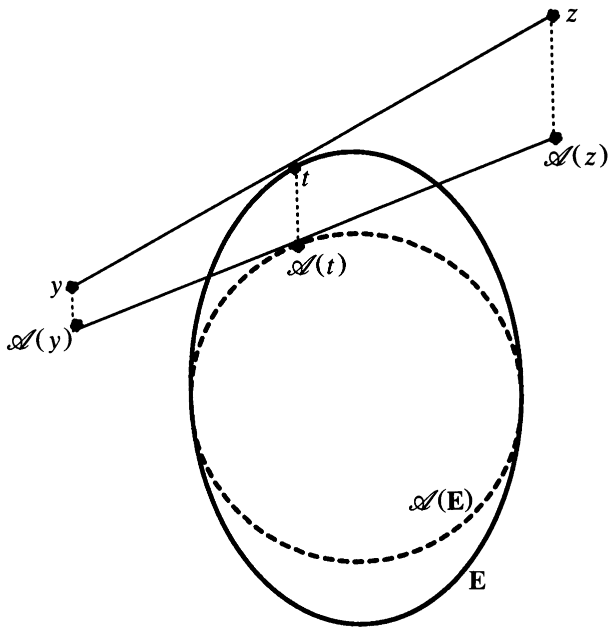

Figure 1.9. Ellipse **E** is carried by linear map $\mathscr{A}$ to a circle $\mathscr{A}(\mathbf{E})$. For artistic convenience, the linear map illustrated simply compresses the major axis (vertical, in the diagram) of **E** until equality with **E**'s minor axis. Each point of chord *ytz* is carried vertically downward to line-segment $\mathscr{A}(y)\mathscr{A}(t)\mathscr{A}(z)$; here t is the point where the yz-chord is tangent to **E**.

respectively, ie.

$$\mathbf{C}\colon x^2 + \frac{y^2}{F^2 + 1^2} = 1^2 \quad \text{and} \quad \mathbf{E}\colon \frac{x^2}{r^2} + \frac{y^2}{F^2 + r^2} = 1^2,$$

with $0 < r < 1$. Then Poncelet-measure μ_h will be the integral, against arclength measure along ellipse **C**, of the function

$$h(x, y) = \text{const} \cdot \frac{1}{\sqrt{(1 + x^2F^2)(r^2 + x^2F^2)}}, \tag{1.10}$$

where the constant is the square root of $(F^2 + r^2)/(1 - r^2)$. □

Remark. What can be said when **CE** has no circuminscribed polygon? —that is, when R is conjugate to an irrational rotation. One starts drawing tangent-chords from some point z, as in Figure 1.1, but the polygon-in-progress never closes up. What will be true is that the vertices of this unfulfilled polygon will densely fill out **C**:

$$\textit{Under an irrational rotation } \rho = \rho_\alpha, \textit{ the orbit of } 0 \textit{ is dense.} \tag{1.11}$$

To establish this, partition the circle $\mathbb{K}$ into M subintervals

$$\left[0, \frac{1}{M}\right), \ldots, \left[\frac{M-1}{M}, 1\right).$$

Take $N \geqslant 1$ smallest such that $\rho^N(0)$ falls inside $[0, (1/M))$. (This certainly will happen for some $N \leqslant M$ since, by the Pigeonhole Principle, some two of the $M + 1$ points $\{0, \rho(0), \rho^2(0), \ldots, \rho^M(0)\}$ fall into the same subinterval.) Thus the ρ^N-orbit of 0 is $(\frac{1}{M})$-dense. So its ρ-orbit is ε-dense, for all ε. □

By the way, it is natural to wonder whether the Poncelet rotation-number α is essentially unique; is it independent of linear map $\mathscr{A}$? That it is, follows from this challenge:

If rotations ρ_α and ρ_β are topologically conjugate, where $0 \leqslant \alpha, \beta \leqslant \frac{1}{2}$, then $\alpha = \beta$.

Another natural question, which is discussed in the appendix, is to wonder

$$\text{Is Poncelet-measure unique? How does } \mu_h \text{ depend on our choice of linear map } \mathscr{A}? \tag{1.12}$$

§2 **THE PLANK PROBLEM.** Experimentation with strips of paper as "planks" tends to suggest that a disk **D** cannot be efficiently covered unless the strips are parallel, which leads one to this

Plank Conjecture. *Suppose $(w_n)_{n=1}^{\infty}$ are the widths of a countable family of planks which cover disk* **D**. *Then*

$$\sum_{n=1}^{\infty} w_n \geqslant \text{Width}(\mathbf{D}).$$

If $\sum_{n=1}^{\infty} w_n$ actually equals Width(**D**), *the diameter of the disk, then the planks must be parallel to one another.*

Measuring area differently. We can take **D** to be the closed radius-1 disk centered at the origin and ask that a *plank* P be the closed region between two parallel lines, *both* of which contact **D**. The *width* of P is the perpendicular distance between these parallel edges.

In order to cover the disk by small total width, one is tempted to use planks which pass near the center of **D**, since such planks P cover more than their fair share of the area of the disk; more than Width(P)/Width(**D**). Yet if many planks pass near the center, they must waste some of their area by overlapping one another. Conversely, while it is easy to arrange that planks passing near the boundary of **D** be disjoint, such planks cover less than their fair share.

An answer to the Plank Problem is not obvious because the area covered by a plank P is not invariant under moving P over the disk. An analogy with Poncelet's theorem is useful. The only "obvious" case where a circuminscribed polygon can be rotated about **E**, is when the two ellipses are concentric circles. And this is exactly the case where normal arclength on **C** is evidently invariant under the transformation R. What saved the day, in the case of general ellipses, was the discovery of a different way of measuring length on **C**, the measure μ_h of (1.6), which indeed is invariant under motion by R.

