ANALYTICAL GROUPS*

BY

GARRETT BIRKHOFF

Introduction

1. **Abstract groups.** The present paper will deal with *abstract* continuous groups. This means that it will discuss symbols which behave like transformations, without specifying the domain on which the transformations operate. The reader will be assumed to be conversant with abstract groups as algebraic entities.

2. **Questions in the large.** It is well-known that the theory of continuous groups in the large differs essentially from the theory in the small. Some things, such as the one-one correspondence between closed subgroups of a Lie group and subalgebras of the Lie algebra of its infinitesimal generators, are true only locally;† others, such as the introduction by Weyl and Haar of invariant mass, are possible only when one deals with groups in the large.

The present paper is a theory *in the small* exclusively; it neither involves implicitly nor resolves explicitly the difficulties in the large. In this it resembles the original theory of Lie.

3. **Actual contents.** Thus the paper avoids two large classes of questions. What questions does it answer—what are its assumptions, and how can one summarize its conclusions?

The paper deals with systems (called "analytical groups") in which an associative multiplication is defined, and which can be so mapped on a Banach parameter-space that if one multiplies all elements by any fixed element near the origin, vector differences are left nearly invariant.‡

It is proved that if G is any analytical group (more properly, analytical group nucleus), then

(1) G is a topological group nucleus in the usual sense.

(2) One can introduce canonical parameters into G.

(3) G has an infinitesimal (Lie) algebra $L(G)$.

(4) The analytical subgroups of G correspond biuniquely to the closed

* Presented to the Society, December 31, 1936; received by the editors December 17, 1936 and, in revised form, May 3, 1937.

† Again, the group of topological automorphisms of the group of the torus differs radically from that of the group of translations of the plane, in spite of the fact that these two groups are locally isomorphic.

‡ A Banach space is of course simply a system having certain prescribed elementary properties of euclidean space which are shared by various important function spaces. Cf. §8.

subalgebras of $L(G)$, a subgroup being normal if and only if the corresponding subalgebra is invariant.

(5) If G is under canonical parameters, then there exists a formal series of polynomials determined by $L(G)$ which expresses the rule for forming group products.

(6) One can define product integrals for functions with values in $L(G)$, which include the Lebesgue integral (the case G is the additive group of real numbers), and all known product integrals (the case G is a group of matrices).

(7) Quite general functions $x(\lambda)$ with real arguments and values in a Banach space B determine formal series in elements of B and their brackets, which express the product integral $\hat{\int} x(\lambda)d\lambda$ under canonical parameters for *any* analytical group G whose parameter-space contains B and whose "commutation modulus" is sufficiently small.

(8) All the operations defined (e.g., vector addition under canonical parameters, product integration) are *topologico-algebraic*—preserved under topological isomorphisms.

4. **Extension to infinite dimensions.** Perhaps the main advantage of the above assumptions, is the fact that many *infinite* continuous groups satisfy them. This marks a real advance in the analytical theory of groups.*

The infinite-dimensional analytical groups treated in the literature are of two kinds: the infinite continuous groups of analytical transformations

$$x_i' = f_i(x_1, \cdots, x_n) \qquad (i = 1, \cdots, n)$$

of n-dimensional space discussed by Lie [10]† and Cartan [4], and the groups of linear operators on Hilbert space recently studied by Delsarte [6]. Each of these authors omits to define the meaning of the convergence $T_n \rightarrow T$ of a sequence of transformations T_n to a limit T—in other words, to define the topological structure of the corresponding abstract groups.

This omission, and the omission to establish a rigorous correlation between the actual transformations of such groups and so-called "infinitesimal" transformations, are not trivial. In fact, although the present paper supplies a complete theory for a class of groups including those studied by Delsarte, the author does not even know what the facts are in the case of the groups studied by Lie and Cartan. Part of the difficulty is that the group manifolds are not metric-linear; part of it is that canonical parameters do not define even locally a one-one representation of the group manifold.

---

* Cf. Abstract 41-3-129 of the Bulletin of the American Mathematical Society (1935); also *Continuous groups and linear spaces*, Matematicheskiĭ Sbornik, vol. 1 (1936), pp. 635–642; an address delivered at the First International Topological Conference, Moscow, September 5, 1935.

† Numbers in brackets refer to the bibliography at the end of the paper.

5. **Continuous groups: topological and analytical.** This illustrates the importance of geometrical properties of group manifolds; we shall now see how continuous groups can be classified on a purely geometrical basis.

A "continuous" abstract algebra (whether group, ring, or field) is a system whose elements are simultaneously the points of a geometrical manifold and the symbols of a formal calculus, and whose algebraic operations determine "smooth" functions of the manifold into itself. By letting the geometry of the manifold suggest the proper definition of smoothness, one is led to a purely geometrical classification of continuous abstract algebras.

Thus with groups whose manifolds are general topological spaces, one naturally regards "smoothness" of the group operations as meaning that group multiplication and passage to the inverse are continuous in the topology of the group manifold. Such groups are called *topological.*

Similarly, with groups whose manifolds are n-dimensional analytical varieties, it is natural to assume that the group operations are analytical in the coordinates; this leads to the usual concept of a *Lie* group.

Now it is a remarkable fact, that two analytical systems which are continuous images of each other, are in general analytical images of each other. This seems to hold even in pure geometry: thus dimensionality, originally known to be invariant only under analytical transformations, is now realized to be a topological invariant. We shall extend the domain of validity of this principle below, by proving that *continuous* isomorphisms between Lie groups are necessarily *analytical.**

6. **Groups as topological algebras.** The result just stated, combined with (8), suggests that one can develop a theory for analytical groups in which group multiplication and passage to the limit are the only notions introduced as undefined primitives.†

Indeed, this program is technically feasible: it is shown below that one *can* give topologico-algebraic definitions of analytical groups. But as the argument is really *metric*, it would be misleading to make it pseudo-topological—even though it is less analytical and more topologico-algebraic‡ than any

* Discontinuous (and hence non-analytical) isomorphisms exist; there is one between the group of translations of the line and the group of translations of the plane. To see this, form in each an independent basis with respect to linear combination with rational coefficients. (However, van der Waerden, Mathematische Zeitschrift, vol. 36 (1933), pp. 780–787, has shown that isomorphisms between compact semi-simple Lie groups are always analytical.) Conceivably two Lie groups which are isomorphic and have homeomorphic manifolds are eo ipso analytically isomorphic.

† Especially since O. Schreier [15] has obtained so much information about group manifolds by such a theory.

‡ Thus pure group algebra—especially that of commutation—is shown to yield many results (especially in Chaps. IV–V) which could not be obtained by general analytical methods.

previous reasoning yielding the same results.

All this relates to the well-known problem of determining the weakest analytical assumptions demonstrably equivalent to an assumption of unrestricted analyticity. The weakest assumption in the literature* (cf. [11]) is that the function of group multiplication has continuous *second* derivatives. It is shown below that if one assumes continuous *first* derivatives, then one can deduce the whole theory† of abstract Lie groups.

Chapter I. Technical machinery

7. **A remark on notation.** It will shorten the argument in the sequel to use the following notational conventions: $M(\lambda)$ for any positive function of a real variable λ such that $\lim_{\lambda\to 0}M(\lambda)=0$; $O(\lambda)$ for any such function satisfying $O(\lambda)\leqq K\cdot|\lambda|$ for some $K<+\infty$; $o(\lambda)$ for any such function satisfying $o(\lambda)\leqq|\lambda|\cdot M(\lambda)$ for some $M(\lambda)$. (The relation of the last two definitions to Landau's well-known o-O notation‡ is obvious.)

Thus let $\phi(x_1, \cdots, x_r)$ and $\psi(x_1, \cdots, x_r)$ be any two real-valued functions of the same (not necessarily numerical!) variables $x_1, \cdots, x_r$. By the preceding definition,

$$\phi(x_1, \cdots, x_r) \leqq M(\psi(x_1, \cdots, x_r))$$

means that given $\eta>0$, $\delta>0$ exists so small that $\psi(x_1, \cdots, x_r)<\delta$ implies $\phi(x_1, \cdots, x_r)<\eta$. The inequalities $\phi(x_1, \cdots, x_r)\leqq O(\psi(x_1, \cdots, x_r))$ and $\phi(x_1, \cdots, x_r)\leqq o(\psi(x_1, \cdots, x_r))$ have similar meanings.

It is obvious that in terms of this notation, the following substitutions are legitimate:

$$(7\alpha)\qquad O(\lambda) \quad \text{for} \quad o(\lambda), \quad \text{and} \quad M(\lambda) \quad \text{for} \quad O(\lambda).$$

$$(7\beta)\qquad M(\lambda) \quad \text{for} \quad M(O(\lambda)), \quad \text{and} \quad O(\lambda) \quad \text{for} \quad O(O(\lambda)).$$

$$(7\gamma)\qquad M(\lambda+\mu) \quad \text{for} \quad M(\lambda)+M(\mu).$$

$$(7\delta)\qquad M(\lambda) \quad \text{for} \quad M(\lambda)/M[1-M(\lambda)].$$

Thus if $\phi(x_1, \cdots, x_r)\leqq M(\psi_1(x_1, \cdots, x_r))+M(\psi_2(x_1, \cdots, x_r))$, then by (7γ), $\phi(x_1, \cdots, x_r)\leqq M(\psi_1(x_1, \cdots, x_r)+\psi_2(x_1, \cdots, x_r))$.

It goes without saying that the M-functions, o-functions and O-functions appearing in the text vary from group to group, and from inequality to inequality—although since only a finite number of such functions are used in

* Except when dealing with compact (von Neumann) and abelian (Pontrjagin) groups, where one need only assume that one has a topological group locally homeomorphic with euclidean space.

† The author announced this result in Abstract 41-5-192 (1935) of the Bulletin of the American Mathematical Society.

‡ Cf. G. H. Hardy, *Pure Mathematics*, 5th edition, Cambridge University Press, 1928, p. 448.

dealing with any one group, there exists a single $M(\lambda)$, $o(\lambda)$, and $O(\lambda)$ which works in all inequalities for that group.

8. **Formal definition of "analytical group."** The properties of "analytical groups" which will be assumed were indicated in §2; they can be stated explicitly as

DEFINITION 1. By an *analytical group* will be meant any region about the origin of a Banach space, in which an associative multiplication is defined for elements near the origin Θ, satisfying

$$(1) \qquad x \circ \Theta = \Theta \circ x = x \quad \text{for all} \quad x.$$

$$(2') \qquad |(xa - xb) - (a - b)| \leqq M(|x| + |b| + |a|) \cdot |a - b|.$$

$$(2'') \qquad |(ay - by) - (a - b)| \leqq M(|a| + |b| + |y|) \cdot |a - b|.$$

In words, the origin is the group-identity e, and vector differences are nearly invariant under group translations $T_y^x : a \to xay$. (By xy or $x \circ y$ is meant the group product of x and y.)

(By a *Banach space* is meant a B-space in the sense of Banach [1]—that is, a linear space in which an absolute value $|x|$ is defined which (1) is positive for $x \neq \Theta$, (2) satisfies the triangle inequality $|x+y| \leqq |x| + |y|$, (3) is multiplied identically by $|\lambda|$ under any scalar expansion $x \to \lambda x$, and (4) makes the space complete*—such that if $\lim_{m,n\to\infty}|x_m - x_n| = 0$, then x exists such that $\lim_{n\to\infty}|x - x_n| = 0$.)

9. **A topological group nucleus.** Combining (2′)–(2″), we get immediately,

$$(2) \qquad |(xay - xby) - (a - b)| = M(|x| + |a| + |b| + |y|) \cdot |a - b|.$$

Again, setting $b = e = \Theta$ in (2′) and $a = e = \Theta$ in (2″), one obtains,

$$(3') \qquad |x \circ y - (x + y)| \leqq M(|x| + |y|) \cdot |y|,$$

$$(3'') \qquad |x \circ y - (x + y)| \leqq M(|x| + |y|) \cdot |x|,$$

which can be combined into the single inequality

$$(3) \qquad |x \circ a \circ y - (x + a + y)| = M(|x| + |a| + |y|) \cdot (|x| + |y|).$$

In words, near the origin group translations T_y^x differ little from the corresponding linear translations $L_y^x : a \to a + x + y$.

(9α) *Multiplication is continuous near* Θ.

Proof. If $|x|$, $|y|$, $|a|$ and $|b|$ are small, then

$$\begin{aligned} |(x \circ y) - (a \circ b)| &= |\{(x \circ y) - (a \circ y)\} + \{(a \circ y) - (a \circ b)\}| \\ &= O(|x - a| + |y - b|). \end{aligned}$$

* Incidentally, $\{x_n\}$ is metrically "fundamental" if and only if it is "fundamental" in the topologico-algebraic sense (of van Dantzig) that $\lim_{m,n\to\infty} x_m^{-1} \circ x_n = 0$. Cf. (9δ).

(9β) *Every sufficiently small element* x *has a unique inverse* x^{-1} *satisfying* $x\circ x^{-1}=x^{-1}\circ x=\Theta$, *and* $|x^{-1}|\leqq 2\cdot|x|$.

Proof. Suppose $M(5|x|)<\frac{1}{2}$. Define $y_0=\Theta$ and by induction $y_{n+1}=y_n-(xy_n)$. Then

$$\begin{aligned}|xy_{n+1}| &= |(xy_{n+1}-xy_n)-(y_{n+1}-y_n)|, && \text{by definition},\\ &\leqq M(|x|+|y_n|+|y_{n+1}|)\cdot|xy_n|, && \text{by } (2'),\end{aligned}$$

since $|y_{n+1}-y_n|=|-xy_n|=|xy_n|$. It follows by induction that $|xy_{n+1}|\leqq\frac{1}{2}^{n+1}\cdot|x|$, and $|y_{n+2}|\leqq(2-\frac{1}{2}^{n+1})\cdot|x|$—whence $M(|x|+|y_{n+1}|+|y_{n+2}|)<\frac{1}{2}$. Hence $\lim_{m,n\to\infty}|y_m-y_n|=0$, and so by completeness a y exists satisfying $|y|\leqq 2\cdot|x|$ and $\lim_{n\to\infty}|y-y_n|=0$. Now by (9$\alpha$), $xy=\Theta$, and so y is a right-inverse of x. Similarly x has a left-inverse z with $zx=\Theta$. Moreover $y=(zx)y=z(xy)=z$; hence $y=z\equiv x^{-1}$ is a full inverse of x; its uniqueness follows since $xx'=\Theta$ implies $x'=(x^{-1}x)x'=x^{-1}(xx')=x^{-1}$, while $x''x=\Theta$ implies $x''=x''(xx^{-1})=(x''x)x^{-1}=x^{-1}$.

(9γ) *Passage to the inverse is continuous near* Θ.

Proof. Let $(x+u)$ be given. Substitute x^{-1} for y_0 and $x+u$ for x in the proof of (9β). By (9α), $(x+u)y\leqq 2\cdot|u|$ in a small enough neighborhood; hence in the construction of $(x+u)^{-1}$ by successive approximation,

$$|(x+u)^{-1}-x^{-1}|\leqq|y-y_0|=4\cdot|u|.$$

We can summarize (9α)–(9γ) in

THEOREM 1. *Every analytical group contains a topological group nucleus in the usual sense.**

A topological space in which an associative multiplication is defined satisfying (9α)–(9γ) everywhere is called a topological group (cf. [15]).

