THE REVERSIBILITY OF A DIFFERENTIABLE MAPPING

F. V. Atkinson

(received March 10, 1961)

1. Introduction. Given n functions of n variables, in the real domain, by the equations

$$(1) \qquad y_r = f_r(x_1, \ldots, x_n), \qquad r = 1, \ldots, n,$$

we have in various contexts to consider whether the equations are soluble for the x_r when the y_r are given. Such questions receive fairly complete answers in complex variable theory; a complex variable relation $w = f(z)$ is of course brought under the heading of the real equations (1) by setting $w = y_1 + iy_2$, $z = x_1 + ix_2$. For example, if $f(z)$ is a polynomial the fundamental theorem of algebra asserts that the equations are soluble, though not in general uniquely. Again, a basic theorem on conformal mapping gives conditions under which the equations are uniquely soluble, to the effect that a (1, 1) mapping of the boundaries of domain and range implies a (1, 1) mapping of the interiors.

The other main context for this question is the "change of variables" in multiple integration. Here it is desirable to be able to say that the x_r are uniquely fixed by the y_r. A standard result covers this point, but in a local sense only; specialized to the form (1), the implicit function theorem asserts that if the derivatives are continuous and the Jacobian not zero, and if (1) holds for a particular set of the x_r, y_r, then corresponding to a slightly perturbed set y_r' there will be a unique perturbed set x_r' so that $y_r' = f_r(x_1', \ldots, x_n')$, $r = 1, \ldots$ n. It should be emphasized that the non-vanishing of the Jacobian does not ensure the unique solubility of the equations (1) in the large.

Canad. Math. Bull. vol. 4, no. 2, May 1961.

In texts on real analysis it is only rarely that the treatment of the implicit function theorem proceeds beyond the local version just mentioned; a notable exception is however the text [2]. While definitive results are obtained in complex variable theory, the methods are often of a special character.

It may therefore be useful if I give here a unified treatment of these topics, treating the general case (1) in such a manner as to cover some of the more basic complex variable results. We have to deal here with matters which are well-known in complex variable theory in a more special case, and again in topology, to take a more general case, but which are less often dealt with at an intermediate level of generality.

After preliminaries I discuss the solubility of (1) without reference to uniqueness, related topics being the maximum and minimum modulus principles and the "fundamental theorem of algebra". Following some subsidiary results on "analytic" continuation, I pass to the two main criteria for the unique solubility of (1). One of these requires the range of the transformation to be simply-connected; another requires the mapping of the boundary of the domain on to the boundary of the range to be (1, 1) at one point. These two results are similar to those of [2], though there are certain differences in the consideration of the boundaries and of the effect of isolated zeros of the Jacobian. Finally, I consider the special cases of a convex or unbounded range, and an application to the existence of a critical value for a scalar function of several variables.

2. Terminology. Using vectorial notation we write (1) in the form

$$y = f(x), \tag{2}$$

where y, f, x denote n-vectors or points in Euclidean n-space, with $n \geq 2$; much of the reasoning applies also with $n = 1$, and to manifolds of more general types than E_n.

With the usual metric in E_n we write $|x| = \sqrt{\Sigma_1^n x_r^2}$. A neighbourhood of x will be the set $U(x; \varepsilon)$ formed by all $x' \in E_n$ such that $|x' - x| < \varepsilon$, for any $\varepsilon > 0$. A "region" in E_n will be an open connected set; if G is a region, and $x \in G$,

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

then $U(x;\varepsilon) \subset G$ for some $\varepsilon > 0$, and for any two points $x^{(1)}, x^{(2)} \in G$, there is to be an arc $x(t)$, $0 \leq t \leq 1$, lying entirely in G, where $x(t)$ is continuous in t and $x(0) = x^{(1)}$, $x(1) = x^{(2)}$.

The boundary of any set $V \subset E_n$ will be denoted by ∂V, and defined analytically as the set of $x \in E_n$ such that for every $\varepsilon > 0$, $U(x;\varepsilon)$ contains both a point of V and a point not in V, that is to say a point of the complement $E_n - V$. In the case when V is unbounded, we consider ∞ as a point of ∂V; here ∞ denotes a unique point at infinity, the same for all directions, not actually in E_n and so not in V. It would of course be possible to bring ∞ on to the same basis as other points by using analogues of the Riemann sphere or projective space. On the present basis, however, a neighbourhood of ∞ will be the set of $x \in E_n$ such that $|x| > R$, for any $R > 0$.