Returning to our planks, the analogy suggests looking for a new way to measure area on **D** so that this "new area" —at least for planks— is invariant under rigid motions. We might expect to construct this measure, ν, as we did with Poncelet-measure, by integration of some nice positive height function $h\colon \mathbf{D} \to \mathbb{R}_+$ against area. If we build ν in this way, then any subset of the disk with zero ν-area would have to have zero area, as well. Also, the desired invariance of ν under rigid motions would mean there is a positive constant $\hbar$ so that

$$\nu(P) = \hbar \cdot \text{Width}(P), \quad \textit{for any plank } P. \tag{2.1}$$

One clarification is in order. Since ν is to be a measure on **D**, we need to interpret a plank P as subset of the disk and agree to regard "P" as the set $P \cap \mathbf{D}$. With this convention, **D** itself is a plank.

Assuming now that someone with outstanding geometric intuition has provided us with such a ν, let's use it to verify the conjecture.

Provisional Proof of the Plank Conjecture. If planks $\{P_n\}_1^\infty$ cover $\mathbf{D}$ then

$$\begin{aligned}\mathrm{Width}(\mathbf{D}) = \frac{1}{\hbar}\nu(\mathbf{D}) &= \frac{1}{\hbar}\nu\Big(\bigcup_n P_n\Big)\\ &\leqslant \frac{1}{\hbar}\sum_n \nu(P_n) = \sum_n \mathrm{Width}(P_n).\end{aligned} \tag{2.2}$$

This shows that no cover can use less total width than a parallel cover.

To show that *only* parallel covers use minimal width, suppose now we have a cover realizing equality in (2.2), that is, $\Sigma_n \nu(P_n) = \nu(\bigcup_n P_n)$. Consequently, the intersection $P_i \cap P_j$ of any pair of planks must have no ν-area and thus no area. This means that in a minimal cover any two planks have disjoint interiors; no two planks *overlap*. So we need but prove that

Any two planks P and R of a non-overlapping cover, are parallel.

To see this, consider any plank Q of the cover which intersects segment L_1 in the figure below.

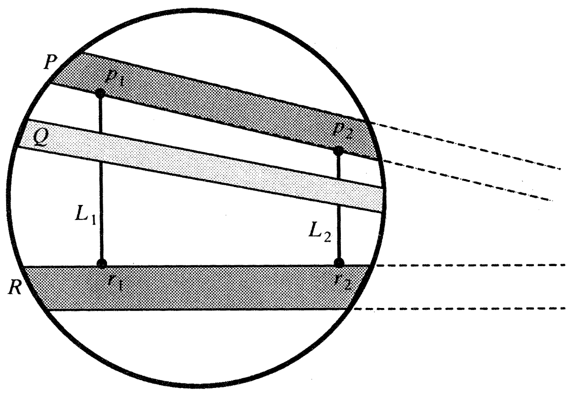

Figure 2.3. Given two planks P and R, choose distinct parallel line-segments $L_1 = \overline{p_1 r_1}$ and $L_2 = \overline{p_2 r_2}$; the L_i are "open" segments, not containing their endpoints. Endpoints p_i and r_i are on neighboring edges of P and R, respectively. Since L_1 and L_2 are disjoint from P and R, they must be covered by the other planks.

We can have chosen the point p_1 to be in the interior of $\mathbf{D}$; so Q cannot have L_1 as an edge. Therefore, since Q overlaps neither P nor R it must run *between* their neighboring edges; hence Q crosses both L_1 and L_2. Since the L_i are parallel, intervals $Q \cap L_1$ and $Q \cap L_2$ have the same length. Consequently,

$$\sum_{Q \in \mathscr{Q}_1} \mathrm{Length}(Q \cap L_1) = \sum_{Q \in \mathscr{Q}_1} \mathrm{Length}(Q \cap L_2),$$

where $\mathscr{Q}_1$ consists of all planks in the cover which overlap L_1. But these planks do not overlap each other. Consequently the closed intervals $\{Q \cap L_i\}_{Q \in \mathscr{Q}_1}$ are non-overlapping, and we may conclude that $\mathrm{Length}(L_1) \leqslant \mathrm{Length}(L_2)$. Reversing the roles of L_1 and L_2 shows that they have equal length. And this means that P and R were parallel all along. □

Both halves of the proof, the inequality and the parallelness, hinge upon our not-yet-known-to-exist plank-measure ν. The rotational symmetry of condition (2.1) suggests looking at rotationally symmetric geometric figures as a possible source of plank-measure. It turns out that the geometric insight which will provide plank-measure was made by Archimedes in his treatise *On the Sphere and the Cylinder*.

Archimedes' area-preserving map of the Earth. The figure below shows how to project the globe onto a cylinder of paper in order to get a map of the Earth on which countries of equal spherical-area are represented by map-regions of equal (planar) area.

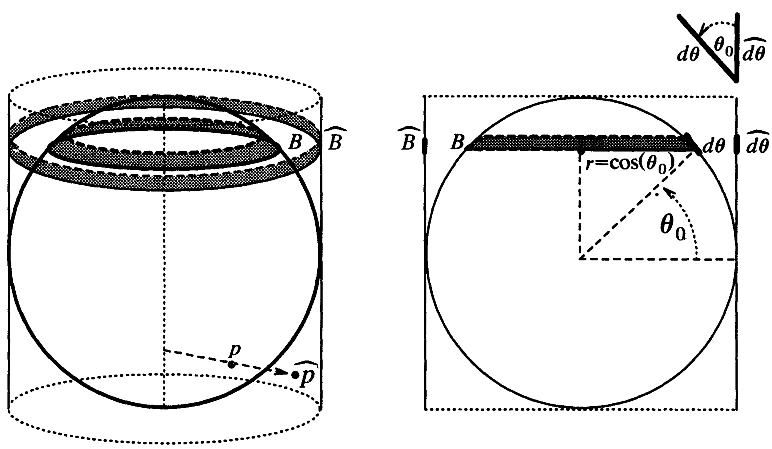

Figure 2.4. A sphere **S** and a (finite) cylinder **C** are tangent along their common equator. The height of **C** equals the sphere's diameter. For a point $p \in \mathbf{S}$, let $\hat{p} \in \mathbf{C}$ be its radial-projection: A ray starting from and perpendicular to the axis-of-symmetry of **C** passes through p and intersects the cylinder at $\hat{p}$.