(9δ) $|x^{-1}|\leqq|x|+o(|x|)$; *in fact*, $|x^{-1}+x|=o(|x|)$.

Proof. By (3′), $|x+x^{-1}|\leqq M(|x|+|x^{-1}|)\cdot|x|$; but by (9$\beta$), $M(|x|+|x^{-1}|)=M(|x|)$. Hence $x^{-1}=-x+u$, where $|u|=M(|x|)\cdot|x|=o(|x|)$, proving the result.

Digression on axiomatics: Setting $y=\Theta$ resp. $x=\Theta$ in (3′) resp. (3″), we obtain (1). Further, near Θ, (3′), (3″) make $by=a$ imply that $|y|$ is nearly $|b-a|$. Hence if we are dealing with a topological group nucleus, then (2′) and (2″) hold. (Proof: By symmetry, it suffices to prove (2′). But by (3′), writing $b^{-1}a=y$, whence $a=by$,

* Cf. B. L. van der Waerden, *Vorlesungen über kontinuierlichen Gruppen*, Göttingen, 1932. For the analogous notion of a Lie group nucleus (alias "germ," cf. [11]).

$$|(a-b)-y| = |by-b-y| \leqq M(|b|+|y|)\cdot|y|$$
$$|(xa-xb)-y| = |xby-xb-y| = M(|x|+|b|+|y|)\cdot|y|.$$

And so by the triangle law, since by continuity $M(|y|)=M(|a^{-1}b|) \leqq M(|a|+|b|)$ (cf. §7),

$$(2') \qquad |(xa-xb)-(a-b)| \leqq M(|x|+|b|+|a|)\cdot|a-b|\ .)$$

10. **Groups in the large.** Let H be any full topological group. Obviously any one-one bicontinuous map of a neighborhood of the identity of H onto a region of a Banach space which satisfies (1), (2′), (2″)—or alternatively, by the last paragraph, (3′), (3″)—defines that neighborhood as an analytical group. We are unable to prove* that any system satisfying Definition 1 is conversely a piece containing the identity of a full topological group.

Full topological groups which can be mapped locally onto Banach space in such a way as to satisfy Definition 1 will be termed *full* analytical groups; this will distinguish full groups from the analytical group nuclei with which we shall be concerned below and which, for brevity, we have called simply "analytical groups."

11. **Changes of parameters.** It is important to know which transformations of Banach spaces play the role of analytical coordinate transformations in the theory of abstract Lie groups—that is, which when applied to a given analytical group G attached to a parameter space, turn G into another topologically isomorphic analytical group H.

One can specify at once two classes of such transformations associated with an arbitrary Banach space B, namely:

(11a) The group of "distortions" of B—that is, of those transformations

---

* This has been proved for finite continuous groups by E. Cartan ([5], p. 18). Cartan omits to mention the decisive fact that if L is any Lie algebra, and N is its largest invariant "integrable" subalgebra, then L contains a semi-simple subalgebra S such that $S \cap N=0$ and $S+N=L$ (cf. J. H. C. Whitehead, Proceedings of the Cambridge Philosophical Society, vol. 32 (1936), pp. 229–238). This omission led Mayer-Thomas to question ([11], p. 806) the validity of Cartan's proof. Cartan has since published another equally technical proof (*Sur la Topologie des Groupes de Lie*, Paris, 1936, p. 22).

Neither of these proofs can be extended to the infinite continuous group nuclei treated below; each depends on lemmas which need not be true in infinite dimensions. For instance the fact that not all closed linear subspaces S of Banach spaces B have complements T such that $S \cap T=0$ and $S+T=B$ prevents one from using Cartan's special proof for solvable groups.

On the other hand Mayer-Thomas' argument (due independently to Paul Smith) for the case of group nuclei which can be embedded in a full group generalizes to infinite continuous groups—one takes the subgroup of the full group generated by the nucleus given, and retopologizes this subgroup by redefining distance as geodesic distance along paths in the subgroup.

Esthetically, one would expect to find a simple proof that every analytical group nucleus can be embedded in a full group, since it is easy to define the full group, if one knows that it exists.

T of B into itself which leave Θ fixed and satisfy (*) $|T(a+x)-T(a)-x| \leqq M(|a|+|x|)\cdot|x|$.

(11b) The class of alterations of the norm function $|x|$ of B to a new norm function $\|x\|$ such that the ratios $|x|/\|x\|$ and $\|x\|/|x|$ are uniformly bounded.

Remark. The latter correspond one-one to choices of bounded open convex regions $\|x\|<1$ of B. (Cf. A. Kolmogoroff, op. cit. in §17.)

THEOREM 2. *Any succession of transformations of types* (11a), (11b) *of the parameter-space of an analytical group G, turns G into a topologically isomorphic analytical group.*

Proof. It is obviously sufficient to prove the theorem for single transformations of types (11a) and (11b); again, the main difficulty is to prove analyticity. With (11b), one needs merely write $|x|/R \leqq \|x\| \leqq R\cdot|x|$, and to replace $M(\lambda)$ in (3) by $R^2M(\lambda/R) \leqq M(\lambda)$. (Cf. §7.)

Consider case (11a). Setting $a+x=b$ in (*), we see that (**) $|T(b)-T(a)-(b-a)| \leqq M(|a|+|b|)\cdot|b-a|$—i.e., vector differences, and hence absolute values near Θ are nearly invariant under T. Hence—the proof as in (9β) is by successive approximation—T is one-one and so by (**) bicontinuous near the origin. Therefore it suffices to prove (3′), (3″)—or even, by symmetry, (3′). This we shall do. Note that T^{-1} is of type (11a), and leaves vector differences near Θ almost invariant. Hence

$$|T^{-1}(a+x)-(T^{-1}(a)+T^{-1}(x))| \leqq M(|a|+|x|)\cdot|x|, \text{ by } (*),$$
$$|T^{-1}(a)\circ T^{-1}(x)-(T^{-1}(a)+T^{-1}(x))| \leqq M(|a|+|x|)\cdot|x|, \text{ by } (3').$$

Hence by the triangle inequality,

$$|T^{-1}(a)\circ T^{-1}(x)-T^{-1}(a+x)| \leqq M(|a|+|x|)\cdot|x|$$

and so, by (**),

$$|T(T^{-1}(a)\circ T^{-1}(x))-(a+x)| \leqq M(|a|+|x|)\cdot|x|.$$

But this *is* (3′) in terms of the new parameters, q.e.d.

We shall regard topologically isomorphic groups as essentially identical—as differing merely in their parametric representation.

12. **Rectifiable paths.** Let us recall a few familiar geometrical notions, so as to have a consistent notation and terminology for subsequent use. These notions are proper to abstract metric spaces,† and so apply to Banach spaces.

By a *path* is meant a continuous image P: $p(\lambda)$ of a line interval‡ $[0, \Lambda]$.

† The ideas go back to Fréchet's Thesis; cf., also, K. Menger, *Zur Metrik der Kurven*, Mathematische Annalen, vol. 103 (1930), p. 471, §5.

‡ As is conventional, $[0, \Lambda]$ denotes the set of real numbers λ which satisfy $0 \leqq \lambda \leqq \Lambda$

Two paths P and Q are called "geometrically equivalent" (written $P\approx Q$) if and only if they can be identified by proper choice of parameters—i.e., if and only if one can establish a one-one sense-preserving correspondence between the intervals of which they are images, such that corresponding points have the same image. Clearly the relation of being geometrically equivalent is reflexive, symmetric and transitive.

Again, by a *segment* ΔP of P is meant the image of any subinterval Δ: $[\lambda_1, \lambda_2]$ of $[0, \Lambda]$. By a *partition* π of P is meant a division of $[0, \Lambda]$ into successive subintervals Δ_k: $[\lambda_{k-1}, \lambda_k]$, where $\lambda_0=0$, $\lambda_n=\Lambda$, and $k=1, \cdots, n$. By the "product" of any two partitions π and π' of P is meant the partition $\pi\cdot\pi'$ whose subintervals are the intersections of the subintervals of π with those of π'. And π is called a "subpartition" of π' (in symbols, $\pi\leqq\pi'$) if and only if $\pi=\pi\cdot\pi'$.

By the π-approximate length of P under any partition π is meant $|P|_\pi\equiv\sum_{k=1}^{n}|P(\lambda_k)-P(\lambda_{k-1})|$, and by the "length" of P is meant $|P| =\sup |P|_\pi$. A path P is called "rectifiable" if and only if $|P|<+\infty$. Obviously

(12α) *If* $P\approx Q$, *then* $|P|=|Q|$.

The "diameter" $|\pi|$ of a partition π of a rectifiable path P is defined as $\sup|\Delta P|$. It is not hard to show

(12β) $|P| =\lim_{|\pi|\to 0}|P|_\pi$.

And since $|\pi|\leqq|\pi'|$ provided $\pi\leqq\pi'$, we see

(12γ) $|P| =\lim_{\pi\downarrow}|P|_\pi$ *in the sense of Moore-Smith.*†

13. **More notation.** We shall now introduce some special but natural notation for handling *rectifiable paths issuing from the origin* (=identity) *of an analytical group nucleus.*

If P_k is any path with domain $[0, \Lambda_k]$, then $t(P_k)$ denotes $p_k(\Lambda_k)-p_k(0)$.

By the *path-sum* of r admissible paths $P_1, \cdots, P_r$, will be meant the path $P=P_1\oplus\cdots\oplus P_r$ formed by adding to $P_1\oplus\cdots\oplus P_{r-1}$ a segment congruent to P_r under linear translation through $t(P_1\oplus\cdots\oplus P_{r-1})$. And by the *path-product* of the P_k, will be meant the path $\tilde P=P_1\circ\cdots\circ P_r$ formed by adding to $P_1\circ\cdots\circ P_{r-1}$ a segment congruent to P_r under group left-translation through $t(P_1\circ\cdots\circ P_{r-1})$. Thus P and $\tilde P$ have $[0, \Lambda_1+\cdots+\Lambda_r]$ for domain, and for $0\leqq\lambda\leqq\Lambda_{k+1}$,

$$(13.1)\qquad \begin{cases} p(\Lambda_1+\cdots+\Lambda_k+\lambda) = t(P_1)+\cdots+t(P_k)+p_{k+1}(\lambda), \\ \tilde p(\Lambda_1+\cdots+\Lambda_k+\lambda) = t(P_1)\circ\cdots\circ t(P_k)\circ p_{k+1}(\lambda). \end{cases}$$

† This means that, given any neighborhood of $|P|$, one can find a π_0 such that $\pi\leqq\pi_0$ implies that $|P|_\pi$ lies in that neighborhood of $|P|$.

Since linear and group translations leave distances invariant resp. nearly invariant, P and $\tilde{P}$ are admissible.

We shall now develop an abstract correspondence between paths which generalizes the correspondence between the sum-integral $\int x(\lambda)d\lambda$ and the product integral $\hat{\int} x(\lambda)d\lambda$ of functions whose values $x(\lambda)$ are matrices. (N.B., $t(\Delta_k P)$ is the analogue of $x(\lambda)\Delta\lambda$.)

Accordingly, let P be any admissible path, and let π be any partition of P into segments $\Delta_1 P, \cdots, \Delta_r P$. Denote by P_k the image of $\Delta_k P$ after linear translation through $-t(P_1 \oplus \cdots \oplus P_{k-1})$, and by Q_k the image of $\Delta_k P$ after left-multiplication by the group-inverse of $t(Q_1 \circ \cdots \circ Q_{k-1})$. Then by construction

$$P = P_1 \oplus \cdots \oplus P_r = Q_1 \circ \cdots \circ Q_r.$$

We shall define the two dualistic paths

$$P_\pi^* = P_1 \circ \cdots \circ P_r \qquad \text{and} \qquad P_\pi\dagger = Q_1 \oplus \cdots \oplus Q_r$$

formed by interchanging the operations of path-addition and path-multiplication. Then we shall prove that the P_π^* and $P_\pi\dagger$ approach fixed limiting positions P^* and $P\dagger$ as $|\pi|$ tends to zero.

14. **Evaluation of paths.** Of course, the meaning of this statement depends on how one defines limits—on how one *topologizes* the "space" of images of a fixed interval.

Let P and Q be any two images of the same interval $[0, \Lambda]$. We shall make the definition

$$|P - Q| = \sup_{0 \leqq \lambda \leqq \Lambda} p(\lambda) - q(\lambda).$$

It is clear that this definition of distance makes the images of $[0, \Lambda]$ the "points" of an abstract metric space, in the sense defined earlier;‡ this depends only on the fact that the images of $[0, \Lambda]$ are themselves in a metric space.

We now come to some statements involving group properties. In stating and proving them we shall write $\prod_{k=1}^r x_k$ for $(x_1 \circ \cdots \circ x_r)$, and $\sum_{k=1}^r x_k$ for $(x_1 + \cdots + x_r)$.

(14α) $|\prod_{k=1}^r x_k - \sum_{k=1}^r x_k| = o(\sum_{k=1}^r |x_k|)$. *Consequently,* $|\prod_{k=1}^r x_k| \leqq O(\sum_{k=1}^r |x_k|)$.

Proof. By the triangle inequality,

$$\left| \prod_{k=1}^{r} x_k - \sum_{k=1}^{r} x_k \right| \leqq \left| \left[\left(\prod_{k=1}^{r-1} x_k \right) \circ x_r \right] - \left[\left(\prod_{k=1}^{r-1} x_k \right) + x_r \right] \right|$$

‡ Fréchet, op. cit., p. 36, introduces this very definition of distance, and shows that it is metric.

$$+\left|\prod_{k=1}^{r-1} x_k - \sum_{k=1}^{r-1} x_k\right|$$

$$\leqq |x_r| \cdot M\left(\left|\prod_{k=1}^{r-1} x_k\right|\right) + M\left(\sum_{k=1}^{r-1}|x_k|\right)\cdot\left(\sum_{k=1}^{r-1}|x_k|\right)$$

by (3) and induction on r. Recombining—since, by induction on r, $|\prod_{k=1}^{r-1}x_k| \leqq O(\sum_{k=1}^{r-1}|x_k|)$—we get

$$\left|\prod_{k=1}^{r} x_k - \sum_{k=1}^{r} x_k\right| \leqq M\left(\sum_{k=1}^{r-1}|x_k|\right)\left(\sum_{k=1}^{r-1}|x_k|\right) \leqq o\left(\sum_{k=1}^{r}|x_k|\right).$$

(14β) *We have the inequality*

$$\left|\prod_{k=1}^{r} x_k - \prod_{k=1}^{r} y_k - \sum_{k=1}^{r}(x_k - y_k)\right| \leqq M\left(\sum_{k=1}^{r}|x_k| + \sum_{k=1}^{r}|y_k|\right)\left(\sum_{k=1}^{r}|x_k - y_k|\right).$$

Hence if $M(\sum_{k=1}^{r}[|x_k| + |y_k|]) < 1$, *then*

$$\left|\prod_{k=1}^{r} x_k - \prod_{k=1}^{r} y_k\right| \leqq 2\left(\sum_{k=1}^{r}|x_k - y_k|\right).$$

Proof. By the triangle inequality iterated,

$$\left|\prod_{k=1}^{r} x_k - \prod_{k=1}^{r} y_k - \sum_{k=1}^{r}(x_k - y_k)\right| \leqq \sum_{k=1}^{r}\left|\left(\prod_{i=1}^{k-1} x_i\right)x_k\left(\prod_{i=k+1}^{r} y_i\right) - \left(\prod_{i=1}^{k-1} x_i\right)y_k\left(\prod_{i=k+1}^{r} y_i\right) - (x_k - y_k)\right|$$

$$\leqq \sum_{k=1}^{r} M\left(\sum_{k=1}^{r}|x_k| + \sum_{k=1}^{r}|y_k|\right)\cdot|x_k - y_k|, \qquad \text{by (2)},$$

since $|\prod_{i=1}^{k-1}x_i| \leqq O(\sum_{i=1}^{r}|x_i|)$ and $|\prod_{i=k+1}y_i| \leqq O(\sum_{i=1}^{r}|y_i|)$ by (14α).

