FUCHSIAN GROUPS AND ERGODIC THEORY*†

BY
EBERHARD HOPF

Introduction. Let Ω be the phase space of a dynamical system. We suppose that every motion can be continued along the entire time-axis. Thus we are concerned with a steady flow in Ω. The following concepts are of fundamental significance for the study of dynamical flows.

(a) There exists a curve of motion everywhere dense on Ω.

The existence of such a motion is known under the name of regional transitivity. We now suppose that a measure m in the sense of Lebesgue invariant under the flow exists on Ω. Such a measure is usually defined by an invariant phase element dm. The following property is stronger than (a).

(b) The curves of motion not everywhere dense on Ω form a point set on Ω of m-measure zero.

Still stronger and more important than (b) is strict ergodicity. We suppose $m(\Omega)$ to be finite.

(c) Let $f(P)$ be an arbitrary m-summable function on Ω. The time-average of $f(P)$ along a curve of motion is then, in general, equal to $\int_\Omega f(P)dm/m(\Omega)$, the exceptional curves forming a point set on Ω of m-measure zero.

How these concepts are interrelated is seen most clearly if we state them in the following way.

(a′) Every open point set on Ω that is invariant under the flow is everywhere dense on Ω.

(b′) Every open point set on Ω that is invariant under the flow has the measure $m(\Omega)$.

(c′) Every m-measurable point set on Ω that is invariant under the flow has either the m-measure zero or $m(\Omega)$.

The latter property of a flow is called *metric transitivity*.‡ Its importance rests

* Presented to the Society, September 13, 1935; received by the editors August 10, 1935.

† To August Kopff.

‡ G. D. Birkhoff and P. Smith, *Structure analysis of surface transformations*, Journal de Mathématiques, (9), vol. 7 (1928), pp. 345–379.

in its equivalence to (c).* The problem whether a given flow is metrically transitive or not is, in general, as difficult as it is interesting. Beyond simple examples progress has recently been made in the direction of certain geodesics problems. Let Σ be a surface (two-dimensional manifold) of class C_3. We denote by p an arbitrary point on Σ and by ϕ the angle measuring directions through p. Every geodesic on Σ is supposed to be continuable indefinitely in both directions. The line elements

$$P = (p, \phi)$$

then constitute the phase space Ω associated with Σ. To the uniform motion along the geodesics on Σ there corresponds a steady flow on Ω. The element of volume

$$dm = d\sigma d\phi,$$

$d\sigma$ being the element of area, is well known to be invariant under the flow.

The particular surfaces† Σ considered in this paper are those of constant negative curvature and of finite connectivity. Their geodesics are supposed to satisfy the above condition of unlimited continuability. Differential geometry shows that there exists a one-to-many correspondence between Σ and the interior $|z| < 1$ of the unit circle such that the elements of length ds and area $d\sigma$ go over into the NE-elements in $|z| < 1$,

$$ds = 2(1 - z\bar{z})^{-1}|dz|, \quad d\sigma = 4(1 - z\bar{z})^{-2}dxdy$$

respectively. The geodesics on Σ go over into the arcs of orthogonal circles within $|z| < 1$ (NE-straight lines). The covering transformations are known to form a Fuchsian group $\mathcal{G}$ of linear substitutions S transforming $|z| \leqq 1$ into itself. $|z| = 1$ is the principal circle of $\mathcal{G}$.‡ A more general notion of the

* See the literature on the ergodic theorem, viz.

G. D. Birkhoff, *Proof of a recurrence theorem for strongly transitive systems*, Proceedings of the National Academy, vol. 17 (1931), pp. 650–660.

T. Carleman, *Application de la théorie des équations intégrales linéaires aus systèmes d'équations différentielles non-linéaires*, Acta Mathematica, vol. 59 (1932), pp. 63–87.

E. Hopf, *On the time average theorem in dynamics*, Proceedings of the National Academy, vol. 18 (1932), pp. 93–100.

A. Khintchine, *Zu Birkhoff's Lösung des Ergodenproblems*, Mathematische Annalen, vol. 107 (1933), pp. 485–488.

J. v. Neumann, *Proof of the quasi-ergodic hypothesis*, Proceedings of the National Academy, vol. 18 (1932), pp. 70–82.

† Surface = Riemannian manifold. We refer, in this connection, to the papers by P. Koebe, *Riemannsche Mannigfaltigkeiten und nichteuklidische Raumformen*, I–V, Sitzungsberichte der Preussischen Akademie, 1927–30.

‡ An elementary introduction into the theory of Fuchsian groups is found in L. R. Ford, *Automorphic Functions*, New York, 1929.

Fuchsian group will be considered here. In order to include one-sided surfaces we shall admit anti-analytic substitutions, i.e., analytic substitutions of $\bar{z}$. On identifying all points of $|z| < 1$ equivalent under any Fuchsian group $\mathcal{G}$ one defines, conversely, an associated surface Σ.