If the region G is the domain of the vectorial function f, we denote its range by $f(G)$, the set of $f(x)$, $x \in G$. It is desirable to have an interpretation for $f(\partial G)$. If, as often happens, f is defined not only in G but also in ∂G, and is continuous in the closed set $G + \partial G$, we define $f(\partial G)$ as the set of $f(x)$, $x \in \partial G$. Failing this, we consider all possible sequences $x^{(k)}$, $k = 1, 2, \ldots$, of points of G which converge to any finite point of ∂G or tend to ∞, if G is unbounded. The corresponding sequence $f(x^{(k)})$, $k = 1, 2, \ldots$, will have one or more limit-points, including possibly ∞. The set of all limit-points obtained from all such sequences is to constitute the set $f(\partial G)$; this agrees with the previous definition where that was applicable.

3. Differentiability. We denote by dy/dx the n-by-n matrix whose r-th row is $\partial f_r/\partial x_1, \ldots, \partial f_r/\partial x_n$, so that $\det(dy/dx)$ will be the Jacobian of the mapping. We shall say that y is differentiable at the point x if firstly dy/dx exists there and if secondly there holds at x the formula of the total differential, namely

$$|\delta y - (dy/dx)\,\delta x| \leq \varepsilon\, |\delta x|; \tag{3}$$

here $\delta y = f(x + \delta x) - f(x)$, and for every $\varepsilon > 0$ there is to be an $\eta = \eta(x, \varepsilon) > 0$ such that (3) holds if $|\delta x| < \eta$.

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

We shall throughout be concerned with the case in which y is differentiable at each point of its domain. Later on, when it is a question of unique solubility, we restrict y to be uniformly differentiable in various subsets V of its domain. By this we mean that for given $\varepsilon > 0$ in (3), we may find $\eta > 0$ which is the same for all $x \in V$. Both these situations are ensured by the standard requirement that dy/dx exist and be continuous. For further discussion of this aspect at various levels the books ([1], [2]) may be cited.

For the case when y is differentiable and has non-zero Jacobian we have the following, which forms part of the conclusion of the local implicit function theorem.

LEMMA 1. Let y be defined in the region G and be differentiable at $x \in G$, with dy/dx being non-singular there. Then there exist $\sigma > 0$, $M > 0$ such that if $x' \in U(x;\sigma)$, $y = f(x)$, $y' = f(x')$, then

$$(4) \qquad |x' - x| \leq M|y' - y|.$$

In particular, $y' \neq y$ if $x' \neq x$, $x' \in U(x;\sigma)$. Furthermore, x cannot be the limit of a sequence $x^{(k)}$, $k = 1, 2, \ldots$, such that $f(x^{(k)}) = y^{(o)}$ for some fixed $y^{(o)}$, $x^{(k)} \neq x$ for all k.

Writing $\delta x = x' - x$, $\delta y = y' - y$ we have, since dy/dx is non-singular, $|(dy/dx)\,\delta x| \geq M_1|\delta x|$ for some $M_1 > 0$. Taking $\varepsilon = \frac{1}{2}M_1$ in (3) we have, for $|\delta x| < \eta$ and some $\eta > 0$,

$$|\delta y| \geq |(dy/dx)\,\delta x| - \varepsilon\,|\delta x| \geq \tfrac{1}{2}M_1\,|\delta x|,$$

which establishes (4) with $M = \frac{1}{2}M_1$ and $\sigma = \eta$.

As regards the last assertion in the lemma, it follows by continuity that $f(x) = y^{(o)}$. Thus taking $x' = x^{(k)}$, $y' = y^{(o)}$ in (4) we find that $x^{(k)} = x$ for all large k, contrary to hypothesis.

4. Properties of a class of mappings. We consider here a fairly general case in which the Jacobian of the mapping may have an infinite number of zeros, but not a continuum of zeros. For this case we establish a number of more or less equivalent

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

properties, from which the fundamental theorem of algebra will be an easy consequence.