Archimedes proved that for any region B on the sphere,

$$\text{Area}_{\mathbf{C}}(\widehat{B}) = \text{Area}_{\mathbf{S}}(B).$$

Since both the sphere and the cylinder are rotationally symmetric, it suffices to establish this when B is a band. Figure 2.4 shows, drawn on a sphere of radius 1, a band B with lower edge at latitude θ_0, and upper edge infinitesimally higher at latitude $\theta_0 + d\theta$. Since the radius of the bottom edge of B is $r = \cos(\theta_0)$,

$$\text{Infinitesimal Area}_{\mathbf{S}}(B) = (2\pi r) \cdot d\theta = 2\pi \cos(\theta_0)\, d\theta.$$

The projected band $\widehat{B}$ has radius 1 and infinitesimal width $\widehat{d\theta} = d\theta \cdot \cos(\theta_0)$, by similar triangles. Consequently

$$\text{Infinitesimal Area}_{\mathbf{C}}(\widehat{B}) = (2\pi \cdot 1) \cdot \widehat{d\theta} = 2\pi\, d\theta \cos(\theta_0).$$

The two infinitesimal areas are thus seen to be equal. □

Constructing plank measure. Area-measure of a sphere whose equatorial-plane is **D**, when projected down to **D**, has the desired plank property. Define ν by

$$\nu(A) := \text{Area}_{\mathbf{S}}(A_{\mathbf{S}}).$$

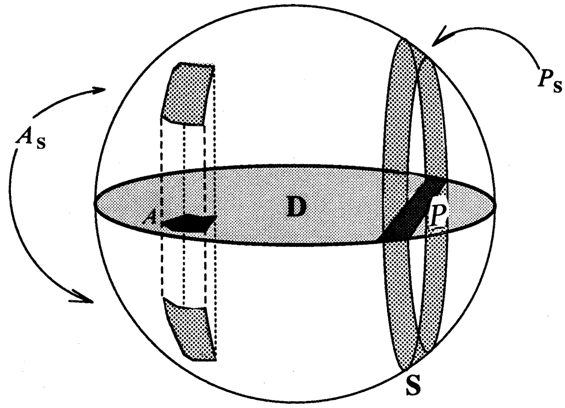

Figure 2.5. A sphere **S** whose equatorial-plane is our disk **D**. For any subset $A \subset \mathbf{D}$, let A_S be its orthogonal projection on the sphere. Thus the projection of a subdisk concentric with **D** would be a pair of "polar icecaps".

The image of a plank $P \subset \mathbf{D}$ is a band P_S on the sphere. The sphere can then be oriented so that the band projects on the cylinder to a cylindrical-band $\widehat{P_S}$ whose width necessarily equals the width of the original plank P. Thus

$$\nu(P) = \text{Area}_S(P_S) = \text{Area}_C(\widehat{P_S}) = 2\pi \cdot \text{Width}(P),$$

which demonstrates that every plank's measure is a *constant* times its width. Moreover, this establishes that the natural value of "Plank's Constant", $\hbar$, is 2π.

§3 GELFAND'S QUESTION. Questions about high-order digits of powers, Table P.3, provide a perfect setting for students to formulate and test numerical conjectures. Here are some "mysterious empirical facts" which can be found by judicious use of a computer. For a positive x, let $\langle\langle x \rangle\rangle$ denote the high-order non-zero digit of the base-ten expansion of x. Thus $\langle\langle$Plank's Constant$\rangle\rangle = 6$ and $\left\langle\left\langle \frac{e}{1000} \right\rangle\right\rangle$ equals 2.

Mysterious observations? Studying the **2**-column of Table P.3,

$$2\,4\,8\,1\,3\,6\,1\,2\,5\,1\,2\,4\,8\,1\,3\,6\,1\,2\,5\ \ldots\langle\langle 2^n \rangle\rangle\ldots$$

one discovers that the nine digits *do* appear to have frequencies‡ —and that they are unequal. Indeed, letting fr(d) denote the *frequency* of digit d,

$$\text{fr}(d) := \lim_{N \to \infty} \frac{1}{N} |\{n \mid 1 \leqslant n \leqslant N \text{ and } \langle\langle 2^n \rangle\rangle = d\}|,$$

the computer suggests that the frequencies decease steadily as d increases from 1 to 9:

$$\text{fr}(1) \approx .301, \quad \text{fr}(2) \approx .176, \quad \cdots \quad \text{fr}(9) \approx .045.$$

The first empirical surprise is that digit d occurs with the same frequency fr(d) in *every* column of the table. This, despite the fact that the columns are decidedly

‡The existence of frequencies is not a foregone conclusion. In the sequence of high-order digits $(\langle\langle n \rangle\rangle)_{n=1}^{\infty}$ of *all* natural numbers, the digit "1" does not have a frequency, since its upper-density is five times its lower-density.

different—they all have patterns which appear in no other column. For example, the sequence "248136" occurs infinitely often in column **2**, yet will never be witnessed by the other columns. Indeed, *no* sequence of length six which appears in a column, *ever* appears in any other column.

Every digit occurs in every row of Table P.3, and yet an all-the-same row never appears. This suggests looking at *joint frequencies*. Let $F_{2:6:9}(d, d', d'')$ be the frequency of rows n which have d in column **2**, digit d' in column **6** and d'' in column **9**. Some pairs of columns seem to be independent —columns **2 : 3** for instance— in that there is approximate equality

$$F_{2:3}(d, d') \approx \text{fr}(d)\text{fr}(d').$$

Yet columns **4** and **5** appear strongly non-independent. Only one digit d ever occurs simultaneously in columns **2 : 5**. Columns **2 : 4** don't have independence, yet columns **3 : 6** do. But **3 : 9** do not. And so on...