$\mathcal{G}$ always possesses a fundamental region R. It can be chosen so as to form a NE-convex polygon bounded by a finite number of segments of NE-straight lines and a finite number of arcs of $|z| = 1$. The images of R under all S of $\mathcal{G}$ cover the whole of $|z| < 1$ simply.

We are to distinguish between two fundamentally different kinds of Fuchsian groups $\mathcal{G}$, and, therefore, of surfaces Σ. $\mathcal{G}$ and Σ are of the first kind if the surface area of Σ or, what is the same, the NE-area of the fundamental region R is finite. In the opposite case, $\sigma(\Sigma) = \infty$, we speak of the groups and surfaces of the second kind.* If $\mathcal{G}$ is of the first kind, R has no arcs of $|z| = 1$ on its boundary. Its vertices lie partly in $|z| < 1$, partly on $|z| = 1$, the angle being zero in the latter case. A well known example is offered by the case where R is bounded by a regular NE-polygon with the sum of the angles equal to 2π. If the $4p$ sides $(p > 1)$ be paired in a certain way, Σ represents a closed two-sided surface of genus p. Another well known example is furnished by the modular group where R is bounded by a NE-triangle with one vertex on $|z| = 1$. The surface Σ has, in this case, a cuspidal singularity. For every group $\mathcal{G}$ of the second kind, however, R has at least one arc of $|z| = 1$ on its boundary and Σ possesses, accordingly, at least one funnel.

For surfaces Σ of the first kind, the regional transitivity (a) has been proved,† in various degrees of generality, by Artin, J. Nielsen, Koebe and Löbell, whereas Myrberg discovered the property (b). It is only recently that Hedlund‡ succeeded in proving the deeper property of metric transitivity of the two examples mentioned above. It is the purpose of the present paper to develop an entirely novel and simple method that yields a proof of the metric transitivity for all surfaces Σ of the first kind.

---

* This definition is readily found to be in agreement with the one usually given.

† E. Artin, *Ein mechanisches System mit quasiergodischen Bahnen*, Abhandlungen des Mathematischen Seminars, Hamburg, vol. 3 (1924), pp. 170–175.

J. Nielsen, *Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen*, Acta Mathematica, vol. 50 (1927), pp. 189–358.

P. Koebe, loc. cit., IV (1929), p. 414.

F. Löbell, *Über die geodätischen Linien der Clifford-Kleinschen Flächen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 572–607.

P. J. Myrberg, *Ein Approximationssatz für die fuchsschen Gruppen*, Acta Mathematica, vol. 57 (1931), pp. 389–409.

‡ G. Hedlund, *Metric transitivity of the geodesics on closed surfaces of constant negative curvature*, Annals of Mathematics, (2), vol. 35 (1934), p. 787; *A metrically transitive group defined by the modular group*, American Journal of Mathematics, vol. 57 (1935), pp. 668–678.

THEOREM. *For all surfaces Σ of the first kind, the flow associated with the geodesics problem on Σ is metrically transitive.*

On surfaces Σ of the second kind (treated in §5), the geodesics show an entirely different behavior. The corresponding theorem can be stated without reference to the phase space Ω.

The geodesics through an arbitrarily given point p of Σ disappear, for almost all directions through p, into the funnels of Σ.

By this we mean that the corresponding NE-straight lines end on one of the arcs of $|z|=1$ belonging to the boundary of R, or on one of the images under $\mathcal{G}$ of those arcs. The theorem, therefore, merely states that those arcs and their images form a set on $|z|=1$ of the same measure as the unit circle itself, which statement, in contrast to the theorem concerning surfaces of the first kind, is most readily proved.

The essential tools used in this paper are potential theory and the NE-metric in $|z|<1$.

1. **Preliminaries on Fuchsian groups.** The cross ratio

$$[z_1, z_2, z_3, z_4] = \frac{z_3 - z_1}{z_3 - z_2}\,\frac{z_4 - z_2}{z_4 - z_1}$$

is unchanged by an analytic linear substitution, whereas it goes over into its conjugate value under an anti-analytic one. The same holds for the differentials

$$\frac{dz_2}{z_1 - z_2}\,\frac{z_3 - z_1}{z_3 - z_2} = -\,[z_1, z_2, z_3, z_2 + dz_2] \tag{1}$$

and

$$(z_1 - z_2)^{-2} dz_1 dz_2 = -\,[z_1, z_2, z_1 + dz_1, z_2 + dz_2]. \tag{2}$$

Only those substitutions S will be considered in the sequel which leave $|z| \leqq 1$ invariant. The relation

$$S(1/\bar{z}) = 1/\overline{S(z)}$$

yields the invariant

$$\left|\frac{w - z}{1 - \bar{z}w}\right|^2 = [z, 1/\bar{z}, w, 1/\bar{w}], \tag{3}$$

and, according to (1), the differential invariant

$$\frac{dz}{1 - z\bar{z}}\,\frac{1 - \bar{z}w}{w - z} = [1/\bar{z}, z, w, z + dz]. \tag{4}$$