(14γ). *Let* $P_1, \cdots, P_r$ *and* $Q_1, \cdots, Q_r$ *be admissible paths, each* P_k *having the same domain as* Q_k. *Further, let* $|P|$ *denote* $\sum_{k=1}^{r}|P_k|$ *and* $|Q|$ *denote* $\sum_{k=1}^{r}|Q_k|$. *Then*

$$|(P_1 \oplus \cdots \oplus P_r) - (Q_1 \oplus \cdots \oplus Q_r)| \leqq \sum_{k=1}^{r}|P_k - Q_k|.$$

And if $|P| + |Q|$ *is so small that* $M(|P| + |Q|) < 1$, *then*

$$|(P_1 \circ \cdots \circ P_r) - (Q_1 \circ \cdots \circ Q_r)| \leqq 2\sum_{k=1}^{r}|P_k - Q_k|.$$

Proof. The first inequality follows from (13.1) and the triangle inequality. The second follows from (13.1) and (14β).

Thus with paths of sufficiently small total length, both path-sums and path-products are uniformly continuous functions of their arguments in our metric "path-space."

LEMMA. *Let P be any sufficiently short path. Then if $\pi' \leqq \pi$, $|P_\pi^* - P_{\pi'}^*| \leqq M(|\pi|)$ and $|P_\pi\dagger - P_{\pi'}\dagger| \leqq M(|\pi|)$.*

Proof. It is an essential preliminary remark that each segment of any P_π^* or $P_\pi\dagger$ has nearly the same length as the corresponding segment of P, since it is obtained from it by group and linear translation of subsegments through relatively small distances—and such translations by (2) leave distances nearly invariant.

Now write $P_\pi^* = P_1 \circ \cdots \circ P_r$. Clearly $P_{\pi'}^*$ is obtainable from P_π^* by replacing each component $P_k = P_{k,1} \oplus \cdots \oplus P_{k,s}$ by the path $\tilde{P}_k = P_{k,1} \circ \cdots \circ P_{k,s}$. But referring to (14$\alpha$), we see that $|\tilde{P}_k - P_k| \leqq o(|P_k|) \leqq M(|\pi|)\cdot|P_k|$. Hence by (14$\gamma$),

$$|P_\pi^* - P_{\pi'}^* \leqq 2M(|\pi|)\cdot\sum_{k=1}^{r}|P_k| = M(|\pi|)\cdot|P|.$$

Similarly, write $P_\pi\dagger = Q_1 \oplus \cdots \oplus Q_r$. Clearly $P_{\pi'}\dagger$ is obtainable from $P_\pi\dagger$ by replacing each component $\tilde{Q}_k = Q_{k,1} \circ \cdots \circ Q_{k,s}$ by a path $\tilde{Q}_k = Q_{k,1} \oplus \cdots \oplus Q_{k,s}$. By (14$\alpha$) and the preliminary remark, $|\tilde{Q}_k - Q_k| \leqq o(|Q_k|) \leqq M(|\pi|)\cdot|Q_k|$. Hence by (14$\gamma$),

$$|P_\pi\dagger - P_{\pi'}\dagger| \leqq M(|\pi|)\cdot\sum_{k=1}^{r}|Q_k| = M(|\pi|)\cdot|P|.$$

THEOREM 3. *Let P be any sufficiently short path. Then paths P^* and $P\dagger$ exist such that*

$$|P_\pi^* - P^*| \leqq M(|\pi|) \qquad \textit{and} \qquad |P_\pi\dagger - P\dagger| \leqq M(|\pi|).$$

Proof. By the above lemma, the P_π^* and $P_\pi\dagger$ converge in the sense of Cauchy-Fréchet. But this means that for every fixed λ, the $p_\pi^*(\lambda)$ and $p_\pi\dagger(\lambda)$ do, and hence (the space being complete) have limits $p^*(\lambda)$ and $p\dagger(\lambda)$. These limits define P^* and $P\dagger$; the inequalities of Theorem 3 then follow from the corresponding inequalities in the lemma and passage to the limit.

(14δ) $(P^*)\dagger = (P\dagger)^* = P$.

Proof. For every partition π, $(P_\pi^*)_\pi\dagger = (P_\pi\dagger)_\pi^* = P$ by definition. And to replace each segment of P_π^* or $P_\pi\dagger$ by the corresponding segment of P^* resp. $P\dagger$ makes by (14γ) a proportionally small change in $(P_\pi^*)_\pi\dagger$ resp. $(P_\pi\dagger)_\pi^*$. Hence $(P^*)_\pi\dagger \to P$ and $(P\dagger)_\pi^* \to P$ uniformly as $|\pi| \to 0$.

(14ε) $|t(P^*)-t(P)|\leqq o(|P|)$ *and* $|t(P\dagger)-t(P)|\leqq o(|P|)$.

Proof. For every π, $t(P_\pi^*)=t(P_1)\circ\cdots\circ t(P_r)$ where $t(P)=t(P_1)\oplus\cdots\oplus t(P)$. Hence the first inequality merely restates (14α). The proof of the second inequality is the same, since (cf. the preliminary remark in the proof of the lemma above) $|P\dagger|\leqq O(|P|)$.

$$(14\zeta)\quad (P_1\oplus\cdots\oplus P_r)^*=P_1^*\circ\cdots\circ P_r^*$$

and

$$(P_1\circ\cdots\circ P_r)\dagger=P_1\dagger\oplus\cdots\oplus P_r\dagger.$$

Proof. $(P_1\oplus\cdots\oplus P_r)^*$ is in particular the limit as $\sup|P_{k,i}|\to 0$ of $P_{1,1}\circ\cdots\circ P_{r,s(r)}$, where

$$P_{k,1}\oplus\cdots\oplus P_{k,s(k)}=P_k.$$

Thus it is the limit as $\sup|\pi_k|\to 0$ of $(P_1)_{\pi_1}^*\circ\cdots\circ(P_r)_{\pi_r}^*$. By (14γ), this limit is $P_1^*\circ\cdots\circ P_r^*$. This proves the first identity; the proof of the second is similar.

Conversely (14ε)–(14ζ) define the correspondences $P\to P^*$ and $P\to P\dagger$.

(14η) *If Q is any path, and Δ: $[\lambda,\mu]$ is any interval of its domain, then* $q^{-1}(\lambda)\circ q(\mu)=t((\Delta Q\dagger)^*)$.

Remark. By $q^{-1}(\lambda)$ is of course meant $[q(\lambda)]^{-1}$.

Proof. Set $Q\dagger=P$; $p^*(\mu)=p^*(\lambda)\circ t((\Delta P)^*)$ by (14ζ); the result follows by transposing $p^*(\lambda)=q(\lambda)$.

(14θ) *If $P\approx Q$, then $P^*\approx Q^*$ and $P\dagger\approx Q\dagger$.*

Proof. Obvious from the definitions.

CHAPTER II. CANONICAL PARAMETERS

15. **Scalar powers.** In §§16–17, we shall consider straight rays P_x: $p_x(\lambda)=\lambda x$ $(0\leqq\lambda\leqq 1)$ and their star correspondents P_x^*.

Obviously $|P_x|=|x|$, and so by (14ε),

(15α) $|t(P_x^*)-x|\leqq o(|x|)$.

(15β) $|t(P_{x+y}^*)-t(P_x^*)-t(P_y^*)|\leqq M(|x|+|y|)\cdot|y|$.

Proof. Let π denote the partition of $[0,1]$ into n equal parts. Then setting $x_k=x/n$ and $y_k=(x+y)/n$ in (14β), we get (15β) for $(P_{x+y})_\pi^*$, $(P_x)_\pi^*$ and $(P_y)_\pi^*$. Passing to the limit as $n\to\infty$, we get (15β).

Combining (15β) with (15α), we see that the so-called "canonical transformation" $T: x\leftarrow t(P_x^*)$ satisfies‡ $|T(x+y)-T(x)-y|\leqq M(|x|+|y|)\cdot|y|$—is of type (11a). Hence (cf. Theorem 2) we have

‡ By $T: x\leftarrow t(P_x^*)$ we mean that the position x is imagined to be occupied by the element $t(P_x^*)$.

THEOREM 4. *The canonical transformation carries any analytical group G into a topologically isomorphic analytical group.*

Again, by definition, $P_{(\lambda\mu)x}=P_{\lambda(\mu x)}$. While unless $\lambda\mu<0$, $P_{(\lambda+\mu)x}\approx P_{\lambda x}\oplus P_{\mu x}$, whence $t(P_{(\lambda+\mu)x})=t(P_{\lambda x}^*)\circ t(P_{\mu x}^*)$. But

(15γ) *Let $R_x=P_x\oplus P_{-x}$. Then $t(R_x^*)=0$.*

Proof. Let π denote the partition of R_x into $2n$ equal segments, and set $\lambda=1/n$. Then

$$\begin{aligned} |t((R_x)_\pi^*)| &= |(\lambda x)^{n-1}\circ(\lambda x\circ-\lambda x)\circ(-\lambda x)^{n-1}| \\ &\leqq |(\lambda x)^{n-1}\circ(-\lambda x)^{n-1}| + 2(|\lambda x\circ-\lambda x|) \end{aligned}$$

by (2), substituting $(\lambda x)^{n-1}$ for x, $(-\lambda x)^{n-1}$ for y, $(\lambda x\circ-\lambda x)$ for a and Θ for b, and requiring x to be so small that $M(|x|+|a|+|y|)<1$, whence

$$|xay| \leqq |xy| + |a| + M(|x|+|a|+|y|)\cdot|a| \leqq |xy| + 2|a|.$$

But by induction on n and (14α), this yields

$$\begin{aligned} |t((R_x)_\pi^*)| &= (n-1)\cdot o(|\lambda x|) + o(|\lambda x|) = n\cdot o(|\lambda x|) \\ &= n\cdot M(|\lambda x|)\cdot|\lambda|\cdot|x| = M(|\lambda x|)\cdot|x|. \end{aligned}$$

To complete the proof, let $n\to\infty$, so that $M(|\lambda x|)\to 0$.

But if $\lambda\mu<0$, $P_{\lambda x}\oplus P_{\mu x}\approx P_{(\lambda+\mu)x}\oplus P_{-\mu x}\oplus P_{\mu x}$; hence in all cases $t(P_{(\lambda+\mu)x})=t(P_{\lambda x}^*)\circ t(P_{\mu x}^*)$, and so we have

THEOREM 5. *For fixed x, the $t(P_{\lambda x}^*)$ are (locally) topologically isomorphic with the additive group of the λ.*

But by Theorem 4 the canonical transformation is one-one; hence the function x^λ defined by making $t(P_y^*)=x$ and $x^\lambda=t(P_{\lambda y}^*)$ is defined and single-valued near the origin.

(15δ) $x^1=x$ (by definition), $x^\lambda\circ x^\mu=x^{\lambda+\mu}$, *and* (since $P_{(\lambda\mu)x}=P_{\lambda(\mu x)}$) $(x^\lambda)^\mu=x^{\lambda\mu}$.

(15ϵ) *x^λ is a topologico-algebraic function of x, in the sense that any topological isomorphism carrying x into y carries x^λ into y^λ.*

Proof. The assertion is true for positive integral $\lambda=n$ since $(x\circ\cdots\circ x)=x^{1+\cdots+1}=x^n$. It is also true for positive rational λ since $y^n=x$ if and only if $y=(y^n)^{1/n}=x^{1/n}$; while since $x\circ y=\Theta$ if (by 15γ)) and only if (by (9β)) $y=x^{-1}$, it is true for all rational $\lambda=m/n$. Finally, since the rationals are dense in the real continuum and x^λ depends continuously on λ, it is true for all λ.

16. **Canonical parameters.** We are now in a position to introduce canonical parameters.

A group will be said to be under "canonical parameters" if and only if the canonical transformation T: $x \leftarrow t(P_x^*)$ is the identical transformation I: $x \rightarrow x$.

(16α) *Any analytical group is transformed into canonical parameters by the canonical transformation—that is, the canonical transformation is idempotent.*

Proof. After T has been performed once, if π denotes the partition of (0, 1) into n equal parts, then by definition and Theorem 5, $t((P_x)_\pi^*) = (x/n)^n = x$, whence, passing to the limit, iteration of T leaves all points fixed.

(16β) *Under canonical parameters, $x^\lambda = \lambda x$; hence scalar multiplication under canonical parameters is an intrinsic topologico-algebraic operation.*

Proof. By definition of x^λ resp. canonical parameters, $x^\lambda = t(P_{\lambda x}^*) = \lambda x$.

(16γ) *In any analytical group, $x + y = \lim_{\lambda \to 0}(\lambda x \circ \lambda y)/\lambda$.*

Proof. Referring to the inequality (3), we get for fixed x and y since $\lambda(x+y) = \lambda x + \lambda y$,

$$|(\lambda x \circ \lambda y) - \lambda(x + y)| \leqq M(|\lambda|) \cdot |\lambda|.$$

Hence, dividing through by the scalar λ,

$$|(\lambda x \circ \lambda y)/\lambda - (x + y)| \leqq M(|\lambda|)$$

which completes the proof.

Combining (16γ) with (16β), we get

THEOREM 6. *If G and H are any two analytical groups under canonical parameters, then any topological isomorphism between G and H is linear—it preserves vector sums and scalar products.*

COROLLARY 6.1. *The group of topological automorphisms of any analytical group is spatially isomorphic with a group of linear transformations of its parameter-space.*

COROLLARY 6.2. *If G and H are any two analytical groups, then any continuous isomorphism between G and H can be expressed as the product of three transformations of the parameter-space of G, of types* (11a), (11b), *and* (11a).‡

COROLLARY 6.3. *Admissible paths* (cf. §13) *are carried into admissible paths under topological isomorphisms between analytical groups.*

(16δ) *An analytical group is under canonical parameters if and only if $x \circ x = x + x$ for all x.*

‡ One should prove further: *Any topological isomorphism between two groups whose function of composition is analytical, amounts to an analytical transformation of coordinates.* To complete the proof, it would suffice to show that in such groups $t(P_x^*)$ is an analytical function of x—a fact already known (from the theory of differential equations) for Lie groups.