THEOREM 1. In the region G let $y = f(x)$ be differentiable. Let the limit-points of the zeros, if any, of the Jacobian be isolated in G, so that if $x \in G$ is such a limit-point, there is a neighbourhood of x containing no other limit-point of zeros of the Jacobian. Then the following hold:

(i) (preservation of regions) the map of a neighbourhood of any $x \in G$ contains a neighbourhood of $f(x)$,

(ii) (minimum modulus principle) for any fixed $u \in E_n$, the minimum of $|y - u|$ for $x \in G$ cannot be attained at a point of G, unless it vanishes there,

(iii) (maximum modulus principle) for any fixed $u \in E_n$, the maximum of $|y - u|$ cannot be attained at a point of G,

(iv) the boundary of the map is contained by the map of the boundary, i.e. $\partial f(G) \subset f(\partial G)$,

(v) if H is a region containing no point of $f(\partial G)$, and if H contains at least one point of $f(G)$, then $H \subset f(G)$.

We first prove these results for the special case in which $\det(dy/dx)$ has no zeros in G.

Starting with (ii), suppose that $|y - u|$ attains a minimum for some $x \in G$; we have to prove that $y - u = 0$ there. At a minimum, assumed at an interior point of the domain, the gradient must vanish. Denoting by $d\psi/dx$ the gradient of a scalar-valued function ψ, that is to say the n-vector formed by $\partial\psi/\partial x_1, \ldots, \partial\psi/\partial x_n$, an easy calculation shows that

$$d|y - u|^2/dx = 2(dy/dx)(y - u),$$

and since dy/dx is non-singular it follows that $y - u = 0$.

We pass to (i), which forms part of the local implicit function; the following argument is a standard method of proof of this theorem. Take any $x^{(0)} \in G$. By lemma 1, there exists a $\rho > 0$ such that on the sphere $|x - x^{(0)}| = \rho$ we have $|y - y^{(0)}| > \zeta$, for some $\zeta > 0$, where $y^{(0)}$ denotes $f(x^{(0)})$. For any u, subject to $|u - y^{(0)}| < \frac{1}{2}\zeta$, consider the minimum of

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

$|y - u|$ as x ranges over the region formed by the disc $|x - x^{(o)}| < \rho$. On the boundary of this disc, that is to say on the sphere $|x - x^{(o)}| = \rho$, we have $|y - u| \geq |y - y^{(o)}| - |u - y^{(o)}| > \frac{1}{2}\zeta$, while at the centre of the disc $|y - u| = |y^{(o)} - u| < \frac{1}{2}\zeta$. This means that the minimum of $|y - u|$ is attained at a point of the disc, and therefore is zero. Thus the map of $|x - x^{(o)}| < \rho$ includes all u for which $|u - y^{(o)}| < \frac{1}{2}\zeta$, which proves (i).

We have included (iii), the maximum modulus principle, for completeness only. If $f(G)$ includes y, it also includes a neighbourhood of y, and so includes points further than y from the assigned point u.

We next prove (iv). Let $y^* \in \partial f(G)$. There is then a sequence $x^{(k)}$, $k = 1, 2, \ldots$, with $x^{(k)} \to \partial G$, such that $f(x^{(k)}) \to y^*$. By selection of a sub-sequence, we may take it that the sequence $x^{(k)}$ converges to a limit x^*, possibly ∞. In the latter event y^* is in $f(\partial G)$, by definition. Suppose again that x^* is finite. If $x^* \in G$, then by continuity we should have $y^* = f(x^*)$, and by (i) y^* would be an interior point of $f(G)$, contrary to hypothesis. Thus $x^* \in \partial G$, and therefore $y^* \in f(\partial G)$, as was to be proved.

Finally, for this special case of theorem 1, we prove (v). Let $y' \in f(G) \cap H$, and let y'' be any other point of H. We have to prove that $y'' \in f(G)$. Since H is a region there is a continuous arc $y(t)$, $0 \leq t \leq 1$, $y(0) = y'$, $y(1) = y''$. Then $y(t) \in f(G)$ for $t = 0$, and denote by t' the lower bound of t in $[0, 1]$ for which $y(t) \notin f(G)$. If then $t' \leq 1$, we should have that $y(t') \in \partial f(G) \subset f(\partial G)$, whereas we assumed that $H \cap f(\partial G)$ was empty. Thus there is no such t', and the arc including y'', lies entirely in $f(G)$.