From (3) and (4) we obtain Poincaré's invariant NE-element of length

$$ds = 2(1 - z\bar{z})^{-1} | dz |. \tag{5}$$

The corresponding NE-element of area is then

$$d\sigma = 4(1 - z\bar{z})^{-2} dxdy. \tag{6}$$

The absolute value of (2) with $z_1 = \zeta$, $z_2 = z$, after being divided by (5), finally furnishes the invariant Poisson differential

$$\frac{1 - z\bar{z}}{|\zeta - z|^2} | d\zeta |. \tag{7}$$

A group $\mathcal{G}$ of substitutions S preserving $|z| < 1$ is called Fuchsian if it is infinite and discontinuous, i.e., if it contains no infinitesimal S. From now on, we suppose $\mathcal{G}$ to be Fuchsian. A point z, $|z| < 1$, and all its equivalent points $S(z)$ determine the same point of the surface Σ. Associated with $\mathcal{G}$ is a group Γ of contact transformations T in the space of the line elements

$$(z, \phi), \quad \phi = \arg (dz),$$

where T has the form

$$(z, \phi) \rightarrow (S(z), \ \phi + \arg S'(z)) \tag{8}$$

in the case of an analytic S and a similar form when S is anti-analytic. Equivalent line elements define the same point P in the phase space Ω. As S preserves angles, the transformations T leave the volume element

$$dm = d\sigma d\phi \tag{9}$$

in the space of line elements invariant.

We now introduce the new coordinates

$$(\eta_1, \eta_2, s); \ | \eta_1 | = | \eta_2 | = 1, \ \eta_1 \neq \eta_2, \ -\infty < s < \infty,$$

in the space of line elements. η_1 and η_2 are initial and end point, respectively, of the sensed NE-straight line passing through (z, ϕ). Let s denote the NE-distance of z from the point z_0 bisecting the circular arc (η_1, η_2), the sign of s being plus (minus) if z is met after (before) z_0 on (η_1, η_2). The correspondence between (z, ϕ) and (η_1, η_2, s) is evidently one-to-one. In the new coordinates, the transformations T are easily seen to be of the form

$$(\eta_1, \eta_2, s) \rightarrow (S(\eta_1), S(\eta_2), s + f_S(\eta_1, \eta_2)). \tag{10}$$

We now prove that the volume element (9) becomes

$$dm = k \frac{| d\eta_1 | | d\eta_2 |}{| \eta_1 - \eta_2 |^2} ds, \tag{11}$$

k being a positive constant. Indeed we can always find a linear substitution preserving $|z|=1$ which transforms an arbitrarily given line element

$$(z, \phi) = (\eta_1, \eta_2, s)$$

into any other one,

$$(z', \phi') = (\eta_1', \eta_2', s').$$

The associated contact transformation is of the form (8) and therefore leaves (9) invariant. Being of the form (10), it also leaves invariant the right-hand side of (11) in view of the invariance of (2) and (5). Hence, the two sides can differ by a constant factor only.

In the new coordinates the flow associated with the geodesics on Σ is described by the simple formulas

$$P = (\eta_1, \eta_2, s) \rightarrow P_t = (\eta_1, \eta_2, s + t). \tag{12}$$

The invariance of dm under the flow is now a trivial consequence of (11) and (12).

The explicit connection between the coordinates is readily established,

$$s = \log\,[\eta_1, \eta_2, z, z_0], \qquad z_0 = \frac{\eta_1 + \eta_2}{2 + |\eta_1 - \eta_2|},$$

$$\phi = \arg \frac{(z - \eta_1)(z - \eta_2)}{\eta_1 - \eta_2},$$

but it is not needed for our purposes.

2. **Another formulation of the theorem.** A Fuchsian group of the first kind possesses a fundamental region R of finite NE-area. To the subdivision of $|z|<1$ into the NE-congruent parts $S(R)$ there corresponds a subdivision of the (η_1, η_2, s) space into cells congruent to each other under the transformations T of Γ. Each of these cells is a representative of Ω, with

$$m(\Omega) = 2\pi\sigma(\Sigma) < \infty.$$

For the proof of the announced theorem it is sufficient to prove

THEOREM A. *A point set A on the (η_1, η_2) torus which is measurable in the sense of ordinary Lebesgue measure, for which*

$$\iint_A |d\eta_1||d\eta_2| > 0$$

and which is invariant under the simultaneous substitutions $S(\eta_1)$, $S(\eta_2)$ of $\mathcal{G}$, has the measure of the entire torus.