Proof. If $x \circ x = x + x$, then by induction $x^{2^n} = 2^n x$, $x = (2^{-n}x)^{2^n}$, whence $t(P_x^*) = x$, and we have canonical parameters. Conversely, under canonical parameters, $x \circ x = x^2 = 2x = x + x$.

17. **Digression: Topologico-algebraic postulates.** It is a curious fact that, by inverting the remarks of the last few sections, one can obtain *topologico-algebraic postulates defining Lie groups*, involving only *intrinsic* operations (i.e., operations invariant under topological isomorphisms). To show this, one need use only superficial reasoning, arguing from the above properties of canonical parameters.†

One can do this even for infinite continuous groups. The general procedure is: 1°: characterize Banach spaces topologico-algebraically (as those complete topological linear spaces possessing a convex open "bounded" set‡); 2°: define linear transformations and thence Fréchet total derivatives (cf. §18) topologico-algebraically;§ 3°: postulate that the group is a Banach space relative to addition under canonical parameters ("canonical addition") and raising to scalar powers; 4°: postulate that an associative operation of multiplication satisfying $x \circ x = x + x$ and continuously differentiable on the Banach space, be defined.

Because of the preceding results and Corollary 2 of Theorem 15, these postulates are satisfied by all analytical groups under canonical parameters. Conversely, by Theorem 8 any system satisfying these postulates is an analytical group, which is by (16δ) under canonical parameters.

In the special case of Lie groups—the case that the parameter space has a finite basis (or equivalently,‖ is locally compact)—one can simplify these postulates to the requirements (i) elements $a_1, \cdots, a_r$ exist such that any element near the identity can be represented uniquely as a product $a_1^{\lambda_1} \circ \cdots \circ a_r^{\lambda_r}$ of small powers $a_k^{\lambda_k}$ of the a_k, and (ii) the function of composition is continuously differentiable¶ in $(\lambda_1, \cdots, \lambda_r)$-space.

18. **Digression: metric postulates.** The present section will be devoted to sketching a proof of

---

† These ideas were announced in Abstract 41-5-192 of the Bulletin of the American Mathematical Society (1935).

‡ For the terminology cf. J. von Neumann, *On complete topological spaces*, these Transactions, vol. 37 (1935), pp. 1–20. For the characterization cf. A. Kolmogoroff, *Zur Normierbarkeit eines allgemeinen topologisches lineares Raumes*, Studia Mathematica, vol. 5 (1935), pp. 29–33.

§ Replace the usual epsilon-delta definitions by "for every given neighborhood there exists a neighborhood so small $\cdots$."

‖ Cf. [1], p. 84, Theorem 8.

¶ This can be phrased topologico-algebraically. For instance, $x \circ y = f(x, y)$ has continuous first derivatives if and only if $\partial f/\partial x(a, b) = \lim_{\lambda \to 0}((a + \lambda x) \circ b)/\lambda$ and $\partial f/\partial y(a, b) = \lim_{\lambda \to 0}(a \circ (b + \lambda y))/\lambda$ exist and are continuous functions of a and b.

THEOREM 7. *One can redefine the class of analytical groups under canonical parameters by weakening the postulates for Banach spaces.*

This result will not be used elsewhere.

Sketch of proof. Substitute "group products" $x \circ y$ for vector sums $x+y$, and "scalar powers" x^λ for scalar products λx, continue to use an (extrinsic) norm function $|x|$, and make the following alterations in the usual postulates (cf. [1]) for Banach space (after first confining their validity to a small region about the identity): (1) replace the two conditions $x+y=y+x$ and $\lambda(x+y)=\lambda x+\lambda y$ by the single weaker condition $|x^\lambda \circ y^\lambda \circ (x \circ y)^{-\lambda}| \leqq |\lambda| \cdot |x| \cdot |y|$, and (2) replace the condition $|x+y| \leqq |x|+|y|$ by the weaker condition $|x \circ y| \leqq |x|+|y|+|x| \cdot |y|$.

The reader should have no difficulty in proving that the altered postulates hold in any analytical group under canonical parameters and under a suitable norm function (cf. Theorems 1 and 5 for the algebraic identities, and (27β)—where it is shown that essentially $|(x \circ y)-(x+y)| \leqq |x| \cdot |y|$—for the strong metric inequalities).

But conversely, if one defines $x \dotplus y = \lim_{\lambda\to 0}(x^\lambda \circ y^\lambda)^{1/\lambda}$ and $\lambda x = x^\lambda$ in any system G satisfying the new postulates, then the space becomes a neighborhood of the origin of a Banach space B, and the map of G on B satisfies (1), (2′), (2″) and $x \circ x = x+x$—completing the outline of the proof.

19. **Digression: differentiability postulates.** We now come to the connection between Definition 1 and differentiability conditions, namely

THEOREM 8. *Let G be any topological group nucleus, some neighborhood of whose identity e is mapped onto a region of a Banach space B, in such a way that $x \circ y = f(x, y)$ has first total derivatives everywhere, which are continuous at e. Then G is an analytical group under the map.*

Proof. Theorem 8 is clearly meaningless until continuous total derivatives have been defined; actually, it refers to the usual definitions due to Fréchet.† Fréchet says that $f(x, y)$ has a total derivative A with respect to x at $x=a$, $y=b$ if and only if there exists a linear transformation A such that

$$(18') \qquad |f(a+x, b) - f(a, b) - Ax| \leqq o(|x|),$$

where Ax denotes the transform of x by A. One similarly defines total derivatives with respect to y. Further, Fréchet calls the two total derivatives $A(x, y)=\partial f/\partial x(x, y)$ and $B(x, y)=\partial f/\partial y(x, y)$ *continuous at* $x=a$, $y=b$ if and only if

† M. Fréchet, *La notion de différentielles dans l'analyse générale*, Annales de l'École Normale Supérieure, (3), vol. 42 (1925), pp. 293–323. For a similar concept of an infinite continuous group, cf. A. D. Michal and V. Elconin, *Abstract transformation groups*, American Journal of Mathematics, vol. 59 (1937), pp. 129–144.

$$(18'') \quad \begin{cases} |A(a+u, b+v)x - A(a, b)x| \leqq M(|u| + |v|) \cdot |x|, \\ |B(a+u, b+v)y - B(a, b)y| \leqq M(|u| + |v|) \cdot |y|. \end{cases}$$

Clearly $A(e, e) = B(e, e) = I$, the identical linear transformation—since irrespective of u, $u \circ e = e \circ u = u$.

Once these definitions and this fact have been stated, the proof of Theorem 8 follows familiar lines. Assuming the existence everywhere and continuity at $x = y = e$ of $A(x, y)$ and $B(x, y)$, one constructs the real functions

$$\phi(\lambda) = |(\lambda x \circ a) - (\lambda x + a)|,$$
$$\psi(\mu) = |(x \circ a \circ \mu y) - (x + a + \mu y)|.$$

Clearly (18′) implies that the upper right-derivatives of $\phi(\lambda)$ and $\psi(\mu)$ are bounded by $|A(\lambda x, a) - I| \cdot |x|$ and $|B(x \circ a, \lambda y) - I| \cdot |y|$, respectively. Hence by the theory of real functions,

$$(3) \quad |(x \circ a \circ y) - (x + a + y)| \leqq K(|a| + |x| + |y|) \cdot (|x| + |y|),$$

where $K(|a| + |x| + |y|)$ is small as long as $|A(\lambda x, a) - I|$ and $|B(x \circ a, \mu y) - I|$ are small identically on $0 \leqq \lambda, \mu \leqq 1$—and so by the continuity of these is an M-function, q.e.d.

Remark. Fréchet's definition obviously specializes to the usual definition of continuous total differentiability when B is finite-dimensional—and is satisfied in this case provided continuous first partial derivatives with respect to all coordinates exist.† (This remark has immediate application to the theory of Lie groups—it shows that if the function $x \circ y = f(x, y)$ has continuous first partial derivatives, then one is dealing with an "analytical group.")

In summary, §§17–19 have contained three alternative definitions of analytical groups, equivalent to Definition 1. One can view these from two angles. They may be regarded from a conceptual angle as giving a better picture of what an analytical group is. Or they may be regarded as giving content to Definition 1 itself—that is, as furnishing examples of analytical groups from other contexts.

Chapter III. Linear groups

20. **Axiomatization.** It is a simple fact, that *one can axiomatize algebras of linear operators‡ on Banach spaces.*

To see this, one must first recall that the operators on *any* linear space B which are defined everywhere, and carry vector sums into vector sums and scalar multiples into scalar multiples, constitute a hypercomplex algebra with

† C. J. de la Vallée-Poussin, *Cours d'Analyse Infinitésimale*, Louvain, 1914, p. 141.

‡ By a "linear operator," we mean ([1], p. 23) any continuous additive, everywhere defined function. This conflicts with the usage for Hilbert spaces, where such operators are called *bounded*.

a principal unit I. (We shall use the notation O for the transformation carrying every $x \epsilon B$ into Θ; I for the identity $x \rightarrow x$; $S, T, U, \cdots$ for other operators.)

One must next observe that if B is a Banach space, then relative to vector sums $T+U$, products λT with scalars, and the "modulus" $\|T\| = \sup_{x \neq \Theta} |Tx|/|x|$ (cf. [1], p. 54), the linear operators on B constitute another Banach space. The proof of this will be left to the reader.†

Finally, $\|T \circ U\| \leqq \|T\| \cdot \|U\|$.

More generally, *any* algebra of linear operators on a Banach space B which contains I and is topologically closed (under the "uniform" topology defined by the metric $\|T-U\|$) has all of the properties just described.

But conversely, let $\mathfrak{H}$ be any system having these properties—i.e., any "metric hypercomplex algebra."‡ Then (applying a classical construction) each element $T \epsilon \mathfrak{H}$ induces a linear transformation θ_T: $X \rightarrow XT$ on the elements $X \epsilon \mathfrak{H}$. Moreover since $\|IT\| = \|I\| \cdot \|T\|$ and $\|XT\| \leqq \|X\| \cdot \|T\|$, the "modulus" of θ_T is precisely $\|T\|$. Thus $\mathfrak{H}$ can be realized as a closed algebra of linear operators on itself, including the identity θ_I: $X \rightarrow XI = X$.

21. **Linear operators with inverses.** Linear operators do not constitute a group under multiplication. But the linear operators S with inverses S^{-1} satisfying $SS^{-1} = S^{-1}S = I$ do. And one can easily prove

THEOREM 9. *Let $\mathfrak{H}$ be any metric hypercomplex algebra. Then the map $(I+T) \rightarrow T$ of the elements $(I+T) \epsilon \mathfrak{H}$ with $\|T\| < \frac{1}{2}$ onto the linear space defined by $\mathfrak{H}$, exhibits these elements as an analytical group $\mathfrak{G}$ under multiplication.*

Proof. Refer to Definition 1. The only properties in any doubt are (2′)–(2″). But

$$\begin{aligned} \Xi &\equiv \|[(I + X) \circ (I + Y) - (I + X) \circ (I + Z)] - [Y - Z]\| \\ &= \|[X \circ Y - X \circ Z]\| = \|X \circ (Y - Z)\|, \quad \text{by algebra,} \\ &\leqq \|X\| \cdot \|Y - Z\|, \qquad \text{by hypothesis,} \end{aligned}$$

proving (2′). One obtains (2″) similarly.

THEOREM 10. *The canonical transformation of $\mathfrak{G}$ is given explicitly by the convergent power series*

$$T \leftarrow \exp(T) - I \equiv T + \frac{1}{2!} T^2 + \frac{1}{3!} T^3 + \cdots = t(P_T^*).$$

† For instance, if $\{T_n\}$ is a fundamental sequence of linear operators, then for any x, $\{T_n x\}$ is a fundamental sequence in B, whose limit we shall define as Tx. By continuity, $T(x+y) = Tx + Ty$ and $|Tx - Ty| \leqq \lim_{n \to \infty} \|T_n\| \cdot |x - y|$.

‡ More properly, any metric *associative* hypercomplex algebra. Omitting the associative law, we get a more general definition (cf. §30), which however yields no realization theorem.

Proof. The questions of convergence are settled by the inequalities $\|T^n\| \leqq \|T\|^n$ and $\|T+U\| \leqq \|T\|+\|U\|$, and the assumption $\|T\| < \frac{1}{2}$. But now dividing P_T into n equal parts, we get by the binomial expansion

$$t(P_T)_\pi^* = \left(I + \frac{1}{n} T\right)^n - I = T + \frac{1}{n^2} C_{n,2}T^2 + \frac{1}{n^3} C_{n,3}T^3 + \cdots$$

which converges (cf. supra) to $\exp(T) - I$.

The inverse of the canonical transformation is of course given by the power series

$$T \leftarrow \log (I + T) = T - \tfrac{1}{2}T^2 + \tfrac{1}{3}T^3 - \cdots ,$$

but we shall not use this fact.†

22. **Generalization of Theorem** 9. A metric algebra $\mathfrak{X}$ need not possess a unit I nor satisfy $\|I\| = 1$ in order that the symbolic elements $I + X$ with $X \epsilon \mathfrak{X}$ and $\|X\| < \frac{1}{2}$ should form an analytical group nucleus when multiplied according to the rule

$$(22.1) \qquad (I + X) \circ (I + Y) = I + (X + Y + X \circ Y).$$

The arguments of §21 do not involve these assumptions.

An important example of such an algebra is due to Delsarte [6], and is also cited by Yosida (op. cit.). It is the algebra $\mathfrak{A}$ of all infinite matrices $A = \|a_{ij}\|$ for which $\sum_{i,j} |a_{ij}|^2 < +\infty$. If we set $\|A\|^2 = \sum_{i,j} |a_{ij}|^2$, we have a Banach space, in which products $C = A \circ B = \|c_{ij}\|$ can be defined by the usual rule $c_{ij} = a_{ik}b_{kj}$ —the series being convergent by Schwarz' inequality, and satisfying, besides, $\|C\| \leqq \|A\| \cdot \|B\|$.

The algebra $\mathfrak{A}$ corresponds of course to the algebra of Schmidt kernels in the theory of integral equations, and is isometric with Hilbert space.

23. **Function of composition.** The formulas of the preceding section lead directly to explicit expressions for the function $X \circ Y$ of composition.

Under the original parameters, $X \circ Y = F(X, Y)$ is analytic since (by the distributivity of multiplication) it is *linear in both variables*—and among the functions between linear spaces, next to constant functions, linear functions

† Remark: The above treatment was suggested by that of J. von Neumann [12]. The main changes are: explicit discussion of transformations as abstract elements, and use (following Banach) of the "modulus" $\|X\|$ for norm.

The concept of a metric hypercomplex algebra ("complete normed vector ring") was announced by the author in Abstract 41-3-104 of the Bulletin of the American Mathematical Society (1935); a similar definition is given by K. Yosida (*On the group embedded in the metrically complete ring*, Japanese Journal of Mathematics, vol. 13 (1936), pp. 7–26). Yosida does not require $\|I\| = 1$; cf. §22.