We now permit $\det(dy/dx)$ to have zeros, whose limit-points are to be isolated in G. Thus for any $x^{(o)} \in G$, there will be a neighbourhood of $x^{(o)}$ in which the zeros of $\det(dy/dx)$ form at most a denumerable sequence, say $x^{(k)}$, $k = 1, 2, \ldots$, whose only limit-point, if they are infinite in number, is $x^{(o)}$. We assert that there still exist $\rho > 0$, $\zeta > 0$, such that if $|x - x^{(o)}| = \rho$, then $|y - y^{(o)}| > \zeta$, where $y^{(o)} = f(x^{(o)})$. Since y is continuous, being differentiable it is sufficient to show that $y \neq y^{(o)}$ for $|x - x^{(o)}| = \rho$, and some $\rho > 0$. Suppose on

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

the contrary that for every ρ, $0 < \rho < \rho_1$, there was an $x = x(\rho)$ such that $|x - x^{(0)}| = \rho$, $y = y^{(0)}$. For every ρ-interval of the form $0 < \rho' \leq \rho \leq \rho'' < \rho_1$, these $x(\rho)$ would have a limit-point, at which, by lemma 1, we should have $\det(dy/dx) = 0$. Thus every closed annulus of the form $0 < \rho' \leq |x - x^{(0)}| \leq \rho'' < \rho_1$ would contain a zero of $\det(dy/dx)$, whose zeros would accordingly not have isolated limit-points.

Having chosen ρ, ζ accordingly, let us denote by I the set of zeros of $\det(dy/dx)$ lying in $|x - x^{(0)}| < \rho$. We apply (v) of theorem 1, taking in place of G the set G* formed by $|x - x^{(0)}| < \rho$ except for points of I. In place of H we take the set H* formed by points of $|y - y^{(0)}| < \zeta$ except for points of f(I).

To apply (v) we note first that G*, H* are both regions; each is an open disc, from which has been removed at most a denumerable sequence of points, whose only possible limit-point is at the centre. Secondly, H* contains no point of $f(\partial G^*)$. In this case ∂G^* consists of the sphere $|x - x^{(0)}| = \rho$, together with I; $f(\partial G^*)$ consists of the maps of these points, and by construction does not intersect H*. Finally, we have to show that f(G*) contains at any rate one point of H*. Take in fact any $x \in G^*$, lying so close to $x^{(0)}$ that f(x) lies in $|y - y^{(0)}| < \zeta$. Then $f(x) \in H^*$ except when f(x) happens to coincide with a point of f(I); in this event we have only to perturb x slightly to x', and then f(x') will, by lemma 1, be distinct from any point of f(I). Thus $f(G^*) \cap H^*$ is not empty, so that $f(G^*) \supset H^*$. Adding to G*, H* the point-sets I, f(I) we find that the map of a neighbourhood of $x^{(0)}$ contains a neighbourhood of $f(x^{(0)})$.

The remainder of the properties (ii) - (v) then follow for the general case, by essentially the same arguments as for the special case in which $\det(dy/dx)$ does not vanish. This completes the proof of theorem 1.

As an example of the failure of property (i) when the Jacobian has a continuous line of zeros, take the transformation of E_2 given by $y_1 = x_1 x_2^2$, $y_2 = x_2$. Plainly the map of E_2 does not contain a neighbourhood of the origin.

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

5. The fundamental theorem of algebra. An n-dimensional extension of this is

THEOREM 2. Let y be differentiable for all x, and let the zeros of $\det(dy/dx)$ have only isolated limit-points. Let $|y| \to \infty$ as $|x| \to \infty$, uniformly in all directions. Then for any $u \in E_n$, the equation $y = f(x) = u$ has at least one solution x.

For $|y - u| \to \infty$ as $|x| \to \infty$, and so must attain its lower bound for some finite x, at which by theorem 1 it must vanish. A similar result with slightly different conditions has recently been given by Reichbach [3].

For mappings onto finite regions we have

THEOREM 3. Let y be differentiable and uniformly bounded in the region G, and let the zeros of $\det(dy/dx)$ have only isolated limit-points in G. Let $f(\partial G)$ be the common boundary of a finite region H' and an unbounded region H'', so that $E_n = H' + H'' + f(\partial G)$. Then $y = f(x)$ is soluble for all $y \in H'$ but not for $y \in H''$.

By (v) of theorem 1, if $f(G)$ contained one point of H'', it would contain H'' entirely, which is impossible since y is uniformly bounded on G and H'' is unbounded. Hence $f(G)$ cannot contain any point of $f(\partial G)$; if so it would contain a neighbourhood of such a point, and therefore a point of H'', which is impossible. Hence $f(G)$ lies in H', and therefore contains H', by (v) of theorem 1.