Suppose Theorem A to be proved. We start with a point set on Ω satisfying the hypothesis of the theorem announced in the introduction. According to (12), this set represents, in the (η_1, η_2, s) space, a cylindrical set, i.e., a set on the (η_1, η_2) torus. According to (11), the part of this set within each of the above cells is of positive measure

$$\int\int \frac{|d\eta_1||d\eta_2|}{|\eta_1 - \eta_2|^2},$$

and, therefore, of positive torus measure. The sum of these parts obviously represents a set A in the sense of Theorem A. Hence its complement has the torus measure zero. Regarded as a cylindrical set in the (η_1, η_2, s) space the latter set is, necessarily, of m-measure zero in that space and, therefore, in Ω.

From now on we may without loss of generality confine ourselves to two-sided surfaces Σ, i.e., to the case where $\mathcal{G}$ contains only analytic substitutions S. For, let S denote the analytic and $\bar{S}$ the anti-analytic substitutions of $\mathcal{G}$. The S's form a subgroup $\mathfrak{g}$ of $\mathcal{G}$ and each $\bar{S}$ can be written in the form $\bar{S} = S\bar{S}_0$ where $\bar{S}_0$ is a fixed $\bar{S}$. As

$$R + \bar{S}_0(R)$$

is evidently a fundamental region for $\mathfrak{g}$, this group is seen to be again a Fuchsian group of the first kind. If Theorem A holds for $\mathfrak{g}$, it holds also for $\mathcal{G}$.

Let now $U(\eta_1, \eta_2)$ be the function on the torus which equals zero on the set A of Theorem A and one elsewhere. U is measurable and invariant under $\mathcal{G}$,

$$U(S(\eta_1), S(\eta_2)) = U(\eta_1, \eta_2). \tag{13}$$

It is to be proved that $U = 0$ except on a torus set of measure zero. We transform our problem once more by introducing harmonic functions. The Poisson integral

$$U(z, \gamma) = \frac{1}{2\pi}\int_{|\zeta|=1} U(\zeta, \gamma)\frac{1 - z\bar{z}}{|\zeta - z|^2}|d\zeta| \tag{14}$$

represents, for amost all γ on $|\gamma| = 1$, a harmonic function of z in $|z| = 1$ and, for every such z, a bounded and measurable function of γ on $|\gamma| = 1$. Furthermore, the function

$$U(z, w) = \frac{1}{2\pi}\int_{|\gamma|=1} U(z, \gamma)\frac{1 - w\bar{w}}{|\gamma - w|^2}|d\gamma| \tag{14'}$$

is, for $|z| < 1$ and $|w| < 1$, harmonic in z as well as in w. (14′), combined with (14), could, of course, be written as one double Poisson integral. In view of the invariance of the Poisson differentials, the invariance (13) under $\mathcal{G}$ of the "torus values" of $U(z, w)$ implies the invariance of the function itself,

$$U(S(z), S(w)) = U(z, w), \tag{15}$$

for all S of $\mathcal{G}$.

Formulas (14) and (14′) show that $U(z, w) \equiv 0$ implies the vanishing of the torus values $U(\zeta, \gamma)$ up to a torus set of measure zero. To prove Theorem A it therefore suffices to prove

THEOREM B. *Suppose that $U(z, w) \geqq 0$ is bounded and harmonic in z as well as in w, $|z| < 1$, $|w| < 1$, and that U satisfies* (15) *for all S of $\mathcal{G}$. If the torus values of U vanish on a set of positive torus measure then U vanishes identically.*

We have to specify in what sense the torus values $U(\zeta, \gamma)$ may be regarded as limit values of $U(z, w)$. If $u(z)$ is a bounded harmonic function of a single point z, $|z| < 1$, we have

$$\text{l.i.m. } u(r\zeta) = u(\zeta) \tag{16}$$

on $|\zeta| = 1$.* An analogue for harmonic functions of two points is quite similarly proved,

$$\underset{r,\rho \to 1}{\text{l.i.m.}}\ U(r\zeta, \rho\gamma) = U(\zeta, \gamma) \tag{16′}$$

on the torus $|\zeta| = |\gamma| = 1$.

3. **Auxiliary theorems.** In the sequel we denote by K_l the interior of the circle about $z = 0$ with the NE-radius l. A simple computation of the NE-area yields the formula

$$\sigma(K_l) = \pi(e^l + e^{-l} - 2). \tag{17}$$

For the validity of Lemma 1 we assume that $z = 0$ is interior to some R.†

LEMMA 1. *If a set B in $|z| < 1$ is invariant under all S of $\mathcal{G}$, and if R is a fundamental region for $\mathcal{G}$, then, for l sufficiently large,*

$$\frac{\sigma(BK_l)}{\sigma(K_l)} < a\sigma(BR),$$

a being a positive constant depending only on $\mathcal{G}$.

This lemma is not quite trivial in the case where R has vertices on $|z| = 1$ and is used mainly to take care of the slight complications arising from this

* l.i.m. means limit in the mean of order two.