The real startler comes, however, when you vary the "seed." Row_{n+1} is obtained by multiplying row_n by the tuple $(2, 3, \ldots, 9)$. Table P.3 starts with a "zero-th row" of $(1, 1, \ldots, 1)$. But there is no reason to use all the same integer—and upon consideration, why even restrict to integers? We could start with a "zero-th row" $(s_2, \ldots, s_9)$ of positive real numbers as *seeds* and then study frequencies; let the symbol $F^{s, s'}_{\mathbf{m}:\mathbf{m}'}(d, d')$ mean the frequency of those n for which simultaneously $\langle\langle s\mathbf{m}^n \rangle\rangle = d$ and $\langle\langle s'\mathbf{m}'^n \rangle\rangle = d'$.

At first this seems like a waste of time, since the seed appears to have no effect on individual columns; no matter which of **2**, ..., **9** one takes for the multiplier **m**, apparently $F^s_{\mathbf{m}}(d) = \text{fr}(d)$ independently of the seed s. But this independence abruptly disappears for joint frequencies. An energetic computer will discover, for example, the oddity that the mapping

$$s \mapsto \mathbf{F}^{s, s}_{2:5}(3, 3) \tag{3.1}$$

is continuous—but fails to be differentiable at exactly three points in the interval $(1, 10)$.

Converting to a dynamical system. The mathematics is simpler if we first just analyze the statistics of a single column—the **2**-column perhaps or, more generally, the **m**-column for some fixed positive multiplier **m**. Effectively we are studying the "multiply map"

$$T : \mathbb{R}_+ \to \mathbb{R}_+ : x \mapsto \mathbf{m}x$$

applied to an initial seed s; we are following the T-orbit of s, taking the measurement $\langle\langle T^n(s) \rangle\rangle$ along it. Since x and $10x$ have the same first digit, we might as well identify numbers whose ratio is a power of 10. Doing this collapses all these half-open intervals,

$$\ldots \left[\tfrac{1}{10}, 1\right), [1, 10), [10, 100), \ldots$$

together, thus effectively wrapping the positive-reals $\mathbb{R}_+$ into a circle. Insofar as our measurements are concerned, then, the multiply map is just some homeomorphism of a circle.

The easiest way to identify $10x$ with x is simply to take logarithms base-ten and then discard the integer part. Letting ψ denote this identification and letting

$\mathbb{K} = [0,1)$ be the circle as before, we get this commutative diagram:

$$\begin{array}{ccc} \mathbb{R}_+ & \xrightarrow{T} & \mathbb{R}_+ \\ \downarrow \psi & & \downarrow \psi \\ \mathbb{K} & \xrightarrow{\rho_\alpha} & \mathbb{K} \end{array} \quad \text{where } \psi(x) := (\log_{10} x)_{\text{mod}\,1} \tag{3.2}$$

Since the logarithm converts multiplication to addition, transformation $\rho_\alpha := \psi \circ T \circ \psi^{-1}$ is simply rotation $x \mapsto x \oplus \alpha$ on the circle by $\alpha := \log_{10}(\mathbf{m})$, where "$\oplus$" means addition modulo 1. Thus the high-order digit of a number x is d if and only if $\psi(x)$ is in the half-open interval

$$I_d := [\log(d), \log(d+1)).$$

(Here, and henceforth, "log" means $\log_{10}$.) These nine intervals $\{I_d\}_{d=1}^{9}$ partition the circle.

An isomorphism. The underlying transformation for Gelfand's Question in the case of a single column, we've discovered, is a rigid rotation of a circle—this is the same transformation found hidden in Poncelet's theorem. Can this isomorphism be made more explicit? Yes it can.

The ellipses **C** and **E** in Figure 1.1 gave rise to a rotation number α and commutative diagram 1.2. Use the φ of that diagram to lift the intervals $\{I_d\}_1^9$ to a partition $\{J_d\}_1^9$ on **C**, where $J_d := \varphi^{-1}(I_d)$, and let $\mathbf{m} := 10^\alpha$.

Effectively $\varphi^{-1} \circ \psi$ is an isomorphism from the Gelfand **m**-system to the Poncelet **CE**-system: Any question about the **m**-column with seed s is mapped to the corresponding question about the Poncelet-orbit of $z := \varphi^{-1}(\psi(s))$. This R-orbit of z lands in intervals

$$J_{d_0}, J_{d_1}, J_{d_2}, \ldots, J_{d_n}, \ldots$$

precisely so that the high-order digit of $s\mathbf{m}^n$ is invariably d_n.

Equidistribution. Now that we know that the **2**-column is isomorphic to the orbit of 0 on the circle under rotation by log(2), what do we know? Since log(2) is irrational this orbit is dense, courtesy remark 1.11, and so every digit appears infinitely often in column **2**. We now find ourselves in the embarrassing circumstance of knowing there are integers n where 2^n starts with "9"—without having the foggiest notion of a single one.

However, *denseness* —a property which ignores the *time* when points appear in the orbit— is not sufficient to determine the *frequency* of "9", or whether it has a frequency at all. A more refined "randomness" property is needed: A sequence $z_0, z_1, \ldots$ in the circle is *equidistributed* if

$$\textit{For any interval } I \subset \mathbb{K}: \quad \lim_{N\to\infty} \frac{1}{N} \left|\{n | 0 \le n < N \,\&\, z_n \in I\}\right| = \text{Length}(I). \tag{3.3}$$

Weyl's Equidistribution theorem for the circle says that for any irrational α:

For any point z, the sequence $z, z \oplus \alpha, z \oplus 2\alpha, \ldots, z \oplus n\alpha, \ldots$ is equidistributed in $\mathbb{K}$.

For a multiplier $\mathbf{m} \in \{2, \ldots, 9, \pi, e, \text{etc.}\}$ which is not a rational power of ten, the corresponding rotation number $\alpha = \log(\mathbf{m})$ is irrational. Weyl's theorem explains the mysterious frequencies fr(d):

$$\text{fr}(d) = \text{Length}(I_d) = \log\left(\frac{d+1}{d}\right), \quad \text{for } d = 1, 2, \ldots, 9.$$

So the frequencies are irrational—in fact, transcendental. Weyl's theorem also explains why these frequencies do not depend on the seed.