Another example, discussed at length by M. H. Stone, consists of the linear operators T_a: $f(x) \rightarrow f(x)a(x)$ on the space of bounded functions on an abstract class. This is a closed subalgebra of the algebra of §20.

are the most purely analytic. Thus the second partial derivatives of $F(X, Y)$ are constant and so the higher derivatives all vanish identically.

Moreover by Theorem 10, if we denote by X^m as usual $X \circ \cdots \circ X$, then we have the following explicit expression for $X \circ Y = G(X, Y)$ *under canonical parameters*,

$$\begin{aligned} G(X, Y) &= \log (I + F(\exp X - I, \exp Y - I)) \\ &= \log\left(I + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{1}{m!n!} F(X^m, Y^n)\right) \\ &= \sum_{k=1}^{\infty} (-1)^{k-1}/k \left\{ \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} \frac{1}{m!n!} F(X^m, Y^n)\right\}^k \end{aligned}$$

whose first terms can be found easily, and are monomials.

It has been shown by J. E. Campbell [3] and F. Hausdorff [8] that this series can also be developed in terms of X, Y, and iterations of the bilinear function $[X, Y] = XY - YX$. The resulting "SCH-series" will be proved in Chapter V to be valid also for non-linear analytical groups.

24. **Digression: polynomials and analyticity.** The *algebraic* significance of the SCH-series will be discussed in Chapter V; what about its *analytical* significance?

It exhibits $G(X, Y)$ as analytical in the strong sense that (1) it is the limit of an absolutely convergent series of polynomials of increasing degrees,† (2) its derivatives all exist and can be found through term-by-term differentiation of the series, (3) hence the Taylor's series for $G(X, Y)$ converges absolutely to $G(X, Y)$—all within a sphere of positive radius.

Although it is not entirely clear when a function between Banach spaces is "analytical"—there may be various generalizations of the established notion for functions between euclidean spaces—it seems undeniable that at least any function with properties (1)–(3) should be called analytical.

25. **Adjoint of an analytical group.** In the present section, we shall show that the notion of the *adjoint* of a Lie group can be extended without real modification to the case of analytical groups. We state this more precisely in the following theorem:

† For polynomial functions between Banach spaces, cf. S. Mazur and W. Orlicz, *Grundlegende Eigenschaften der polynomische Operatoren*, Studia Mathematica, vol. 5 (1935), pp. 50–68 and pp. 179–189. One can define polynomials through continuity + the identical vanishing of $(n+1)$st differences, through the identical vanishing of $(n+1)$st derivatives, or as sums of multilinear functions in a variable repeated $0, \cdots, n$ times; and these definitions are equivalent.

Unlike these authors, we are concerned with functions of two variables. N.B.: A polynomial function on r variables which is homogeneous of degree k in each, is homogeneous of degree kr (and not of degree k) on the product-space of the variables.

THEOREM 11. *Let G be any analytical group under canonical parameters. Then each element $g\epsilon G$ determines a linear transformation θ_g: $x \to g^{-1}xg$ on the parameter-space of G, and the correspondence $g \to \theta_g$ is continuously homomorphic.*

Proof. Since G is a topological group θ_g is a topological automorphism. Hence by Corollary 6.1 it is a linear transformation on G. Moreover by the well-known identity $(gh)^{-1}x(gh) = h^{-1}(g^{-1}xg)h$, the correspondence $g \to \theta_g$ is homomorphic. It remains only to show that it is continuous under the *uniform* topology. But

$$\begin{aligned}
| h^{-1}xh - g^{-1}xg | &= O(| h^{-1}x^{-1}hg^{-1}xg |), \quad \text{by (3)}, \\
&= O(| g^{-1}[(hg^{-1})^{-1}x^{-1}(hg^{-1})x]g) \\
&= O(| (hg^{-1})^{-1}x^{-1}(hg^{-1})x) \\
&\qquad\qquad (\text{since } \|\theta_g\| = O(|g|), \quad \text{by } (14\beta)) \\
&= O(| g - h |) \cdot | x |, \quad \text{by } (27\beta),
\end{aligned}$$

whence $\|\theta_g - \theta_h\| = O(|g-h|)$, completing the proof.

The validity of the proof of course depends on proving (27β) without the aid of Theorem 11. We shall do this in §27.

The correspondence $g \to \theta_g$ does not always carry open sets into open sets: it need not be "gebietstreu" in the sense of Freudenthal.

CHAPTER IV. COMMUTATION

26. **Outline.** The present chapter will be devoted to showing how every analytical group G possesses a bilinear "commutation function" $[x, y]$. In Chapter V, it will be shown that $[x, y]$ determines G to within local isomorphism.

The commutation function $[x, y]$ belonging to a given group G is most easily defined as the bilinear asymptote at $x = y = 0$ to the purely algebraic commutator function

$$(x, y) = K(x, y) \equiv x^{-1}y^{-1}xy.$$

The fact that (x, y) *has* a bilinear asymptote is proved below (in §28) from the relations (deduced in §27)

$$(27\alpha) \qquad \begin{cases} | (u \circ x, y) - (x, y) - (u, y) | \leqq M(| x | + | y |) \cdot | (u, y) |, \\ | (x, v \circ y) - (x, y) - (x, v) | = M(| x | + | y |) \cdot | (x, v) |, \end{cases}$$

$$(27\beta) \qquad | (x, y) | = O(| x | \cdot | y |),$$

while the fact that $[x, y]$ is a topologico-algebraic invariant associated with G is almost obvious (cf. §29).

Moreover one can deduce the familiar identities

$$(30\alpha)\qquad \begin{cases}[x, y] + [y, x] = 0,\\ [[x, y], z] + [[y, z], x] + [[z, x], y] = 0\end{cases}$$

as corollaries of formal identities on group products. These results can be summarized in the statement that G possesses a *metric Lie algebra* $L(G)$. Chapter IV concludes with various applications of $L(G)$, to the case that G is under canonical parameters.

In the proofs of Chapter IV, group algebra plays a novel and essential role.

27. **The approximate bilinearity of** $K(x, y)$. The present section will be devoted to showing that $(x, y) = K(x, y)$ is approximately bilinear at $x = y = 0$, in the sense that (27α)–(27β) are true.

The proof of (27α) is almost immediate. One has the formal identity

$$(27\gamma)\qquad \begin{aligned}(u \circ x, y) &= x^{-1}u^{-1}y^{-1}uxy\\ &= (u, y)_x \circ (x, y)\end{aligned}$$

under the convention that g_x denotes $x^{-1}gx$. But by the fundamental inequality (2) of §8,

$$(27\delta)\qquad |(u, y)_x - (u, y)| \leqq M(|x|)\cdot|(u, y)|,$$

whence $|(u, y)_x| = O(|(u, y)|)$. Hence

$$\begin{aligned}\Xi &\equiv |(u \circ x, y) - (x, y) - (u, y)|\\ &\leqq |(u \circ x, y) - (x, y) - (u, y)_x| + |(u, y)_x - (u, y)|\\ &\leqq M(|(x, y)|)\cdot|(u, y)_x| + M(|x|)\cdot|(u, y)|\end{aligned}$$

(by (27γ) and (3) of §9, and (27δ))

$$\leqq M(|x| + |y|)\cdot|(u, y)|,$$

since $|(u, y)_x| = O(|(u, y)|)$. But this is the first half of (27α); the second half follows from the symmetry between left- and right-multiplication.

As a special instance of (27α), we have

$$|(x^m, y) - (x^{m-1}, y) - (x, y)| \leqq M(|x^m| + |y|)\cdot|(x, y)|.$$

Hence, since $|x^m| = O(|mx|) = O(m|x|)$, by induction

$$|(x^m, y) - m(x, y)| \leqq M(m|x| + |y|)\cdot m\cdot|(x, y)|.$$

Combining with the symmetric formula in (x, y^n), we get

$$(27\epsilon)\qquad |(x^m, y^n) - mn(x, y)| = M(m|x| + n|y|)\cdot mn|(x, y)|.$$

Consequently, within some small radius ρ of the origin,

$$(27\zeta)\qquad |(x, y)| \leqq O(|(x^m, y^n)|/mn).$$

But clearly within this sphere, given $x\neq 0$ and $y\neq 0$, one can so choose m and n that $\frac{1}{2}\rho<|x^m|$, $|y^n|<\rho$—whence, $|(mx, ny)|$ being bounded within this sphere, we get $|(x^m, y^n)|\leqq O(|mx|\cdot|ny|)$, and so by (27ζ),

$$(27\beta)\qquad |(x, y)| \leqq O(|x|\cdot|y|).$$

It is a corollary, since $y\circ x\circ K(x, y)=x\circ y$, and likewise $(y\circ x)+(x\circ y-y\circ x)=x\circ y$, that (by (3))

$$(27\beta')\qquad |xy - yx| \leqq O(|x|\cdot|y|).$$

28. **The asymptote** $[x, y]$. Substituting from (27β) in (27α), and recalling that $x+w=u\circ x$ implies† $|w|\sim|u|$, we get

$$(28\alpha)\qquad \begin{cases} |(x+w, y)-(x, y)-(w, y)| \leqq M(|x|+|y|)\cdot|w|\cdot|y|, \\ |(x, y+w)-(x, y)-(x, w)| = M(|x|+|y|)\cdot|x|\cdot|w|, \end{cases}$$

from which there follows

$$(28\alpha')\qquad |(x+u, y+v)-(x, y)| \leqq O(|u|+|v|).$$

Now start anew with (28α)–$(28\alpha')$, and use the same algebraic analysis used in proving (27ϵ). By (28α),

$$|(mx, y)-((m-1)x, y)-(x, y)| \leqq M(m|x|+|y|)\cdot|x|\cdot|y|.$$

Hence by induction on m, we get

$$|(mx, y)-m(x, y)| \leqq M(m|x|+|y|)\cdot m|x|\cdot|y|.$$

Combining with the symmetric formula in (x, ny), we have

$$(28\beta)\qquad |(mx, ny)-mn(x, y)| \leqq M(m|x|+n|y|)\cdot mn\cdot|x|\cdot|y|.$$

By double use of (28β), we get for $0<h/m, k/n<1$,

$$\left|\left(\frac{h}{m}x, \frac{k}{n}y\right)-\frac{hk}{mn}(x, y)\right| \leqq M(|x|+|y|)\cdot\frac{hk}{mn}\cdot|x|\cdot|y|,$$

whence, by rational approximation and passage to the limit, using $(28\alpha')$ to establish continuity, we have

$$(28\gamma)\qquad \left|\frac{1}{\lambda\mu}(\lambda x, \mu y)-(x, y)\right| \leqq M(|x|+|y|)\cdot|x|\cdot|y|$$

for $0<\lambda, \mu<1$. Therefore if $\lambda+\mu+\lambda'+\mu'<\epsilon$, then

† By $|w|\sim|u|$ we mean that $\big||w|-|u|\big|\leqq M(|x|+|u|+|w|)\cdot|u|$; this relation is evidently reflexive, symmetric, and transitive.

$$\left| \frac{1}{\lambda\mu}(\lambda x, \mu y) - \frac{1}{\lambda'\mu'}(\lambda' x, \mu' y) \right| = M(\epsilon)\cdot |x| \cdot |y|$$

and so, by the completeness of the parameter-space

$$[x, y] \equiv \lim_{\lambda,\mu \downarrow 0} \frac{1}{\lambda\mu} K(\lambda x, \mu y) \tag{28δ}$$

exists. Furthermore, by (28γ),

$$|[x, y] - (x, y)| \leqq M(|x| + |y|)\cdot |x| \cdot |y|. \tag{28ϵ}$$

Finally, since $(-x+x, y) = (0, y) = 0$, by (28α)

$$|(-x, y) - [-(x, y)]| \leqq M(|x| + |y|)\cdot |x| \cdot |y|.$$

whence we see that

$$[x, y] = \lim_{\lambda,\mu \to 0} \frac{1}{\lambda\mu}(\lambda x, \mu y) \tag{28ζ}$$

exists.

29. **The bilinearity of** $[x, y]$, **etc.** In this section, we shall prove the bilinearity and topologico-algebraic invariance of $[x, y]$.

The invariance of $[x, y]$ under continuous isomorphisms between groups under canonical parameters follows from the definition and Theorem 6. And by (28α), "distortions" of type (11a) change $(\lambda x, \mu y)$ by $o(|\lambda| \cdot |\mu|)$, from which invariance under general continuous isomorphisms follows by Corollary 6.2.

As for the bilinearity of $[x, y]$, by (28α)

$$\Xi_1 \equiv \left| \frac{1}{\lambda\mu}(\lambda[x+u], \mu y) - \frac{1}{\lambda\mu}(\lambda x, \mu y) - \frac{1}{\lambda\mu}(\lambda u, \mu y) \right|$$
$$= M(|\lambda x| + |\mu y|)\cdot \frac{1}{\lambda\mu}\cdot |\lambda u| \cdot |\mu y|$$

whence, passing to the limit, $[x+u, y] - [x, y] - [u, y] = 0$. Hence $[x+u, y] = [x, y] + [u, y]$; and $[x, y+v] = [x, y] + [x, v]$ by symmetry. Also, by (27β) and (28ϵ), $[x, y] = O(|x| \cdot |y|)$, and so is bounded. Hence it is bilinear.

In summary of the above results,

THEOREM 12. *(x, y) has a bilinear asymptote $[x, y]$ which is a topologico-algebraic invariant of G.*

Remark. In a linear group, algebra based on the expansion $(I+\lambda X)^{-1} = I - \lambda X + \lambda^2 X^2 - \lambda^3 X^3 + \cdots$ shows

$$\frac{1}{\lambda^2}(\lambda X, \lambda Y) = (XY - YX) + \text{terms of higher order}$$

whence, passing to the limit, $[X, Y] = XY - YX$.

30. **Metric Lie algebras.** One can now deduce relations (30α) from algebraic identities on group products.

In the first place, since $(y, x) = (x, y)^{-1}$, and $u^{-1} + u$ is nearly zero, clearly $[x, y] + [y, x] = 0$. That is, $[x, y]$ is skew-symmetric.

The proof that $[x, y]$ satisfies Jacobi's identity,

$$[[x, y], z] + [[y, z], x] + [[z, x], y] = 0$$

is less simple. It depends very essentially on realizing that by (27β) and (28ε),

$$\begin{aligned} \Xi_2 &\equiv |((x, y), z) - [[x, y], z]| \\ &\leqq |((x, y), z) - ([x, y], z)| + |([x, y], z) - [[x, y], z]| \\ &\leqq M(|x| + |y| + |z|)\cdot|x|\cdot|y|\cdot|z| \end{aligned} \tag{30β}$$

and, besides, on remarking that since $v \circ u = u \circ v \circ (v, u)$, to permute two commutators in a group product changes the value by an amount which is by (27β) small to the fourth order.