† If it is not, by a suitable linear substitution, we can always move a given interior point of R into the origin. It may well be mentioned that the origin plays here a mere auxiliary role.

case in the proof of Theorem B. The proofs of this and of the following lemmas 2 and 3 are given in the next §4.

LEMMA 2. *Let $u(z) \geqq 0$ be bounded and harmonic in $|z| < 1$, and let*

$$u(S(z)) = u(z)$$

hold for all S of $\mathcal{G}$. If the boundary values of u on $|z| = 1$ vanish on a set of positive measure, $u(z)$ vanishes identically.

This is a very simple special case of Theorem B. The principal difficulty in the proof of that theorem is surmounted by the main

LEMMA 3. *If $U(z, w)$ satisfies all the hypotheses of Theorem* B, *the measure on $|\gamma| = 1$ of the set where $U(0, \gamma) = 0$ is necessarily positive.*

Proof of Theorem B. The set E on $|\gamma| = 1$ where $U(0, \gamma) = 0$ is of positive measure according to Lemma 3. We now make use of Harnack's inequalities for a non-negative harmonic function $u(z)$, $|z| < 1$,

$$e^{-s(z,z')}u(z) \leqq u(z') \leqq e^{s(z,z')}u(z),$$

where $s(z, z')$ denotes the NE-distance of the two points. These inequalities being applied to $U(z, w) \geqq 0$ yield

$$e^{-s(0,z)}U(0, w) \leqq U(z, w) \leqq e^{s(0,z)}U(0, w), \tag{18}$$

which shows that, for a fixed z, the set where the boundary values $U(z, \gamma)$ of $U(z, w)$ on $|w| = 1$ vanish is independent of z; in fact it coincides with E except for a set of measure zero. Now according to (15) we have, for an arbitrary S of $\mathcal{G}$,

$$U(S(0), w) = U(0, S^{-1}(w)).$$

On replacing w by $S^{-1}(w)$ and γ by $S^{-1}(\gamma)$ in (14′) and on taking into account the invariance of the Poisson differential, we infer from (14′) that

$$U(0, S^{-1}(w)) = \frac{1}{2\pi}\int_{|\gamma|=1} U(0, S^{-1}(\gamma)) \frac{1 - w\bar{w}}{|\gamma - w|^2} |d\gamma|.$$

Hence $U(0, S^{-1}(\gamma))$ are the boundary values of $U(0, S^{-1}(w))$. Since $S(E)$ is the set where these boundary values vanish, the equation $S(E) = E$ holds for all S of $\mathcal{G}$ apart from a null set on the unit circle. Considering, in the same way as before, the harmonic function $u(z)$, $|z| < 1$, whose boundary values are zero on E and one elsewhere, we infer that u satisfies the hypothesis of Lemma 2 and, therefore, that $u \equiv 0$, i.e., that E has the measure of the entire unit circle. It then follows from the definition of E that the boundary values of $U(0, w)$ vanish almost everywhere. Hence $U(0, w) \equiv 0$ and, according to (18), $U(z, w) \equiv 0$, which is the desired result.

4. **Proof of Lemmas 1 and 2.** We denote by $N(z, l)$ the number of points $S(z)$ congruent to z which lie in K_l. We first show that, for l sufficiently large,

$$\frac{N(z, l)}{\sigma(K_l)} < a, \tag{19}$$

where $a>0$ depends on $\mathcal{G}$ only. $N(z, l)$ is the number of points $S(z)$ whose NE-distance from the origin does not exceed l,

$$s(0, S(z)) \leqq l.$$

Since

$$s(0, S(z)) = s(S^{-1}(0), z),$$

$N(z, l)$ is not greater than the number of points $S^{-1}(0)$ congruent to the origin with a NE-distance $\leqq l$ from z, provided that the points $S^{-1}(0)$ are different for different substitutions S. This is the case, as the origin is interior to R and as an interior point of R cannot be a fixed point for any S except for the identity transformation. We furthermore know that

$$s(S(0), 0) > b, \qquad S(0) \neq 0,$$

holds, where $b>0$ depends on $\mathcal{G}$ only. Therefore a circle of NE-radius b about any point congruent to the origin contains no other such point. This implies that the number of the different points $S(0)$ with a NE-distance $\leqq l$ from z is less than the number of mutually enclusive circles of NE-radius b which can be placed within a circle of NE-radius $l+b$. Hence

$$N(z, l) < \frac{\sigma(K_{l+b})}{\sigma(K_b)} = \frac{\sigma(K_{l+b})}{\sigma(K_l)\sigma(K_b)}\,\sigma(K_l).$$

Here the first factor tends to $\pi^{-1}(1+e^{-2b}-2e^{-b})^{-1}$ as $l\to\infty$, which proves (19).