Randomness and Ergodicity. An irrational rotation ρ is "topologically random" in the sense that each orbit is dense. Restated, the orbit of any point $a \in \mathbb{K}$ hits any non-empty open set U: *There exists n so that* $a \in \rho^{-n}(U)$.

An analogue of this, a notion of measure-theoretic "randomness", is that the orbit of any set A of positive length hits any other positive-length set U: *There exists n so that* $\lambda(A \cap \rho^{-n}(U)) > 0$. This measure-theoretic property makes sense for any measure-preserving transformation $(T : X, \mu)$, and has this equivalent formulation:

Any T-invariant† *set B is —up to a nullset— either empty or is the whole space. Either B or* $X \stackrel{c}{-} B$ *is a nullset.*

The connection between our irrational rotation ρ and ergodicity is twofold. On the one hand, an irrational rotation is ergodic. On the other hand, any ergodic transformation satisfies an abstract equidistribution theorem, due to Birkhoff, called the POINTWISE ERGODIC THEOREM. This ergodic theorem is sufficiently strong to imply both Weyl's theorem and —a result we will need shortly— the KRONECKER-WEYL THEOREM. A formulation of Birkhoff's theorem appears in the appendix, as well as a proof that irrational rotations are ergodic.

It is the possibility that a transformation can *fail* to be ergodic which gives rise to the instability of frequencies $F^{s,s'}_{\mathbf{m}:\mathbf{m}'}(d, d')$ as the seeds s and s' are varied.

Non-equidistribution. When $\mathbf{m} = 10^{p/q}$ is the multiplier then $\rho = \rho_\alpha$ is a rational rotation. Take, for example, $\alpha = 1/3$. The ρ-orbit of any "seed" $z \in \mathbb{K}$ thus has exactly 3 points. The observed frequency of a digit d does not remain constant as the seed is varied; for example, the frequency that the z-orbit hits I_2 is $\frac{1}{3}$ for any $z \in I_2$, but is zero for $z \in I_3$.

We also notice another failure—that $\rho_{1/3}$ is not ergodic. Given any $A \subset \mathbb{K}$, the set $B := A \cup \rho^{-1}(A) \cup \rho^{-2}(A)$ is ρ-invariant. Taking an A of positive length less than $1/3$ gives a non-trivial invariant set B. This explains lack of constancy for frequencies; the orbit of any seed $z \in B^c$ hits B with zero frequency and yet the measure of B is positive.

Happily, any transformation can be decomposed into disjoint ergodic transformations. In the case of rotation by $1/3$, write the circle as a disjoint union

$$\mathbb{K} = \bigsqcup_{0 \le c < 1/3} K_c, \quad \text{where } K_c := \{c, c \oplus \tfrac{1}{3}, c \oplus \tfrac{2}{3}\}, \tag{3.4a}$$

of ρ-invariant 3-point sets K_c. For a point z in the circle, the particular $K = K_c$ containing z is called the *ergodic component* of z because, when K is equipped with this ρ-invariant probability measure ν_K,

$$\nu_K(\{c\}) = \nu_K(\{c \oplus \tfrac{1}{3}\}) = \nu_K(\{c \oplus \tfrac{2}{3}\}) = \tfrac{1}{3}, \tag{3.4b}$$

then the system $(\rho : K, \nu_K)$ is ergodic. As K ranges over the ergodic components of ρ, the measures ν_K form a disintegration of arclength measure λ. This is why (3.4) is called the *ergodic decomposition* of $(\rho : \mathbb{K}, \lambda)$.

Joint frequencies. Frequencies $F_{2:3}(d, d')$ are measurements made on the direct product of two "multiply maps" and so $\psi \times \psi$ (see commutative diagram 3.2)

† "T-invariant" can be taken to mean either $T^{-1}(B) = B$, or the superficially weaker statement that the symmetric difference $B \Delta T^{-1}B$ is a nullset. It is easy to check that if B satisfies the latter invariance, then B can be altered by a nullset to fulfill the stronger invariance.

carries this system to a rotation of the torus $\mathbb{K}^{\times 2} = \mathbb{K} \times \mathbb{K}$,

$$\rho_\alpha \times \rho_\beta: (z, z') \mapsto (z \oplus \alpha, z' \oplus \beta),$$

where $\alpha = \log(2)$ and $\beta = \log(3)$. So $F_{2:3}^{s;s'}(d, d')$ is simply the frequency that the $\rho_\alpha \times \rho_\beta$-orbit of $(z, z') := (\psi(s), \psi(s'))$ hits rectangle $I_d \times I_{d'}$.

Analogous to the one-dimensional case, if $\rho_\alpha \times \rho_\beta$ is "random" —is ergodic— one might expect this frequency to be the measure of the rectangle, that is, its area $\mathrm{fr}(d)\mathrm{fr}(d')$. This, as in Weyl's theorem, is a statement of equidistribution and is enunciated in the well-known Kronecker-Weyl theorem.

Real numbers $\{\alpha_1, \ldots, \alpha_L\}$ are *rationally independent* if they are linearly independent over $\mathbb{Q}$; the only integral solution to $N_1\alpha_1 + \cdots + N_L\alpha_L = 0$ is the all-zero $N_1 = 0, \ldots, N_L = 0$ solution. On $\mathbb{K}^{\times L}$, the L-dimensional torus of L-dimensional volume 1, a sequence of points $x_0, x_1, \ldots$ is *equidistributed* if it hits every sub-block $I_1 \times \cdots \times I_L$ of the torus with frequency equal to the subblock's volume.