But direct computation based on cancellation proves†

$$(x, y)((x, y), z)(z, y)(z, x)((z, x), y)(y, x)(y, z)((y, z), x)(x, z) = \Theta.$$

Therefore, permuting terms, and cancelling

$$(x, y)(y, x) = (z, y)(y, z) = (z, x)(x, z) = \Theta,$$

we get by the preceding remark, the inequality

$$|((x, y), z)((y, z), x)((x, z), y)| \leqq O(|x| + |y| + |z|)\cdot|x|\cdot|y|\cdot|z|.$$

Hence by (30β) and the fundamental relation (3),

$$|[[x, y], z] + [[y, z], x] + [[z, x], y]| = M(|x| + |y| + |z|)\cdot|x|\cdot|y|\cdot|z|.$$

Replacing x, y, z by $\lambda x, \lambda y, \lambda z$ where λ is small, and using linearity, we get Jacobi's Identity in the limit.

Summarizing, we may say (in the language of Chapter III),

THEOREM 13. *Relative to sums $x+y$, scalar products λx, and "brackets" $[x, y]$, the parameter-space of any analytical group nucleus G is a metric Lie algebra $L(G)$.*

Remark 1. In §§26–30 we have nowhere assumed that G was under canonical parameters.

† This formula was suggested to the author by identities in §2.3 of [7].

Remark 2. Since $|[x, y]| \leqq O(|x| \cdot |y|)$, after changing the scale (i.e., multiplying the norm by a suitable constant) we can assume simply $|[x, y]| \leqq |x| \cdot |y|$.

Remark 3. Brackets $[x, y]$ are defined for all x, y in the Banach space B, unlike $x \circ y$ which is defined only locally.

We shall show (Corollary 15.1) that G is determined to within local isomorphism by $L(G)$, and that conversely any metric Lie algebra belongs to an analytical group nucleus. This shows that the problem of enumerating the analytical group nuclei with a given parameter-space is equivalent to that of enumerating the different metric Lie algebras on the same linear space.

31. **Subgroups and normal subgroups.** The results of §§31–32 will refer to analytical groups G under canonical parameters. A subset S of elements of G will be called an *analytical subgroup nucleus* if and only if, relative to the topology and group multiplication table of G, S is itself an analytical group nucleus. An analytical subgroup nucleus S will be called *normal* (or invariant) if and only if for every $g \epsilon G$, $g^{-1}Sg$ contains some neighborhood of the identity of S.

If S is an analytical subgroup nucleus, then each $x \epsilon S$ must lie on a one-parameter subgroup x^λ in S, and hence (by Theorem 6) a segment λx in G must lie in S. Again, the length of this segment must exceed some fixed positive constant; otherwise we could find $\{x_n\}$ such that $\lambda_n x_n \epsilon S$ implies $\lambda_n x_n \to 0$, and this is impossible in an analytical group nucleus.

Therefore S must contain with x and y, $k(\lambda x, \lambda y)/\lambda^2$ for some fixed $k > 0$ and all λ on $[0, 1]$; hence it must contain with x and y, $k[x, y]$ (since, being *complete*, it is *closed* in G). Similarly, it must contain with x and y, $x + y = \lim_{\lambda \to 0}(\lambda x \circ \lambda y)/\lambda$. And finally, if two such subgroup nuclei contain elements on the same class of segments $\lambda x \epsilon G$, then they clearly generate (in case G is a group) the same subgroup of G, and so may be identified.

These facts may be summarized in

(31α) *Let G be any analytical group nucleus under canonical parameters. The analytical subgroup nuclei S of G are pieces of closed subalgebras of the metric Lie algebra $L(G)$, two subgroups being identical if and only if the subalgebras are.*

If S is "normal" (i.e., invariant under all inner automorphisms), then $x \epsilon S$ and $g \epsilon G$ imply that for some $k > 0$, $k[g, x] = \lim_{\lambda \to 0}(k\{\lambda x \circ \lambda g_{\lambda x}\}/\lambda^2) \epsilon S$. Furthermore,

$$g + x = \lim_{n \to \infty} \left[\left(\frac{1}{n} g \right) \circ \left(\frac{1}{n} x \right) \right]^n$$

$$= g \circ \prod_{k=1}^{n} \left\{ \left(\frac{1}{n} g \right)^{k-n} \circ \left(\frac{1}{n} x \right) \circ \left(\frac{1}{n} g \right)^{n-k} \right\},$$

$$\epsilon g S = Sg.$$

Hence $g+S=Sg$, and

(31β) *If S is normal, then the associated subalgebra of $L(G)$ is invariant*† *and the cosets of S are the hyperplanes parallel to the manifold of S.*

We shall prove converses to (31α)–(31β) in (32β) and corollary 15.5.

32. **The adjoint group.** Using the commutation function, one can easily deduce an explicit series for the adjoint group of §25.

Define T: $u \to T(u) = (y/n)^{-1} \circ u \circ (y/n)$. Then T is a linear transformation, and

$$\begin{aligned} \left| T(u) - \left(u + \left[u, \frac{1}{n} y \right] \right) \right| &\leqq \left| u \circ \left(u, \frac{1}{n} y \right) - u - \left(u, \frac{1}{n} y \right) \right| \\ &\quad + \left| \left(u, \frac{1}{n} y \right) - \left[u, \frac{1}{n} y \right] \right| \\ &\leqq \frac{1}{n} \cdot M \left(|u| + \frac{1}{n} |y| \right) |u| \cdot |y| \end{aligned}$$

(by (3), (27β) and (28ε)).

Hence by n-fold iteration and the binomial expansion,

$$\begin{aligned} \left| T^n(u) - \left\{ u + [u, y] + C_{n,2} \cdot \frac{1}{n^2} [[u, y], y] + \cdots \right\} \right| \\ \leqq M \left(|u| + \frac{1}{n} |y| \right) |u| |y|. \end{aligned}$$

Whence, since $T^n(u) = y^{-1} \circ u \circ y$, passing to the limit, we have

$$\begin{aligned} |w(u, y)| \equiv \left| \{ y^{-1} \circ u \circ y \} - \left\{ u + [u, y] + \frac{1}{2!} [[u, y], y] + \cdots \right\} \right| \\ = o(|u|) \cdot |y|. \end{aligned}$$

But since the terms are all linear in u, clearly

$$|w(u, y)| = n \left| w \left(\frac{1}{n} u, y \right) \right| = n \cdot o(1/n) \cdot |u| \cdot |y|.$$

That is, letting $n \uparrow \infty$, $|w(u, y)| = 0$, and so

† In the usual sense, that $x \epsilon S$ and $g \epsilon L(G)$ imply $[g, x] \epsilon S$.

$$(32\alpha)\quad y^{-1}xy = x + [x, y] + \frac{1}{2!}[[x, y], y] + \frac{1}{3!}[[[x, y], y], y] + \cdots .$$

From (32α) we deduce as a corollary,

(32β) *If the subalgebra associated with a given subgroup S of an analytical group G is invariant, then S is a normal subgroup.* (Converse of (31β).)

Chapter V. Function of composition

33. **Introduction.** The main purpose of this chapter is to show in §§34–36 how the function $x \circ y = f(x, y)$ of composition of any analytical group under canonical parameters, can be written as the sum of an infinite series of polynomials determined by the commutation† function $[x, y]$—and to deduce in §37 various corollaries from this fact.

F. Schur [16] first showed that this series was valid in all groups under canonical parameters. Campbell [3] and Hausdorff [8] have since obtained it by other methods,‡ and so we shall call it the "*SCH*-series."

The present exposition is preferable on three grounds to those cited. It applies to infinite-dimensional groups. It paraphrases identities on pure group products [§§38–40], and does not require Taylor's series or manipulations with matrix polynomials (which are unnatural in non-linear groups). And most important, it generalizes easily to yield similar series expressing the definite (product) integrals over fixed time intervals of variable linear combinations of infinitesimal transformations, in a form which (like the *SCH*-series§) is independent of the group which they generate.

In §§38–40, the paraphrases in terms of identities on group products, of the *SCH*-series and other identities in the theory of continuous groups, are developed. They are not a part of the technical argument—unlike the paraphrases of the identities of Lie-Jacobi, which are actually used in proving the latter. They have been included because they correlate the theories of discrete groups and continuous groups in a way essential to the full understanding of either.

---

† Expressions (x, y) or $[x, y]$ will be called "simple" commutators and brackets, respectively; the commutator (ϕ, ψ) of any two commutators ϕ and ψ of "lengths" $w(\phi)$ and $w(\psi)$—where for uniformity individual letters are regarded as commutators of length one—will be called a "complex" commutator of "length" $w(\phi)+w(\psi)$. Similarly with complex brackets $[\phi, \psi]$ of "length" $w(\phi)+w(\psi)$.

‡ Schur starts with the obvious identity $f(x, (\lambda+\delta)y) = f(x, \lambda y) \circ \delta y$, determines $d/d\lambda\{f(x, \lambda y)\}$ from this, and integrates the resulting differential equation. Campbell and Hausdorff develop the series by setting $e^x e^y = e^{f(x,y)}$, and use the algebra of matrices to solve for $f(x, y) = \log[1+(e^x e^y - 1)]$—thus introducing an extraneous operation of addition.

§ The *SCH*-series is the case where x operates first for a unit of time, followed by y operating for a unit of time.

34. **Product-equivalence of paths.** Let G be any analytical group under canonical parameters. Then

(34α) *The problem of determining* $x \circ y = f(x, y)$ *is equivalent to that of determining, given two short paths* P *and* Q, *whether or not* $t(P^*) = t(Q^*)$.

(We shall express the relation $t(P^*) = t(Q^*)$ by writing $P \sim Q$, and saying P is *product-equivalent* to Q. By (14θ), $P \approx Q$ implies $P \sim Q$.)

Proof. Since $x \circ y = t((P_x \oplus P_y)^*)$ under canonical parameters, we have found the $z = f(x, y)$ when‡ we have found the $P_z \sim P_x \oplus P_y$. While conversely, $t(Q^*)$ is approximated arbitrarily closely and hence determined by the $t(Q_\pi^*) = t(Q_1) \circ \cdots \circ t(Q_r)$ for the different partitions π of Q—and the $t(Q_1) \circ \cdots \circ t(Q_r)$ are determined by Q and the function of composition.

From (34α) and the known existence of an SCH-series expressing $f(x, y)$ in terms of the commutation function, we certainly can infer that $t(Q^*)$ is determined by Q and the commutation function in a way valid in all groups G under canonical parameters. But it by no means gives us explicit series for Q^* (except when $Q = P_x \oplus P_y$)—and it is such series that we shall finally obtain, getting the SCH-series as a special case (cf. §36).

Our first step will be to determine, given Q, all the $P \sim Q$. To this end we prove

(34β) *Let* P *and* Q *be any admissible paths with domain* $[0, \Lambda]$. *Then* $P \sim Q$ *if and only if some* U: $u(\lambda)$ *exists, such that* $u(0) = u(\Lambda) = 0$ *and*

$$| \delta p - [u^{-1}(\lambda) \circ \delta q \circ u(\lambda) + \delta u\dagger] | \leqq o(| \Delta Q | + | \Delta U |).$$

Proof. Suppose $P \sim Q$, and write $p^*(\lambda) = q^*(\lambda) \circ u(\lambda)$. Since $p^*(0) = 0 = q^*(0)$ and $p^*(\Lambda) = t(P^*) = t(Q^*) = q^*(\Lambda)$, $u(0) = u(\Lambda) = 0$. Define $R = P^*$, so that $P = R\dagger$. Clearly if Δ: $[\lambda, \mu]$ is any interval, then by (14η)

$$\begin{aligned} t((\Delta P)^*) &= t((\Delta R\dagger)^*) = r^{-1}(\lambda) \circ r(\mu) \\ &= \{u^{-1}(\lambda) \circ [q^{*-1}(\lambda) \circ q^*(\mu)] \circ u(\lambda)\} \circ \{u^{-1}(\lambda) \circ u(\mu)\} \\ &= \{u^{-1}(\lambda) \circ t((\Delta Q)^*) \circ u(\lambda)\} \circ t((\Delta U\dagger)^*). \end{aligned}$$

But $|\Delta P| = O(|\Delta R|) \leqq O(|\Delta Q^*| + |\Delta U|) \leqq O(|\Delta Q| + |\Delta U|)$; besides $|t((\Delta P)^*) - t(\Delta P)| \leqq o(|\Delta P|)$, and similarly with ΔQ and ΔU (by (14ϵ)) even after the inner automorphism induced by $u(\lambda)$. Moreover by (3) $|x \circ y - (x+y)| = o(|x| + |y|)$; consequently if we write $\delta p = t(\Delta P)$, $\delta q = t(\Delta Q)$ and $\delta u = t(\Delta U)$, we get

$$| \delta p - [u^{-1}(\lambda) \circ \delta q \circ u(\lambda) + \delta u\dagger] | = o(| \Delta Q | + | \Delta U |).$$

‡ We recall the notation P_x for the path $p_x(\lambda) = \lambda x$ defined on $[0, 1]$, and $P_x \oplus P_y$ for the broken line R: $r(\lambda) = \lambda x$ on $[0, 1]$, and $r(\lambda) = x + (\lambda - 1)y$ on $[1, 2]$.

Conversely, suppose that this inequality is satisfied for some $u(\lambda)$ (of bounded variation‡) with $u(0)=u(\Lambda)=0$. Then, when we write $r(\lambda)=q^*(\lambda)\circ u(\lambda)$, obviously $t((R\dagger)^*)=t(R)=t(Q^*)$. Moreover by the argument above, $r\dagger(\lambda)$ satisfies the given inequality. Therefore by the triangle inequality,

$$|\delta p-\delta r\dagger| \leqq M(|\Delta Q|+|\Delta U|)\cdot(|\Delta Q|+|\Delta U|).$$

Hence if π is any partition of $[0,\lambda]$, writing $\|\pi\|$ for sup $(|\Delta Q|+|\Delta U|)$, and summing inequalities, we get

$$\begin{aligned}|p(\lambda)-r\dagger(\lambda)| &= |[p(\lambda)-p(0)]-[r\dagger(\lambda)-r\dagger(0)]|\\ &\leqq M(\|\pi\|)\cdot(|Q|+|U|),\end{aligned}$$

whence in the limit $p(\lambda)\equiv r\dagger(\lambda)$.

We can rewrite (34β) perhaps more suggestively in the notation of differentials, as

$$dp = u^{-1}(\lambda)\circ dq\circ u(\lambda)+du\dagger. \tag{34γ}$$

35. **Devices for calculation.** Consider the terms of this formula. By (32α), $u^{-1}\circ dq\circ u$ can be calculated explicitly from U, Q and the commutation function.

Again, although we have not shown how to calculate $U\dagger$ from U explicitly§ by using the commutation function, we can now do so in case U is *unidimensional.*

(A path U: $u(\lambda)$ will be called "unidimensional" if and only if it is confined to a straight line—i.e., if and only if for some u_0, $u(\lambda)=\alpha(\lambda)u_0$. If U is unidimensional, then by Theorem 5, $U_\pi^*=U_\pi\dagger=U$ identically, whence in the limit $U^*=U\dagger=U$. By a "unidimensional alteration" of any path Q with domain $[0,\Lambda]$, will be meant any path $P=R\dagger$ determined from an R: $r(\lambda)=q^*(\lambda)\circ[\alpha(\lambda)u_0]$ for which $\alpha(0)=\alpha(\Lambda)=0$. In this case, clearly $P\sim Q$ and furthermore by (34γ),

$$dp = u^{-1}(\lambda)\circ dq\circ u(\lambda)+du. \tag{35α}$$

And so P is determined by Q, $u(\lambda)$ and the commutation function.)