We now return to the set B of Lemma 1. By means of the function

$$\phi(z) = \begin{cases} 1 & \text{in } K_l, \\ 0 & \text{elsewhere}, \end{cases}$$

we obtain

$$N(z, l) = \sum_S \phi(S(z)),$$

and therefore

$$\iint_{RB} N(z, l)d\sigma_z = \sum_S \iint_{RB} \phi(S(z))d\sigma_z = \sum_S \iint_{S^{-1}(RB)} \phi(z)d\sigma_z. \tag{20}$$

Since $B = S(B)$ for all S of $\mathcal{G}$,

$$\sum_S S^{-1}(RB) = \sum_S BS^{-1}(R) = B\sum_S S(R) = B,$$

and the right hand side of (20) equals

$$\int\int_B \phi(z)d\sigma_z = \sigma(BK_l). \tag{20'}$$

Lemma 1 obviously follows from (19), (20) and (20′).

Proof of Lemma 2. In the particular case where the boundary of R lies within $|z| < 1$ the lemma is obvious, since then $u(z)$ attains its extrema at some points of $|z| < 1$, i.e., at interior points of $|z| < 1$. If R has vertices on $|z| = 1$ an elementary proof could still be given. We prefer, however, to use tools which seemed unavoidable in the further course of the proof of Theorem B. The auxiliary function

$$h_\epsilon(t) = \begin{cases} (1 - t/\epsilon)^3, & 0 \leqq t < \epsilon, \\ 0, & \epsilon \leqq t, \end{cases} \tag{21}$$

is concave and possesses a continuous second derivative, $t \geqq 0$. We first show that

$$\lim_{l\to\infty} \frac{1}{\sigma(K_l)} \int\int_{K_l} h_\epsilon(u(z))d\sigma_z = \frac{1}{2\pi}\int_{|\zeta|=1} h_\epsilon(u(\zeta))\,|d\zeta| \tag{22}$$

and that

$$\lim_{\epsilon\to 0} \int_{|\zeta|=1} h_\epsilon(u(\zeta))\,|d\zeta| = \operatorname*{meas}_{|\zeta|=1}[u(\zeta) = 0]. \tag{23}$$

The integral average on the left of (22) is, evidently, an average of

$$\frac{1}{2\pi}\int_{|\zeta|=1} h_\epsilon(u(r\zeta))\,|d\zeta| \tag{24}$$

over a certain range of r corresponding to the range from 0 to l for the NE-radius. In this average large values of l, i.e., values of $r\,l$ near one, are of dominating weight. Since, according to (16), (24) tends to the right-hand side of (22) as $r \to 1$ it follows that (22) must also be true. Finally, (23) follows from the obvious inequalities

$$\operatorname*{meas}_{|\zeta|=1}[u(\zeta) = 0] \leqq \int_{|\zeta|=1} h_\epsilon(u(\zeta))\,|d\zeta| \leqq \operatorname*{meas}_{|\zeta|=1}[u(\zeta) < \epsilon].$$

For later purposes, we need the analogous relations resulting from (16′). On setting

$$M_\epsilon(l) = \frac{1}{\sigma^2(K_l)}\iint_{K_l}\iint_{K_l} h_\epsilon(U(z, w))d\sigma_z d\sigma_w, \tag{25}$$

we obtain quite similarly

$$\lim_{\epsilon\to 0}\lim_{l\to\infty} M_\epsilon(l) = (4\pi^2)^{-1} \operatorname*{meas}_{|\zeta|=|\gamma|=1} [U(\zeta, \gamma)] = 0. \tag{26}$$

Returning to the proof of Lemma 2 we note that the integral average on the left in (22) is less than

$$\frac{\sigma(B_\epsilon K_l)}{\sigma(K_l)}, \ l \text{ sufficiently large}, \tag{27}$$

where B_ϵ is the set of all z in $|z|<1$ satisfying $u(z)<\epsilon$. The invariance under $\mathcal{G}$ of the function $u(z)$ implies that of the point set B_ϵ. By Lemma 1, (27) and therefore the left-hand average in (22) does not exceed the value

$$a\sigma(B_\epsilon R). \tag{28}$$

From the hypothesis of the present lemma, viz. that the right-hand side of (23) is positive, it then follows that (28) remains, for all $\epsilon>0$, above a positive constant. In particular, the set common to all B_ϵ is not empty. There exists therefore a point in $|z|\leqq 1$ where $u=0$, i.e., where $u\geqq 0$ attains its minimum, which completes the proof of the lemma.