Kronecker-Weyl Theorem. *Numbers $1, \alpha_1, \ldots, \alpha_L$ are rationally independent if and only if under the action of rotation $\rho_{\alpha_1} \times \cdots \times \rho_{\alpha_L}$ on the L-dimensional torus, every orbit is equidistributed. (This is also equivalent to ergodicity of this toral-rotation.)*

As an application, since $\{1, \log(2), \log(3)\}$ are rationally independent, we have that

$$F_{2:3}^{s;s'}(d, d') = \log\left(1 + \frac{1}{d}\right)\log\left(1 + \frac{1}{d'}\right),$$

regardless of the seeds s and s'.

What happens with non-independence? Take, for example, $\alpha = \log(2)$ and $\beta = \log(5)$. Here, the equality $2 \cdot 5 = 10$ translates to the non-zero rational relation $\alpha + \beta = 1$, and so $\rho_\alpha \times \rho_\beta$ is not ergodic. To see this, regard the 2-dimensional torus $[0, 1) \times [0, 1)$ as a square. Then the ergodic components are "wrap-around" diagonals running northwest to southeast on this square; each number $c \in [0, 1)$ gives us a diagonal K_c which is $(\rho_\alpha \times \rho_\beta)$-invariant:

$$\mathbb{K}^{\times 2} = \bigsqcup_{0 \le c < 1} K_c, \quad \text{where } K_c := \{(x, y) | x \oplus y = c\}.$$

Each diagonal K_c is topologically a circle whose ergodic measure λ_c is simply normalized arclength,

$$\lambda_c(A) := \frac{\mathrm{Length}(A \cap K_c)}{\mathrm{Length}(K_c)}, \quad \text{where } A \subset \mathbb{K}^{\times 2}.$$

In this **2:5** case, the system $(\rho_\alpha \times \rho_\beta: K_c, \lambda_c)$ is isomorphic to irrational rotation on the diagonal "circle" K_c by rotation number α (or, equivalently, β).

Weyl's theorem then asserts that $F_{2:5}^{s;s'}(d, d')$ is the λ_c-length of the intersection of rectangle $I_d \times I_{d'}$ with *the* diagonal K_c which contains the ψ-image of (s, s'). This is the diagonal with $c := \psi(s) \oplus \psi(s')$. For example, the **2:5** column-pair of Table P.3 corresponds to $s = s' = 1$ and so $c = 0$; therefore $K = K_c$ is the main diagonal $x + y = 1$. It should be easy to figure out the frequency of seeing a doubled digit (d, d), no?

Scanning the **2** and **5** columns of Table P.3 shows that "3" occurs doubled, $\mathbf{2}^5 = \underline{3}2$ and $\mathbf{5}^5 = \underline{3}125$. It is not surprising that no other digit occurs doubled since, if $d \neq 3$, then the main diagonal does not pass through the square $I_d \times I_d$ and consequently the frequency of (d, d) is zero. Conversely, if you traverse the main diagonal K by heading southeast, you will enter the $I_3 \times I_3$ square at the point $(\log(3), 1 - \log(3))$ and exit the square at $(1 - \log(3), \log(3))$. The fraction of your time spent inside the square was

$$F_{2:5}(3,3) = [1 - \log(3)] - \log(3) = 1 - \log(9),$$

which is irrational. (By the way, this picture of a northwest to southeast diagonal explains the three points of non-differentiability of the function $g(s) := F_{2:5}^{s:s}(3,3)$ mentioned in (3.1). As the seed s increases, the corresponding diagonal moves northeast and the length of it within the $I_3 \times I_3$ square varies linearly *except* when the diagonal passes over a corner of $I_3 \times I_3$.)

The general case. What are the ergodic components of toral rotation $\rho_{\alpha_1} \times \cdots \times \rho_{\alpha_L}$?—it being clear now that to compute joint-frequencies of the form $F_{\alpha_1:\cdots:\alpha_L}^{z_1,\dots,z_L}$, one needs a geometric description of these ergodic components. The Kronecker-Weyl theorem tells us how, in principle, to do this. However, computing frequencies for a specific set of α_i will require a bit of geometry.

Consider a maximal subset of $\{\alpha_1, \dots, \alpha_L\}$ which, were "1" adjoined, would comprise a rationally independent collection; suppose it has $M \geq 1$ members. Let G be the closure of the orbit of $(0, \overset{L}{\dots}, 0)$; in other words, the closure in the L-dimensional torus of tuples

$$((n\alpha_1)_{\text{mod}\,1}, (n\alpha_2)_{\text{mod}\,1}, \dots, (n\alpha_L)_{\text{mod}\,1})$$

as n ranges over the integers. This G will be an M-dimensional subtorus (indeed, a subgroup of $\mathbb{K}^{\times L}$). So the ergodic components of $\rho_{\alpha_1} \times \cdots \times \rho_{\alpha_L}$ are merely the translates of this subtorus. Consequently, the frequency

$$F_{\alpha_1:\cdots:\alpha_L}^{z_1,\dots,z_L}(d_1, \dots, d_L)$$

is the M-dimensional "area" of the cross-section of the block (rectangular parallelepiped) $I_{d_1} \times \cdots \times I_{d_L}$ which is sliced by the particular subtorus that contains the point $(z_1, \dots, z_L)$. This is the subtorus $G \oplus (z_1, \dots, z_L)$.

Even for the case of three rotation numbers with exactly two being rationally independent, determining frequencies might be tricky. Here one wants to compute the area of the region which is the intersection of a plane slicing through a 3-dimensional block. Depending on the tilt of this plane, the intersection could be a triangle, a quadrilateral, a pentagon or a hexagon.

So we end up by finding another point of commonality among the problems of Poncelet, Tarski and Gelfand: Existence of a natural invariant measure does not, alas, imply that the measure is going to be easy to compute...

§A APPENDIX. Here we give a terse history of the three questions, as well some related conundrums. A technical note: In this article all measures were tacitly Borel measures, and all sets and functions were Borel measurable.