Since (32α) gives an infinite series in any case, the fact that only unidimensional alterations can be computed explicitly suggests the following procedure: decomposing a given Q into undimensional constituents, altering these one at a time, and justifying the computations by proving general properties of paths represented by infinite series. This we shall do, first proving

‡ I.e., such that the curve U: $u(\lambda)$ is rectifiable.

§ N.B.: $U\dagger$ differs from U by $M(|U|)$—and hence one can deform a given Q little by little into any desired shape (e.g., a straight ray), whose final position will be *determined* by Q and the commutation function. But its *calculation* involves integrating a (highly involved) differential equation.

(35β) *Let* $u_1(\lambda)$, $u_2(\lambda)$, $u_3(\lambda)$, $\cdots$ *be any twice differentiable functions with domain* $[0, \Lambda]$ *and values in a Banach space. Suppose that the* sup $|u_k| \equiv \sup_{0\leqq\lambda\leqq\Lambda}|u_k(\lambda)|$, *the* sup $|u_k'|$, *and the* sup $|u_k''|$ *all form convergent series. If* $[\lambda, \lambda+d\lambda]$ *is any subinterval of* $[0, \Lambda]$, $u(\lambda)$ *denotes* $\sum_{k=1}^{\infty} u_k(\lambda)$, *and* δu_k *denotes* $u_k(\lambda+d\lambda)-u_k(\lambda)$, *then*

$$\left| \delta u - d\lambda\left[\sum_{k=1}^{\infty} u_k'(\lambda)\right]\right| \leqq O(|d\lambda|^2) \leqq o(|d\lambda|).$$

Remark. It is a corollary that u is differentiable and has $\sum_{k=1}^{\infty} u_k'(\lambda)$ for derivative.

Proof. Since by the comparison test, all the series involved converge absolutely (and uniformly!), and the terms of absolutely convergent series can be permuted, clearly $\delta u = \sum_{k=1}^{\infty}\delta u_k$. Moreover for every k,

$$|\delta u_k - u_k'(\lambda)d\lambda| \leqq \tfrac{1}{2}[\sup |u_k''|]\cdot d\lambda^2.$$

Summing, we get by the triangle law

$$\left| \delta u - d\lambda\left[\sum_{k=1}^{n} u_k'(\lambda)\right]\right| \leqq d\lambda \sum_{n+1}^{\infty} \sup |u_k'(\lambda)| + d\lambda^2\cdot \sum_{k=1}^{n} \sup |u_k''|.$$

When we pass to the limit, this becomes

$$\left| \delta u - d\lambda\left[\sum_{k=1}^{\infty} u_k'(\lambda)\right]\right| \leqq d\lambda^2\cdot \sum_{k=1}^{\infty} \sup |u_k''| \leqq O(|d\lambda|^2).$$

It will be convenient to signify that the hypotheses of (35β) are satisfied by writing

$$U = U_1 + U_2 + U_3 + \cdots = \sum_{k=1}^{\infty} U_k.$$

We shall now get a path $R\dagger \sim P_x \oplus P_y$, from which we shall be able to calculate $f(x, y)$ by using an algorithm applicable to all analytical combinations of unidimensional paths. (The analyticity of $P_x \oplus P_y$ is concealed.)

THEOREM 14. *Let* R: $r(\lambda) = \lambda x \circ \lambda y$ *be defined on* $[0, 1]$. *Then* $P_x \oplus P_y \sim R\dagger$. *And* (*assuming* $|[x, y]| \leqq |x|\cdot|y|$ *by* §30) *if* $|x|+|y| < 1/10$, *then*

$$r\dagger(\lambda) = \lambda y + \lambda x + \frac{\lambda^2}{2!}[x, y] + \frac{\lambda^3}{3!}[[x, y], y] + \cdots \equiv s(\lambda).$$

Proof. It is obvious from identities established in §14 that

$$t((P_x \oplus P_y)^*) = t(P_x) \circ t(P_y) = x \circ y = t(R) = t((R\dagger)^*).$$

The proof is complete if we can show that $|\delta r\dagger - \delta s| \leq o(|d\lambda|)$. For if this is so, then the upper right-derivative of $|r\dagger(\lambda) - s(\lambda)|$ is zero everywhere, and so $r\dagger(\lambda) = s(\lambda)$. But by ($35\beta$), δs differs from $d\lambda\{y + x + \lambda[x, y] + \cdots\}$ by $o(|d\lambda|)$—and this is by (32α) $d\lambda\{y + y^{-\lambda} \circ x \circ y^{\lambda}\}$. Again, by (3) $d\lambda\{y + y^{-\lambda} \circ x \circ y^{\lambda}\}$ differs from

$$t((\delta R\dagger)^*) = (x^{\lambda} \circ y^{\lambda})^{-1} \circ (x^{\lambda + d\lambda} \circ y^{\lambda + d\lambda})$$
$$= (y^{-\lambda} \circ x^{d\lambda} \circ y^{\lambda}) \circ y^{d\lambda}$$

by $M(|d\lambda| \cdot |x|) \cdot |d\lambda| \cdot |y| \leq o(|d\lambda|)$. And by ($14\epsilon$) we have $|t((\delta R\dagger)^*) - \delta r\dagger| \leq o(|d\lambda|)$—completing the chain of links of length $o(|d\lambda|)$ between δs and δr, and hence the proof.

36. **Evaluation of regular paths.** We can now find $f(x, y) = t(R) = t((R\dagger)^*)$ by a process which enables one to find series expressing $t(P^*)$ for *any* short path P which is "regular" in a sense defined below.

Accordingly, let G be any analytical group under canonical parameters, in which a scale of length has been so chosen that $[x, y] \leq |x| \cdot |y|$. Let P be any path in G which can be written

$$P = P_1 + P_2 + P_3 + \cdots \qquad \text{(in the sense of } (35\beta)),$$
$$P_i\colon\ p_i(\lambda) = \rho_i(\lambda) \cdot b_i \qquad (0 \leq \lambda \leq 1),$$

where (1) the $\rho_i(\lambda)$ are analytical scalar functions with $\sum_{i=1}^{\infty} \int |d\rho_i| < 1/10$, (2) the b_i are brackets in elements $x_1, \cdots, x_r$ arranged in order of increasing length, and containing with any b_i and b_j, also $[b_i, b_j] = b_{f(i,j)}$. Such a path will be called *regular*.

Remark 1. By inserting dummy terms $0 \cdot b_i$, one can make *any* sum of scalar multiples of brackets in $x_1, \cdots, x_r$ satisfy (2) simply because the number of different brackets of any preassigned length w in $x_1, \cdots, x_r$, is finite.

Remark 2. If $|x| + |y|$ is small enough, then the $r\dagger(\lambda)$ of Theorem 14 is regular.

Remark 3. Since $|[x, y]| \leq |x| \cdot |y|$, $|b_i| \leq 1$ identically if $|x_1| \leq 1, \cdots, |x_r| \leq 1$.

THEOREM 15. *Let P be any regular path. Then $t(P^*)$ is $\sum_{i=1}^{\infty} \gamma_i b_i$, where each γ_i can be calculated from $\rho_1(\lambda), \cdots, \rho_i(\lambda)$ in a finite number of rational operations, integrations, and differentiations. The calculations are independent of G.*

Outline of proof. We shall construct paths $P'' \sim P' = P$, $P''' \sim P''$, $P^{\text{iv}} \sim P'''$, $\cdots$ by successive unidimensional alterations. Each $P^{\nu+1}$ will be "regular" in the same sense that P is, except that $1/10$ may be replaced by some other constant $< 1/5$. Moreover the $\rho_i^{\nu+1}(\lambda)$ for $i \leq \nu$ will be of the form $\lambda\gamma_i$—where γ_i is independent of ν—and the $\rho_i^{\nu+1}(\lambda)$ for $i>\nu$ will be increasingly negligible—whence $t(P^*)=\sum_{i=1}^{\infty}\gamma_i\cdot b_i$.

Definition of $P^{\nu+1}$ by induction. If one sets

$$u_\nu(\lambda) = [\lambda\rho_\nu^\nu(1) - \rho_\nu^\nu(\lambda)]\cdot b_\nu \equiv \beta_\nu(\lambda)\cdot b_\nu$$

and can obtain a $P^{\nu+1}=\sum_{i=1}^{\infty}\rho_i^{\nu+1}(\lambda)\cdot b_i$ from P^ν through unidimensional alteration by $u_\nu(\lambda)$, then assuming the term-by-term differentiability of all series, by (32α) and (35α), we obtain heuristically

$$(*)\qquad dp^{\nu+1} = dp^{\nu} + d\beta_\nu(\lambda)b_\nu + \sum_{j,k=1}^{\infty} d\rho_j^{\nu}\cdot[\beta_\nu(\lambda)]^k\cdot b_{i(j,k)}\cdot\frac{1}{k!},$$

where $b_{i(j,1)}\equiv[b_j, b_\nu]$ and $b_{i(j,k)}\equiv[b_{i(j,k-1)}, b_\nu]$. But clearly $i(j, k)=i$ has in no case an infinity of solutions (j, k). Hence we can certainly *define*

$$\rho_i^{\nu+1}(\lambda) = \rho_i^{\nu}(\lambda) + \sum_{i(j,k)=i}\int\frac{1}{k!}[\beta_\nu(\lambda)]^k\cdot d\rho_j^{\nu}$$

with the assurance of obtaining analytical $\rho_i^{\nu+1}(\lambda)$—and using only rational operations, integration, and differentiation.

Actual proof. Let us do this. Then—since the length of no $b_{i(j,k)}$ exceeds that of b_ν—certainly by construction $\rho_\nu^{\nu+1}(\lambda)=\lambda\rho_\nu^\nu(1)=\lambda\gamma_\nu$, and for $i<\nu$, $\rho_i^{\nu+1}(\lambda)=\rho_i^\nu(\lambda)=\lambda\gamma_i$ by induction. Furthermore

(36α) *The series* (*) *converge in the sense of* (35β). *Consequently* (*collecting terms*) $P^{\nu+1}=\sum_{i=1}^{\infty}P_i^{\nu+1}$ *in the same sense. Moreover* $\sum_{i=1}^{\infty}\int|d\rho_i^{\nu+1}|<1/5$.

Remark. They even converge absolutely if we replace each bracket by the product of the absolute values of its entries.

Proof. If $\sigma(\lambda)$, $\beta(\lambda)$ and $\rho(\lambda)$ are *any* real analytical functions, then certainly

$$\left\{\begin{aligned}
\sup\ |\sigma| &\leqq \int|d\sigma| = \int|\beta|^k\cdot|d\rho|\\
&\leqq \left[\int|d\beta|\right]^k\cdot\left[\int|d\rho|\right],\\
\sup\ |\sigma'| &\leqq \left[\int|d\beta|\right]^k\cdot\sup\ |\rho'|,\\
\sup\ |\sigma''| &\leqq k\cdot\left[\int|d\beta|\right]^{k-1}\cdot\sup\ |\rho'| + \left[\int|d\beta|\right]^k\cdot\sup\ |\rho''|
\end{aligned}\right.$$

(differentiation is indicated by superscribing primes). Hence by induction on ν,—since $\sum_{k=1}^{\infty}\lambda^k=\lambda/(1-\lambda)$ and $\sum_{k=1}^{\infty}k\lambda^k<+\infty$ if $|\lambda|<1$—the series (*) converges in the sense of (35β). Moreover (since grouping terms never increases sums of absolute values) for the same reasons $\sum_{i=1}^{\infty}\int|d\rho_i^{\nu+1}|$ (which bounds $\sum_{i=1}^{\infty}\sup|\rho_i^{\nu+1}|$) does not exceed the corresponding sum for P^ν by a proportion of more than $\int|d\rho_\nu^\nu|/(1-\int|d\rho_\nu^\nu|)$. And by induction this is at most $5\int|d\rho_\nu^\nu|/4$. Consequently

$$\sum_{i=\nu+1}^{\infty}\int|d\rho_i^{\nu+1}| \leqq \sum_{i=\nu}^{\infty}\int|d\rho_i^{\nu}| - \int|d\rho_\nu^{\nu}| + \frac{1}{5}\left(\frac{5}{4}\int|d\rho_\nu^{\nu}|\right)$$
$$\leqq \sum_{i=\nu}^{\infty}\int|d\rho_i^{\nu}| - \frac{3}{4}\int|d\rho_\nu^{\nu}|$$

and

$$\sum_{i=1}^{\nu}\int|d\rho_i^{\nu+1}| \leqq \sum_{i=1}^{\nu-1}\int|d\rho_i^{\nu}| + \int|d\rho_\nu^{\nu}|.$$

But four-thirds of the first sum, plus the second sum, is non-increasing as $\nu\uparrow\infty$—whence the second sum is always bounded by $\frac{4}{3}\cdot\frac{1}{10}<\frac{1}{5}$, and the first tends to zero.

This proves (36α). Hence (regrouping the terms of (*) through (32α)), by (32α) and (35α), $P^{\nu+1}\sim P^\nu\sim P$. And since by inequalities just proved, $|t((P^{\nu+1})^*)-\sum_{i=1}^{\nu}\gamma_i b_i|$ tends to zero as ν increases, $t(P^*)\equiv t((P^\nu)^*) =\sum_{i=1}^{\infty}\gamma_i\cdot b_i$.

This completes the proof of Theorem 15.

37. **Corollaries of Theorem** 15. Theorem 15 has several immediate corollaries of primary theoretical importance. We shall list some of these now.

COROLLARY 15.1. *One can write $f(x, y)$ as the sum of an infinite series of scalar multiples of brackets of x and y arranged in order of increasing weight; each coefficient can be computed after a finite number of rational operations, and are rational numbers.*

Proof. In Theorem 14, $r\dagger(\lambda)$ is (cf. Remark 2 above) a regular path whose $\rho_i(\lambda)$ are polynomials (of degree at most the length $w(b_i)$ of b_i) with rational numbers as coefficients. These properties are preserved under the rational operations, differentiations, and integrations performed above—any polynomial can be differentiated or integrated by rational operations on its coefficients.

(The reader will find it instructive to compute the terms of degrees two and three.)

Caution. Because of the linear interdependence (due to the identities of Lie-Jacobi) between the brackets of length w, the series of Theorem 15 is not unique; its computation depends on the arrangement of the brackets of each length w.

COROLLARY 15.2. *The function $x \circ y = f(x, y)$ of composition of any analytical group G under canonical parameters is analytical.*

Proof. By §24, brackets are polynomial functions.