5. **Proof of Lemma** 3. We first confine ourselves to the simpler case where the boundary of the fundamental region R of $\mathcal{G}$ lies entirely in $|z|<1$. Of all the images $S(R)$ we call R_0 the particular one that contains the origin in its interior or on its boundary,

$$z = 0 \subset R_0. \tag{29}$$

R_0 and all its images have the same finite NE-diameter D. We enumerate all these congruent parts of $|z|<1$ in an arbitrary way, $R_0, R_1, R_2, \cdots$, and we call S_ν the substitution of $\mathcal{G}$ that transforms R_ν into R_0,

$$S_\nu(R_\nu) = R_0. \tag{30}$$

We now consider all R_ν lying entirely in the closed circular disc K_l. They evidently cover the whole of K_{l-D}. From (25) we then obtain

$$M_\epsilon(l-D) \leqq q(l)\frac{1}{\sigma^2(K_l)}\sum_{\nu,\mu}\iint_{R_\nu}\iint_{R_\mu} h_\epsilon(U(z, w))d\sigma_z d\sigma_w, \tag{31}$$

where

$$q(l) = [\sigma(K_l)/\sigma(K_{l-D})]^{-2} \tag{31′}$$

and where integration and summation are carried out so as to satisfy the conditions

$$z \subset R_\nu \subset K_l, \qquad w \subset R_\mu \subset K_l \tag{31''}$$

in all possible ways. By (29), (30), and (31''),

$$s(0, S_\nu(z)) \leqq D,$$

whence, by Harnack's inequality,

$$U(z, w) = U(S_\nu(z), S_\nu(w)) \geqq e^{-D} U(0, S_\nu(w)),$$

and, as $h_\epsilon(t)$ nowhere increases,

$$h_\epsilon(U(z, w)) \leqq h_\epsilon\{e^{-D} U(0, S_\nu(w))\}, \quad z \subset R_\nu. \tag{32}$$

Now, the function of w

$$h_\epsilon\{e^{-D} U(0, w)\} \tag{33}$$

is a concave function of a harmonic function and, therefore, *subharmonic* in $|w| < 1$. This is most easily proved by verifying the non-negativeness of the Laplacian.* Hence, (33) nowhere exceeds the function $V(w) \equiv V_\epsilon(w)$ which is harmonic in $|w| < 1$ and which possesses, on $|w| = 1$, the same boundary values. The slight difficulty brought about by the fact that these boundary values are merely measurable is readily surmounted by considering first smaller circles and by proceeding then to the limt as $r \to 1$. We note that

$$V(0) = \frac{1}{2\pi} \int_{|\gamma|=1} h_\epsilon\{e^{-D} U(0, \gamma)\} \, |d\gamma|, \tag{34}$$

and that, by (32),

$$h_\epsilon(U(z, w)) \leqq V(S_\nu(w)). \tag{35}$$

From (31), (31''), and (35) we obtain, with regard to $V_\epsilon \geqq 0$,

$$M_\epsilon(l - D) \leqq q(l) \frac{1}{\sigma^2(K_l)} \sum_{\nu,\mu} \int\int_{R_\nu} \int\int_{R_\mu} V(S_\nu(w)) d\sigma_z d\sigma_w$$

$$\leqq g(l) \frac{1}{\sigma(K_l)} \sum_\nu \int\int_{R_\nu} \left\{ \frac{1}{\sigma(K_l)} \int\int_{K_l} V(S_\nu(w)) d\sigma_w \right\} d\sigma_z.$$

Since $V(S_\nu(w))$ is harmonic in $|w| < 1$, we have by Gauss's mean-value theorem,

$$\frac{1}{\sigma(K_l)} \int\int_{K_l} V(S_\nu(w)) d\sigma_w = V(S_\nu(0)),$$

* This is where the existence and continuity of h_ϵ'' is used.

whence

$$(36)\qquad \begin{aligned} M_\epsilon(l-D) &\leqq q(l)\frac{1}{\sigma(K_l)}\sum_\nu \sigma(R_\nu)V(S_\nu(0)) \\ &= \frac{q(l)}{\sigma(K_l)}\sigma(R_0)\sum_\nu V(S_\nu(0)), \end{aligned}$$

the summation being extended over all ν for which $R_\nu \subset K_l$. On setting

$$(37)\qquad S_\nu(R_0) = R'_\nu,$$

we infer from (29) that

$$(38)\qquad S_\nu(0) \subset R'_\nu.$$

For two different R_ν the corresponding R'_ν are obviously different. Furthermore, it follows from (30) and (37) that the NE-distance of R'_ν from R_0 is the same as that of R_ν from R_0. According to (31″) the NE-distance of R_ν from R_0 is less than l. Hence all regions R'_ν considered here must lie within the circle K_{l+2D}. Thus (36) can be written as

$$(38)\qquad M_\epsilon(l-D) \leqq \frac{q(l)}{\sigma(K_l)}\sum_\nu \int\int_{R'_\nu} V(S_\nu(0))d\sigma_z,$$