Poncelet's theorem, history. Jean-Victor Poncelet was an officer of engineers in Napoleon's army during the invasion of Russia. Like other invasions of Russia this one failed, and on November eighteenth of 1812, Poncelet was captured in the retreat from Moscow. During his 1812–1814 captivity, which he later wrote about in PONCELET 1862, he developed the notions of Projective Geometry which led him to the Closure theorem. (For a fuller history, see BELL 1937.)

Commentary. The idea of using an invariant to prove Poncelet's theorem goes back to Jacobi and Bertrand, according to SCHOENBERG 1983; and Schoenberg's beautiful paper gives an efficient proof using an *invariant integral*—which in this context is the same thing as an invariant measure.

Nonetheless, Schoenberg's proof uses non-elementary notions: The Brouwer fixed-point theorem and facts in projective geometry concerning the polar line of a point. In contrast, by focusing on an invariant *measure* for Poncelet's transformation, even the small amount of computation in Schoenberg is entirely avoided. This is merely a change in viewpoint, not a change in proof, yet it has the advantage of yielding a theorem even in the case that the given pair of ellipses have no circuminscribed polygon, and thereby suggests the link with Gelfand's question. By avoiding Brouwer's theorem, which is non-constructive in nature, one can actually compute the invariant density, eg. (1.10), and thus in principle determine whether the rotation number is rational so as to ascertain whether a circuminscribed polygon exists.

An extensive history and "pre-history" of Poncelet's theorem, as well as several proofs, appear in a paper by Bos, Kers, Oort, Raven 1987, a paper which I came upon after this article was written.

There is a connection between the Poncelet-measure determined by confocal ellipses **C** and **E**, and the invariant "billiard measure" of an elliptical billiard table. The ergodic components of the billiard measure of **C** turn out to be the 1-parameter family of Poncelet **CE**-measures obtained by varying **E**, the inner confocal ellipse. Billiard measure is defined in a companion article to this one, [King], where it is used to study a billiard question on a table with a cusp at infinity.

Plank Problem, history. Archimedes proved in proposition 3 of Book II of *On the Sphere and the Cylinder*, see Heath 1897, that a plane orthogonal to a diameter of a sphere cuts the sphere into two regions with areas in the same ratio as the lengths of the two pieces of the diameter. Since the areas of the sphere and cylinder were known to Archimedes, this proposition is tantamount to showing that radial projection is area-preserving. I am indebted to C. E. Thompson for indicating this proposition to me.

Given here was the elementary case of *Tarski's Plank Problem*, proposed in 1932. What Tarski had asked was this:

Suppose a convex compact region **D** *in the plane is covered by countably many planks. Must the sum of their widths dominate the width of* **D**?

(The *width* of a shape is the width of the narrowest single plank which covers it.) This problem does not immediately follow from the disk case, since some convex shapes —the equilateral triangle being a notorious example— have a width exceeding the diameter of its inscribed circle. Almost 20 years later, Tarski's problem was solved by an ingenious argument of Thøger Bang, 1951. The elementary version of Tarski's problem suggests the following question which, to my knowledge, is open.

Question A.1. What is the infimum, $W(r)$, of total widths $\sum_{n=1}^{\infty} \text{Width}(P_n)$ taken over all systems $\{P_n\}_1^{\infty}$ of planks which cover the annulus of outer-radius 1 and inner-radius r?

Gelfand's Question, history. The "frequencies of $\langle\langle 2^n \rangle\rangle$" problem appears on page 37 of Avez 1966, where it is attributed to Gelfand.

Weyl's theorem is usually proved by noting that definition (3.3) of equidistribution is equivalent to this formulation: The sequence $x_0, x_1, \ldots$ is *equidistributed* in the circle $\mathbb{K}$ if

$$\textit{For any continuous } g\colon \mathbb{K} \to \mathbb{C}\colon \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} g(x_n) = \int g\, d\lambda \tag{A.2a}$$

Weyl's theorem concerns those g for which, for each z, the sequence $\{z \oplus n\alpha\}_{n=0}^{\infty}$ is equidistributed. Since the set of such g forms a closed subspace (in the supremum-norm) of the space of all continuous functions, the Stone-Weierstrass theorem tells us that it is enough to check this when g is a group-character, where it is easily verified.

The Pointwise Ergodic Theorem. There is a more general "equidistribution theorem" which applies when T is a measure-preserving transformation on probability space (X, μ). In the special case when T is ergodic, Birkhoff's theorem states that for any $g \in L^1(\mu)$ the limit and equality below exist.

$$\text{For a.e. } x \in X\colon \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} g(T^n x) = \int g\, d\mu. \tag{A.2b}$$

(The nullset of "bad" x violating equality will generally depend on g; there is no universal full-measure set working for all functions. An earlier ergodic theorem, convergence in the L^2-norm, was proven by von Neumann.) The pointwise ergodic theorem is a far-reaching generalization of the Strong Law of Large Numbers (SLLN). It also quickly implies Weyl's theorem —that continuous functions g have *no* bad points— once irrational rotations are shown to be ergodic. We show that now.

Proposition A.3. *Any irrational rotation $\rho = \rho_\alpha$ is ergodic.*

Proof: Suppose ρ-invariant set B has positive mass. Take a short open interval I in the circle such that

$$\lambda(I \cap B) \geq (.99)\lambda(I).$$

By the density of the orbit of an endpoint of I, we can find times $k_1, \ldots, k_M$ such that the translated intervals $\rho^{k_m}(I)$ are disjoint, and

$$\lambda\left(\bigsqcup_{m=1}^{M} \rho^{k_m}(I)\right) \geq .98.$$

But $B = \bigcup_{m=1}^{M} \rho^{k_m}(B) \supset \bigsqcup_{m=1}^{M} \rho^{k_m}(I \cap B)$. Consequently

$$\begin{aligned} \lambda(B) &\geq \sum_{m=1}^{M} \lambda(\rho^{k_m}(I \cap B)) \\ &\geq \sum_{m=1}^{M} (.99)\lambda(\rho^{k_m}(I)) \geq (.99)(.98) > .97. \end{aligned}$$

Of course this ".97" could have been made as close to 1 as desired, so $\lambda(B^c) = 0$. □

The upshot is that Birkhoff's theorem simultaneously contains the probabilistic equidistribution of SLLN and the topological equidistribution of Weyl.