In complex variable theory, the fundamental theorem of algebra is often deduced from Liouville's theorem, that an analytic function which is bounded at infinity is a constant. It is natural to expect that Liouville's theorem admits extension to n dimensions; this has been achieved by G. S. Young, ([4], where further references are given).

6. The method of continuation. In what follows we explore the method of the continuation of solutions of $y = f(x)$, where y moves along an arc and x varies accordingly. The process is very similar to that of analytic continuation in complex variable theory. Under suitable conditions, it can proceed indefinitely, except in the neighbourhood of a boundary, and gives a unique result in simply-connected regions. We need to impose severer

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

restrictions on y, assuming it uniformly differentiable, with dy/dx having an inverse which is uniformly bounded in suitable regions. These conditions are ensured by the usual assumptions that dy/dx is continuous and non-singular.

The basis of continuation is of course the local implicit function theorem, which we cite as

LEMMA 2. In any bounded closed subset V of the region G let y be uniformly differentiable and let the inverse of dy/dx exist and be uniformly bounded, in the sense that all its elements are uniformly bounded. Then there are positive numbers σ, τ, dependent only on V, such that if $x \in V$, $y = f(x)$, $|y' - y| < \tau$, there is a unique $x' \in G$ such that $y' = f(x')$, $|x' - x| < \sigma$. Moreover we have (4), for some $M > 0$, dependent only on V.

This is contained in (i) of theorem 1, for the special case in which dy/dx is non-singular, together with lemma 1. By the assumptions of lemma 2, the determination of ρ, ζ, M_1, η, can proceed independently of the choice of $x \in V$. This is essentially the proof of the local implicit function theorem given, for example, in [5].

We shall need to apply this result repeatedly so as to cover an arc, and need for this purpose conditions which ensure that x does not approach ∂G. We have

LEMMA 3. Let the conditions of lemma 2 hold, and let C be any bounded and closed set containing no point of $f(\partial G)$. Then the set of $x \in G$, $f(x) \in C$, is contained in a certain bounded and closed subset V of G.

The set V is characterized by the requirements that if $x \in V$, then $|x| < R$ and furthermore $U(x; \varepsilon) \in G$, for some fixed positive R and ε. Supposing the contrary, let there be a sequence $x^{(k)} \in G$, $k = 1, 2, \ldots$, $f(x^{(k)}) \in C$, violating one of these two requirements. We may suppose that $f(x^{(k)}) \to y^* \in C$. Supposing first that $|x^{(k)}| \to \infty$, we have at once that $y^* \in f(\partial G)$, contrary to hypothesis. If the sequence $x^{(k)}$ is bounded, we may assume it tends to x^*, either in G or in ∂G. We assume then that for a sequence ε_k, with $\varepsilon_k > 0$, $\varepsilon_k \to 0$ as $k \to \infty$, $U(x^{(k)}; \varepsilon_k)$ is not entirely contained in G. This is impossible if $x^{(k)} \to x^* \in G$, since G has only interior points. Thus $x^* \in \partial G$,

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

and so again $y^* \in f(\partial G)$, contrary to hypothesis.

Our result on the possibility of continuation is then lemma 4. Let C be a continuous arc $y(t)$, $0 \leq t \leq 1$, which contains no point of $f(\partial G)$, and let the assumptions of lemma 2 hold. If then there exists an $x(0) \in G$ such that $f(x(0)) = y(0)$, there then exists a continuous function $x(t)$, $0 \leq t \leq 1$, such that $x(t) \in G$, $f(x(t)) = y(t)$. This function is unique.

For the bounded closed subset $V \subset G$ corresponding to this choice of C, according to lemma 3, we choose $\tau > 0$, according to lemma 2, and subdivide C at a finite number of points $y(t_r)$, $0 = t_0 < t_1 < \ldots < t_N = 1$, chosen so that if $t_{r-1} \leq t \leq t_{r+1}$, then $|y(t) - y(t_r)| < \tau$. By lemma 2 we can continue $x(t)$ as a function of t successively over the intervals (t_r, t_{r+1}), $r = 0, 1, \ldots, N-1$, as required. Moreover, by the last part of lemma 2, asserting that (4) holds, we have that $x(t)$ depends continuously on $y(t)$, and so also on t. The same reason ensures the uniqueness of $x(t)$. If there existed a second function $x^*(t)$ with the same properties, then $x(t) = x^*(t)$ for $t = 0$, and so for $0 \leq t \leq t_1$. By induction, $x(t) = x^*(t)$ also for $t_r \leq t \leq t_{r+1}$, $r = 1, \ldots, N-1$, so that $x(t) = x^*(t)$ for $0 \leq t \leq 1$.