where

$$(38')\qquad S_\nu(0) \subset R'_\nu \subset K_{l+2D}.$$

By (38′) on applying Harnack's inequality to $V \geqq 0$ we have

$$V(S_\nu(0)) \leqq e^D V(z), \qquad z \subset R'_\nu,$$

which being combined with (38) yields

$$\begin{aligned} M_\epsilon(l-D) &\leqq \frac{q(l)}{\sigma(K_l)} e^D \sum_\nu \int\int_{R'_\nu} V(z)d\sigma_z \\ &\leqq \frac{e^D q(l)}{\sigma(K_l)} \int\int_{K_{l+2D}} V(z)d\sigma_z = q(l)e^D \frac{\sigma(K_{l+2D})}{\sigma(K_l)} V(0), \end{aligned}$$

and, according to (17),

$$\lim_{l\to\infty} M_\epsilon(l) \leqq e^{5D}V(0).$$

On taking account of (26) and (34) we infer from this inequality that

$$(39)\qquad (4\pi^2)^{-1} \operatorname*{meas}_{|\zeta|=|\gamma|=1} [U(\zeta,\gamma)=0] \leqq (2\pi)^{-1}e^{5D} \operatorname*{meas}_{|\gamma|=1} [U(0,\gamma)=0],$$

which proves Lemma 3 in the case where R has no vertices on $|z|=1$.

In the general case by cutting off the vertices lying on $|z|=1$ we can always divide R_0 into two parts,

$$R_0 = R_0^* + R_0^{**}, \tag{40}$$

such that

$$\sigma(R_0^{**}) < \delta \tag{41}$$

and that R_0^* has a finite NE-diameter, say $D=D(\delta)$. We may always suppose R_0^* to contain the origin. The set

$$B = \sum_S S(R_0^{**})$$

is invariant under $\mathcal{G}$. The set R_0^* and all sets R_ν^* congruent to it cover the complement of B in $|z|<1$. We note that all R_ν^* have the same NE-diameter D. Consider all

$$R_\nu^* \subset K_l.$$

Every point of K_{l-D} belongs either to one of these R_ν^* or to B. For, if it belongs to any R_ν^* at all, this set must be contained in K_l since its diameter equals D. Hence

$$K_{l-D} \subset \sum R_\nu^* + BK_{l-D}, \qquad R_\nu^* \subset K_l.$$

Since $h_\epsilon \leqq 1$ we obtain from (25)

$$M_\epsilon(l-D) \leqq \left[\frac{\sigma(BK_{l-D})}{\sigma(K_{l-D})}\right]^2 + \frac{q(l)}{\sigma^2(K_l)} \sum_{\nu,\mu} \iint_{R_\nu^*} \iint_{R_\mu^*},$$

the summation being confined to all

$$R_\nu^* \subset K_l, \qquad R_\mu^* \subset K_l.$$

Here the second term satisfies the same inequalities as before. The first term is less than

$$a^2\sigma^2(BR_0) = a^2\sigma^2(R_0^{**}) < a^2\delta^2,$$

by Lemma 1 and (41). On proceeding as before to the limit as $l \to \infty$, $\epsilon \to 0$, we obtain inequality (39) with the additional term $a^2\delta^2$ in the right-hand side. A suitable choice of δ evidently leads to the general proof of Lemma 3. Theorem B, and therefore the proposed theorem on surfaces Σ of the first kind, is herewith completely proved.

6. **Surfaces of the second kind.** For a group $\mathcal{G}$ of the second kind, the fundamental region has on its boundary one or several arcs of the unit circle. We shall consider only the case where R has no zero angle vertices on $|z|=1$,

i.e., where the surface Σ has no cusps. These arcs and their images under $\mathcal{G}$ are known to lie everywhere dense on $|z| = 1$. We shall prove that this set ω has the measure of the whole of $|z| = 1$. Let α be one of the complete arcs of $|z| = 1$ belonging to the boundary of R. There will be two images of R adjacent to R along the two sides of α which end at the two end points of α, respectively. In particular, there are two arcs of the set ω immediately adjacent on both sides of α. This shows that the end points of any arc α are interior points of the set ω introduced above. Since ω is invariant under $\mathcal{G}$, the Poisson integral $u(z)$ whose boundary values on $|z| = 1$ are zero on ω and one elsewhere must also be invariant under $\mathcal{G}$,

$$u(S(z)) = u(z)$$

for all S of $\mathcal{G}$. All we have to prove is that $u \equiv 0$. Indeed, $u(z)$ has, in the sense of ordinary convergence, the boundary value zero on every closed arc α. On account of its invariance, u takes all its values in R. Since a harmonic function always attains its extrema on the boundary, $u(z)$ must attain them on the (closed) part of $|z| = 1$ belonging to the boundary of R, whence $u \equiv 0$.

If R has vertices on $|z| = 1$ an elementary proof could still be given as well as for Lemma 2, for instance by applying Green's formula

$$\iint \operatorname{grad}^2 u \, dxdy = \int u \frac{\partial u}{\partial n} ds$$

to a region obtained by diminishing R suitably (cutting off the vertices on $|z| = 1$ in a suitable way).

Massachusetts Institute of Technology,
Cambridge, Mass.