Poncelet-measure is unique. Although we will not go into detail (see for example PETERSEN 1983), ergodicity and equidistribution ideas answer question 1.12 affirmatively by showing that

Under an irrational rotation ρ_α, arclength λ is the unique invariant Borel probability measure.

This gives uniqueness of Poncelet-measure μ_h when the rotation number α, coming from ellipses **C** and **E**, is irrational. But even in the case when α is rational, μ_h must still simply be the image, under φ, of the circle's arclength measure: Dilating the inner ellipse **E** slightly causes the rotation number to vary continuously, and only requires a small change in the affine map $\mathscr{A}$. Hence the Poncelet-measure for a **CE**-pair with rational α is the limit of Poncelet-measures from nearby irrational rotation numbers, and all these measures are the φ-image of arclength.

The Kronecker–Weyl theorem. The preceding proposition derived ergodicity of an irrational rotation directly from the density of every orbit. Exactly the same proof would show that a toral rotation is ergodic, once orbits are known to be dense. Also analogous to the 1-dimensional case of the circle, ergodicity together with Birkhoff's theorem can be used to demonstrate equidistribution, which in the multi-dimensional case is Kronecker-Weyl. Thus the fundamental idea in proving K – W is to show that rational independence of the rotation numbers implies that the orbit of the origin is dense.

Let us conclude by proving density in the 2-dimensional case, so as to illustrate how density can be lifted from the 1-dimensional case.

Warmup for the K – W theorem. *Suppose numbers $\alpha, \beta, 1$ are rationally independent. Then under toral-rotation $\rho_\alpha \times \rho_\beta$, the orbit of $(0,0)$ is dense.*

Proof: Fixing ε we show that the (α,β)-orbit of $(0,0)$ is ε-dense. Pick $N \geq 1$ such that

$$(\hat{\alpha}, \hat{\beta}) := [\rho_\alpha \times \rho_\beta]^N (0,0)$$

is within ε of $(0,0)$ in the torus; this, by the same Pigeon-hole argument used in (1.11). This $\hat{\alpha}$ is of the form $N\alpha$ minus an integer, and likewise for $\hat{\beta}$. Rational independence implies that $(\hat{\alpha}, \hat{\beta})$ is not $(0,0)$ and thus

the $(\hat{\alpha}, \hat{\beta})$-orbit of $(0,0)$ is ε-dense in the line

$$L := \{(x,y) \in \mathbb{R} \times \mathbb{R} \mid x\hat{\beta} = y\hat{\alpha}\}$$

in the "unwrapped" torus ie., in the plane. Wrapping the plane back up, this line L winds densely around the torus, since its slope $\hat{\alpha}/\hat{\beta}$ is irrational; this again follows from rational independence. In consequence, the $(\hat{\alpha}, \hat{\beta})$-orbit of $(0,0)$ is ε-dense in the torus, and therefore so is the (α,β)-orbit. □

Final remark. We know the frequency of leading-digit "9" in the doubling sequence $2, 4, 8, 16, 32, \ldots$. But what is the frequency of leading-digit "9" in the sequence

$$2, 4, 16, 256, 65536, \ldots \qquad (*)$$

where —instead of doubling— the next term is always the *square* of the current term?

Map ψ of diagram 3.2 carries squaring to the map $S(x) := 2x \pmod 1$ on $[0, 1)$. This 2-to-1 map, an endomorphism of the circle group, preserves Lebesgue measure and is ergodic, and so Birkhoff's theorem tells us that the orbit of *almost* every point x visits each I_d interval with Lebesgue frequency. But sequence (*) asks this question of a *specific* point, $x = \log(2)$. Is $\log(2)$ a "bad" point (relative to the I_d intervals) for Birkhoff's theorem?

To this day, nobody knows...

References

Library call numbers, where available, follow each reference.

A. Avez, Ergodic theory of Dynamical Systems (vol. 1), University of Minnesota, Institute of Technology, 1966.

Thøger Bang, *A solution to the "Plank Problem"*, Amer. Math. Soc. Proceedings 2 (1951). cn: *QA1.A5215*

E. T. Bell, Men of Mathematics, Simon and Shuster, New York, 1937. cn: *925.1 B433m*

H. J. M. Bos, C. Kers, F. Oort, D. W. Raven, *Poncelet's closure theorem*, Expo. Math 5 (1987), 289–364. cn: *QA 1 .E96*

T. L. Heath, The works of Archimedes, Cambridge University Press, 1897. cn: *QA31A72*

J. L. King, *Billiards inside a cusp*. Math. Intelligencer (to appear). cn: *QA1.M38*. An earlier version appears in: MSRI Preprint Series, #035-94 (1994).

K. Petersen, Ergodic Theory, Cambridge University Press, 1983. cn: *QA313 .P47*

Jean-Victor Poncelet, Traité des propriétés projectives des figures; ouvrange utile a ceux qui s'occupent des applications de la geometrie descriptive et d'operations geometriques sûr le terrain, Gauthier-Villars, Paris, 1866. Second edition. (First ed. published 1822) cn: *QA471.P65*

Jean-Victor Poncelet, Applications d'analyse et de géométrie qui ont servi, en 1822, de principal fondement au Traité des propriétés projectives des figures, Paris, 1862.

I. J. Schoenberg, *On Jacobi-Bertrand's proof of a Theorem of Poncelet*, in "Studies in Pure Mathematics," Birkhauser, 1983, pp. 623–627. cn: *QA7 .S845*

A. Tarski, *Further remarks about the degree of equivalence of polygons* (in Polish), Odbitka Z. Parametru. 2 (1932), 310–314.

Department of Mathematics
University of Florida
Gainesville, FL 32611-2082
squash@math.ufl.edu