7. The domain as a covering space for the range. In ensuring the uniqueness of "analytic" continuation for a given path, according to lemma 4, we need to rely not so much on the differentiability assumptions of lemma 2, but rather on the following consequences of these assumptions. Let H' be a region containing no point of $f(\partial G)$, and let G' denote $f^{-1}(H')$, the set of $x \in G$, $f(x) \in H'$. For any $y' \in H'$, let $x^{(k)}$, $k = 1, 2, \ldots$, be the solutions of $y' = f(x)$, $x \in G'$. Then there is a neighbourhood $U(y')$ of y', and neighbourhoods $U(x^{(k)})$ of the $x^{(k)}$, each of which is mapped topologically onto $U(y')$ by $y = f(x)$. Furthermore every $x \in G'$ such that $f(x) \in U(y')$ belongs to at least one of the $U(x^{(k)})$. Finally, for every $y' \in H'$ there is to be at any rate one $x \in G'$, $f(x) = y'$.

With these latter three properties we have the essentials for G' to be a covering space for H', under the continuous mapping f; we cite [6], chapter 8, for a general discussion. In this and the next section we outline some of the further developments, insofar as they are relevant to the reversibility

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

of a differentiable mapping. Side by side with the general topological development, one may consider the special case of complex variable theory, in particular the theory of the Riemann surface [7].

We first consider the number of "sheets" in the covering. For each $y' \in H'$, let $\nu(y')$ be the number of solutions of $y' = f(x)$, $x \in G'$. Then it turns out that ν is a constant in H'. We may say that G' is a ν-sheeted covering of H'. A simple example in the complex variable case is given by $w = z^2$, which gives a two-sheeted covering of the finite z-plane less the point $z = 0$. An allied concept is that of the degree of a continuous mapping [6], [8], or more specially that of a differentiable mapping. For our still more special case of an oriented and differentiable mapping the main result that $y = f(x)$ has a constant number of solutions is given in [2] and [9]. Formally expressed, we have

LEMMA 5. With the assumptions of lemma 2, let $H' \subset E_n$ be a region such that $H' \cap f(\partial G)$ is empty. Then $\nu(y)$, the number of solutions of $y = f(x)$, $x \in G$, is constant for $y \in H'$.

We remark first that the result also holds when $\nu = 0$; that is to say if $y = f(x)$, $x \in G$ is insoluble for one $y \in H'$, it is then insoluble for all.

Suppose that for some $y', y'' \in H'$ we have $\nu(y') > \nu(y'')$, so that $\nu(y') \geq 1$. Let y', y'' be connected by a continuous arc in H', say $y(t)$, $0 \leq t \leq 1$. Let the solutions of $y' = f(x)$ be formed at $t = 0$ and continued as functions of t for $0 \leq t \leq 1$, say $x^{(k)}(t)$, $k = 1, \ldots, \nu(y')$, being solutions of $y(t) = f(x)$. These will constitute $\nu(y')$ solutions of $y'' = f(x)$ when $t = 1$, giving a contradiction, provided that the $x^{(k)}(1)$, $k = 1, \ldots, \nu(y')$, are all distinct. This must be so; if two of them coincided, then by continuation in the reverse sense from $t = 1$ to $t = 0$ we would get a contradiction to the uniqueness of continuation along a path, which is ensured by lemma 4.

8. <u>Homotopy, fundamental and monodromy groups</u>. A basic result in the theory of analytic continuation is the "monodromy principle", according to which the result of analytic continuation is independent of the path followed, at least for a class of homotopic paths, any one of which can be derived from any

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

other by continuous variation, within a region in which the function concerned is analytic. Likewise for our case

LEMMA 6. Let the assumptions of lemma 2 hold, and let H' be a region containing a point of $f(G)$ and no point of $f(\partial G)$. Let $y(s,t)$, $0 \leq s, t \leq 1$, lie in H' and be continuous in s and t, and such that $y(s,0) = y'$, $y(s,1) = y''$, for $0 \leq s \leq 1$. Let $x(s,t)$ be continuous in t and a solution of $y(s,t) = f(x(s,t))$ for $0 \leq s,t \leq 1$ such that $x(s,0) = x'$, for some $x' \in G$. Then $x(s,1)$ has the same value for $0 \leq s \leq 1$.