COROLLARY 15.3. *If the Lie albebra of G is "w-nilpotent" (that is, if all brackets of length w vanish), then $f(x, y)$ is a polynomial of degree at most r.*

COROLLARY 15.4. *Two analytical groups having topologically isomorphic Lie algebras are locally topologically isomorphic (and so analytically isomorphic).*

Proof. Within some neighborhood of the identity, and under canonical parameters, they have the same function of composition.

COROLLARY 15.5. *Let L be the Lie algebra of any analytical group G, and let S be any closed subalgebra of L. Then the elements in S near the origin are an analytical subgroup nucleus.*

Proof. They are a subgroup (by Corollary 15.1), satisfy (1), (2), (2′), and are a complete linear subspace of L.

From Corollary 5 and (31α), we get

COROLLARY 15.6. *The analytical subgroup nuclei of any analytical group G under canonical parameters, are the closed subalgebras of its metric Lie algebra.*

COROLLARY 15.7. *A locally compact analytical group is a Lie group in the usual sense.*

Proof. Any locally compact Banach space is finite-dimensional by [1], p. 84, and the function of composition is by Corollary 2 analytical under canonical parameters.

COROLLARY 15.8. *A commutative analytical group nucleus under canonical parameters is a neighborhood of the origin in a Banach space.*

38. **Digression: paths and group-products.** Since to assert $x_1 \circ \cdots \circ x_n = y_1 \circ \cdots \circ y_n$ is to assert

$$P_{x_1} \oplus \cdots \oplus P_{x_n} \sim P_{y_1} \oplus \cdots \oplus P_{y_n},$$

and since every admissible path can be approximated arbitrarily closely by broken lines, one would expect product-equivalences‡ $P \sim Q$ between images of an interval $[0, \Lambda]$ to correspond to algebraic identities between group products. We shall sketch in §38 some crude examples of such correspondences.

The identity $xy = yx(x, y)$ shows that if Q is any broken line, one can replace any two segments of Q by the opposite sides of the parallelogram which

‡ We recall the notation $P \sim Q$ meaning $t(P^*) = t(Q^*)$.

they determine, without altering $t(Q^*)$, provided a small deviation (x, y) is inserted.

The graphical principle (essential in the classical proofs of Green's and Stokes' Theorems) that any path-deformation can be split up into elementary deformations across parallelograms, is analogous to the algebraic principle that any permutation of terms in a sequence is the product of transpositions.

The derivation in §34, given a path Q, of paths $P \sim Q$ by choosing $v^*(0) = v^*(\Lambda) = 0$ and setting

$$dp = v^{*-1}(\lambda) \circ dq \circ v^*(\lambda) + dv$$

corresponds to taking a product $x_1 \circ \cdots \circ x_n$ and a second product $u_1 \circ \cdots \circ u_n = e$, defining $v_k^* = u_1 \circ \cdots \circ u_k$ and proving by induction $\prod_{i=1}^{k} [(v_{i-1}^{*-1} \circ x_i \circ v_{i-1}^*) \circ u_i] = x_1 \circ \cdots \circ x_k \circ v_k^*$, and thus concluding that

$$\prod_{k=1}^{n} x_k = \prod_{k=1}^{n} [(v_{k-1}^{*-1} \circ x_k \circ v_{k-1}^*) \circ u_k]. \tag{38α}$$

39. **Digression: the Rearrangeability Principle.** In correlating the argument of §§34–36 with formal identities on group products, let us begin by recalling a recent result of P. Hall ([7], Theorem 3.1), namely

$$(xy)^n \equiv x^n y^n z_1^{\phi_1(n)} \cdots z_t^{\phi_t(n)} \pmod{H_w}, \tag{39α}$$

where the z_k are complex commutators in x and y of lengths $< w$ arranged in order of increasing length, the exponents ϕ_k are polynomials of degree $w(z_k)$, and H_w is the normal subgroup whose elements are the products of commutators of lengths $\geqq w$.

That there exist (not necessarily polynomial) *functions* $\phi_k(n)$ such that (39α) is satisfied, is very easy to show. For since $uv = vu(u, v)$, one can transpose any two adjacent terms in any product involving x, y, and their commutators, by inserting commutators of lengths greater than the length of either transposed term. Hence one can first shift all the occurrences of x in such a product to the extreme left, then all the occurrences of y to positions just to the right of these, and similarly with $z_1, \cdots, z_t$.

This method, combined with the rule that any permutation can be accomplished by successive transpositions, obviously yields a general

Rearrangeability Principle. If one is given *any* product ψ involving elements $x_1, \cdots, x_n$ and their commutators, *any* integer w, and *any* ordering ρ of the x_k and their commutators of weights $< w$, then ψ is congruent modulo commutators of weights $\geqq w$ to a product of powers of the x_k and their commutators arranged in the sequence ρ.

More than this, one can *distribute* their occurrences according to any preassigned distribution function.

These principles are the key to the algebraic situation. Using them, one can show for instance that if $m=n^w$, then

$$(39\beta) \qquad x^m y^m \equiv \prod_{k=1}^{n} v_k \pmod{H_w},$$

where each v_k is of the form $x^{m/n}y^{m/n}z_1^{\zeta_1(m;k)} \cdots z_t^{\zeta_t(m;k)}$ and $|\zeta_h(m;i)-\zeta_h(m;j)| \leqq 1$, which means that the v_k are all nearly equal.

Proof. Write $x^m y^m = (x^{m/n})(x^{m/n}) \cdots (x^{m/n})(x^{m/n}y^m)$. Then transpose the occurrences of y (inserting commutators, of course) until you obtain the identity

$$x^m y^m = (x^{m/n}y^{m/n}u_1)(x^{m/n}y^{m/n}u_2) \cdots (x^{m/n}y^{m/n}u_3),$$

where the u_k are congruent (mod H_w) to products of the commutators z_k of lengths $<w$. Proceed by induction, dividing the occurrences of each z_h $(h=1, \cdots, t)$ into n nearly equal lots, and you will get (39β).

Now suppose x and y are elements of a continuous group under canonical parameters. Write $x^m=\bar{x}$ and $y^m=\bar{y}$; since the v_k are nearly equal, if we know by (27β) that the elements of H_w are relatively small, we see that $|f(\bar{x}, \bar{y})-nv_1|$ is small, where for n large v_1 is nearly determined by x, y, and the commutation function $[x, y]$.

40. **Digression (cont.): analyticity and other remarks.** We do not have to go far beyond the same principles to see from an algebraic standpoint even why an *SCH*-series exists, in the way that it does.

To see this, observe that for fixed $x^m=\bar{x}$, $y^m=\bar{y}$ and very large n, since x and y are correspondingly small, (1) products are nearly sums, and (2) commutators are nearly equal to the corresponding brackets. Hence if b_h denotes the bracket in $\bar{x}$ and $\bar{y}$ corresponding to the commutator in x and y denoted by z_h, and $\lambda_h=n\zeta_h(m;1)/n^{w(z_h)}$, then the smallness of $|f(\bar{x}, \bar{y})-nv_1|$ implies the smallness of $|f(\bar{x}, \bar{y})-\{\bar{x}+\bar{y}+\sum_{h=1}^{t}\lambda_h b_h\}|$. This gives one the first $(t+2)$ terms of an *SCH*-series, approximately.

Actually, the λ_h are polynomials whose dominating terms are independent of m, although the reasons for this are number-theoretical and not at all trivial, and the calculation of the dominating terms is not even impossibly laborious.

Similar reasoning yields an algebraic paraphrase of Theorem 15. Take any path X: $x(\lambda)=\sum_{i=1}^{r}\rho_i(\lambda)\cdot x_i$. Divide X into $m=n^w$ equal parts, set

$\mu_i^k = \rho_i(k,n) - \rho_i([k-1]/n)$ and $x = y^m$, and obtain through the Rearrangeability Principle (as with (39β)), an identity‡

$$(40\alpha)\qquad \prod_{k=1}^{n}\left(\prod_{i=1}^{r} y_i^{[m\mu_i^k]}\right) \equiv \prod_{k=1}^{n} v_k \pmod{H_w},$$

where the v_k are products $\prod_{h=1}^{t} z_h^{\xi_h(m;k)}$ of nearly equal powers of commutators z_h in the y_i. But replacing each commutator z_h of length w_h in the y_i by $(m^{-w_h})\cdot b_h$, where b_h denotes the bracket in the x_i corresponding to z_h—the substitution is nearly one of equals for equals—and setting $\gamma_h(m) = nm^{-w_h}\zeta_h(m; 1)$, (40$\alpha$) becomes

$$(40\alpha)\qquad t(X^*) \text{ is nearly } \sum_{h=1}^{t} \gamma_h(m)\cdot b_h,$$

the calculation of $\gamma_h(m)$ being the same for all groups.

41. **Every metric Lie algebra belongs to a group.** We can now prove by considerations of convergence, that

THEOREM 16. *Every metric Lie algebra L is the Lie algebra of an analytical group nucleus.*

Proof. Define group products $x \circ y = f(x, y)$ in L through the SCH-series. There are three points to establish: the convergence of the series, the validity of the inequalities (2′)–(2″), and the associative law $f(f(x, y), z) = f(x, f(y, z))$.

By Remark 2 of §30, we can assume $|[x, y]| \leqq |x| \cdot |y|$. Then by the proof of Theorem 15 (cf. the remark after (36α)), if we substitute for each bracket in the SCH-series, the product of the absolute values of its entries, and if these are $<1/10$, then the sum of the absolute values of the resulting series is bounded by $2(|x| + |y|)$. The *convergence* of $f(x, y)$ provided $|x| + |y| < 1/10$ is a weak corollary of this.

Again, expanding $[f(x, a) - f(x, b)] - (a - b)$ in SCH-series, we have by Theorem 14 after cancellation and pairing off of corresponding terms, scalar multiples of differences such as

$$\begin{aligned}\Phi &= [x, [a, x], a] - [x, [b, x], b] \\ &= [x, [a - b, x], a] + [x, [b, x], a - b]\end{aligned}$$

whose magnitude is bounded by $|x| \cdot |a - b|$ times the number n_i of entries in the bracket, times what we would get if we replaced every bracket by the product of the absolute values of all but one of its entries. But the sum

‡ $[m\mu_i^k]$ denotes conventionally the integral part of $m\mu_i^k$.

of the products of these last two factors still converges absolutely provided $|a|+|b|+|x|<1/20$, so that $n_i(1/20)^{n_i-1}<10(1/10)^{n_i}$. Hence

$$|(xa - xb)| - |(a - b)| \leqq K\cdot |x| \cdot |a - b|$$

within this region. This implies (2′); (2″) follows by symmetry. It remains to prove the associative law.

Here we do the obvious thing: substitute the *SCH*-series for u in the *SCH*-series for $f(u, z)$, and likewise the *SCH*-series for $f(y, z)$ for v in the *SCH*-series for $f(x, v)$, and expand in both cases by the distributive law. We will get two series of monomial brackets in x, y, z, with possible repetitions. If they are absolutely convergent, then by the continuity implied in (2′)–(2″) they will converge to $f((x, y), z)$ and $f(x, f(y, z))$ respectively. We shall next prove that they are absolutely convergent.

If $|x|+|y|+|z|<1/80$, and we replace each bracket in the series for $f(f(x, y), z)$ by the product of the absolute values of its entries, then the sum of the absolute values of what we get is by the distributive law (on scalars) what we *would* get if we replaced brackets by products in the *SCH*-series for $f(u, z)$, replaced z by $|z|$, and u by the sum of the absolute values of the terms in the *SCH*-series for $f(x, y)$. And since both of these are $<1/40$, the series for $f(f(x, y), z)$ is absolutely convergent. The absolute convergence of the series for $f(x, f(y, z))$ follows by symmetry.

Hence to prove that $f(f(x, y), z)=f(x, f(y, z))$ we need only show that irrespective of n, the sum of the terms of length $\leqq n$ is the same for the two series. The demonstration of this essentially algebraic fact completes the proof.

Demonstration. Form the multiplicative group of all non-commutative polynomials $I+U=I+\lambda_1X+\lambda_2Y+\lambda_3Z+\cdots$ in X, Y, Z, ignoring terms of degree $>n$. This is a (4^n-1)-parameter Lie group, in which $(I+U)^{-1}=I-U+U^2-\cdots+(-1)^nU^n$. Since the group is analytical, the functions $f(f(X, Y), Z)$ and $f(X, f(Y, Z))$ are identically equal near $X=Y=Z=0$, and hence formally equal. Moreover as in all linear groups, $[U, V]=VU-UV$. But by Theorem 3 of the author's *Representability of Lie algebras and Lie groups by matrices*, Annals of Mathematics, vol. 38 (1937), pp. 526–532, any identity between alternants $VU-UV$ follows formally from the identities of Lie-Jacobi. Hence the equality between the sums of the terms of degree $\leqq n$ in the two series follows formally from the identities of Lie-Jacobi (which we assumed at the beginning).

Bibliography

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.

2. G. Birkhoff, *Integration of functions with values in a Banach space*, these Transactions, vol. 38 (1935), pp. 357–378.

3. J. E. Campbell, *On a law of combination of operators*, Proceedings of the London Mathematical Society, vol. 28 (1897), pp. 381–390.

4. E. Cartan, *Notice sur les Travaux Scientifiques de M. Élie Cartan*, Chap. VIII, Paris, 1931.

5. E. Cartan, *Les groupes continus et l'analysis situs*, Mémorials des Sciences Mathématique, Fasc. 42, Paris, 1930.

6. J. Delsarte, *Les groupes de transformations linéaires dans l'espace de Hilbert*, Mémorials des Sciences Mathématiques, Fasc. 57, Paris, 1932.

7. P. Hall, *A contribution to the theory of groups of prime-power orders*, Proceedings of the London Mathematical Society, vol. 36 (1933), pp. 29–95.

8. F. Hausdorff, *Die symbolische Exponentialformel in der Gruppentheorie*, Leipzig Berichte, vol. 58 (1906), pp. 19–48.

9. S. Lie, *Transformationsgruppen*, Leipzig, 1888.

10. S. Lie, *Unendlich continuirliche Gruppen*, Abhandlungen, Sachsische Gesellschaft der Wissenschaften, vol. 21 (1895), pp. 43–150.

11. W. Mayer and T. Y. Thomas, *Foundations of the theory of continuous groups*, Annals of Mathematics, vol. 36 (1935), pp. 770–782.

12. J. von Neumann, *Gruppen linearen Transformationen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 3–42.

13. S. Saks, *Théorie de l'Intégrale*, Warsaw, 1933.

14. L. Schlesinger, *Neue Grundlagen für eine Infinitesimalkalkul der Matrizen*, Mathematische Zeitschrift, vol. 33 (1931), pp. 33–61.

15. O. Schreier, *Abstrakte kontinuierliche Gruppen*, Hamburger Abhandlungen, vol. 4 (1926), pp. 15–32.

16. F. Schur, *Neue Begründung der Theorie der endlichen Transformations gruppen*, Mathematische Annalen, vol. 35 (1889), pp. 161–197.

17. V. Volterra, *Sulle equazione differenziali lineari*, Rendiconti dei Lincei, vol. 3 (1887), pp. 391–396.

Harvard University,
Cambridge, Mass.