By lemma 3 the $x(s,t)$ must all lie in a certain bounded and closed subset V of G. For this V we choose positive numbers σ, τ and M according to lemma 2. We also choose $\theta > 0$ so that $|s' - s| \leq \theta$ ensures that

$$(5) \qquad |y(s',t) - y(s,t)| < \min(\tfrac{1}{2}M\sigma, \tau)$$

for $0 \leq t \leq 1$. Then, by lemma 1, either

$$(6) \qquad |x(s',t) - x(s,t)| \geq \sigma,$$

or else

$$(7) \qquad |x(s',t) - x(s,t)| \leq M^{-1}|y(s',t) - y(s,t)|$$

whence, by (5),

$$(8) \qquad |x(s',t) - x(s,t)| \leq \tfrac{1}{2}\sigma .$$

Since there can be no continuous transition between (6) and (8) as t goes from 0 to 1, and since (8) holds when $t = 0$, it follows that (8), and so also (7) as the alternative to (6), holds also when $t = 1$. Thus $x(s',1) = x(s,1)$ if $|s' - s| \leq \theta$, and since this is so for the whole s-interval $[0,1]$ we have the result.

Of particular importance is the situation in which we continue x as a function of y along a closed path $y(t)$, $0 \leq t \leq 1$, $y(0) = y(1) = y'$, say, in the region H'. In the first place we note that the result of this will be the same, for fixed y' and for fixed initial x', $f(x') = y'$, for any two closed paths which are homotopic in H'. Arranging the closed paths in H', beginning and ending at y', in homotopic classes we derive the

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

fundamental group of H'. The identity element in this group consists of a closed curve formed by the single point y', together with all closed curves homotopic to it, that is to say which are continuously contractible to a point within H'. Multiplication in the group, though we shall not actually need this, is of course given by making one closed path follow another.

Suppose now that, as in the proof of lemma 5, we form the solutions of $y' = f(x)$ and continue them as functions $x^{(k)}(t)$, $k = 1, \ldots, \nu$, for $0 \le t \le 1$, where now $y(t)$ describes a closed path, $y(0) = y(1) = y'$. In this case the $x^{(k)}(1)$ must also be solutions of $y' = f(x)$, and so must coincide with the $x^{(k)}(0)$, being a permutation of them. This group of permutations forms the monodromy group, and is clearly a representation of the fundamental group of H', with correspondence in particular of the identity elements.

The monodromy group is transitive, in the sense that it contains an element taking any one root $x^{(i)}(0)$ of $y' = f(x)$ into any other $x^{(j)}(0)$, so that $x^{(j)}(0) = x^{(i)}(1)$, provided that H' coincides with $f(G)$ and contains no point of $f(\partial G)$. To form the required closed path in H', we take a path in G which has $x^{(i)}(0)$, $x^{(j)}(0)$ as its end-points, and then form the map of this path in H'. This will be a closed path, beginning and ending at y', and has the required property.

As a result we have that if $y' = f(x)$ has more than one root for x, then the fundamental group of H' does not reduce to the identity. This provides one of the main criteria for our mapping to be $(1,1)$.

9. <u>The first global implicit function theorem</u>. We now obtain one set of conditions that $y = f(x)$ have a single-valued inverse. This is based on the result of continuation of x along closed y-paths, as just described. A second test will depend on the nature of the mapping of the boundary.

In this first test we shall require the range $f(G)$ to be simply-connected. This is to mean that any two arcs in $f(G)$ with the same end-points can be continuously deformed into one another within $f(G)$, in the manner described in lemma 6. Equivalently, any closed curve in $f(G)$ is to be homotopic to a point. This definition of simple-connectedness is more

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

appropriate to our method of analytic continuation than some others, such as that a region is simply-connected if its boundary or its complement is connected. Moreover, these latter definitions are equivalent to the one we use here only if $n = 2$.

The criterion is then

THEOREM 4. Let $y = f(x)$ be defined in the region G and let

(i) y be uniformly differentiable in any bounded and closed subset V of G,

(ii) dy/dx have an inverse which is uniformly bounded in any such V,

(iii) $f(\partial G)$ be the common boundary of two regions H', H'', of which H' is simply-connected, and such that $E_n = H' + H'' + f(\partial G)$,

(iv) H'' be not contained in $f(G)$, as for example if H'' is unbounded and $f(x)$ uniformly bounded in G.

Then $y = f(x)$, $x \in G$, is uniquely soluble for $y \in H'$.

As proved in connection with theorem 3, $f(G)$ in this case coincides with H', and we have only to dispose of the possibility that $\nu(y) \geq 2$ in H'. If this were so, then, by the argument of § 8, H' could not be simply-connected. To recapitulate, for some $y' \in H'$, take two roots of $y' = f(x)$, join them by an arc in G, and form the map of this arc, which will be a closed curve beginning and ending at y'. This closed curve can then be shrunk to a point within H'; continuation along the original closed curve of x as a function of y and along the point version of it will give different results, contrary to lemma 6.

For the case $n = 2$ this result is essentially that of de la Vallee Poussin [10]; see also [11], [2], pp. 193-194, and [6], p. 193, Aufgabe 3.

If $n \geq 3$, the simple-connectivity of H' will not be affected by the removal of a restricted set of points, such as a finite number, and accordingly theorem 4 can be extended for this case. We need

LEMMA 7. Let $G \subset E_n$ be a region, and let $I \subset G$ be a

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

point set. Let I_1 be the set of limit-points of I which lie in G, and let $I_1 \subset I$. Let furthermore I_1 have no limit-point in G. Then, if $n \geq 2$, G - I is a region; if $n \geq 3$, and G is simply-connected, then so is G - I.

This may be proved by direct constructions. Two points of G - I can be joined by an arc in G, which can contain at most a finite number of points of I_1. If $n \geq 2$, these points of I_1 may be avoided by introducing detours round circles, which if sufficiently small will introduce no further points of I. Any remaining points of $I - I_1$ on the modified arc will be finite in number, and again may be avoided by further detours around small circular arcs. Again, any two arcs in G - I with the same end-points may be joined by a family of arcs of the form $y(s,t)$, $0 \leq s \leq 1$, $0 \leq t \leq 1$, lying in G. As in the previous case, this family of arcs will contain at most a finite number of points if I_1, which may be avoided by perturbations of the arcs; the remaining points of $I - I_1$ on the perturbed family of arcs may be avoided similarly.

More generally, we may rely on the fact that such a set I is denumerable, and the known fact that a denumerable set I has the property claimed, insofar as connectivity is concerned. The arcs, or families of arcs as the case may be, each admit a continuous, and so non-denumerable, set of perturbations, of which at least one in each case must be free of points of the denumerable set I. We shall not, however, reproduce the details.

Finally, we verify that G - I is a region in respect of being an open set. This follows from the fact that I contains all its limit points which are in G.

The extension of theorem 4 is then

THEOREM 4'. The conclusion of theorem 4 remains in force if $n \geq 3$ and if for (ii) we substitute

(ii') dy/dx has a uniformly bounded inverse in any closed and bounded subset of G - I, where $I \subset G$, the set I_1 of limit-points of I in G has no limit-point in G, and $I_1 \subset I$.

We apply theorem 4 with $G^* = G - I$, $H^* = H' - f(I)$. It is only necessary to verify that G^*, H^* are regions and that H^*

is simply-connected. It will be sufficient to verify that f(I) has, in regard to H', the properties postulated of I in lemma 7, in regard to G. Consider the limit-points of f(I). Any convergent sequence in f(I) has, as its inverse image, a sequence in I, which will have a limit-point, either in G or on ∂G, possibly ∞; if this limit-point is in G, it will be in I_1, if not, the corresponding limit-point of f(I) will be in $f(\partial G)$, and so not in H'. Thus the limit-points of f(I) which are in H' are the set $f(I_1)$, and have, by the same argument, no limit-point in H'. This completes the proof.

What we have proved, strictly speaking, is that $y = f(x)$ is uniquely soluble for given $y \in H' - f(I)$, and for $x \in G - I$ to be found. Clearly then $y = f(x)$ is soluble for $y \in H'$, and $x \in G$ to be found. The possibility that for some $y^* \in f(I)$ there might be two $x \in G$, say x' and x'', is easily disposed of. By theorem 1, neighbourhoods of x', x'' would be mapped into regions containing a neighbourhood of y', and so both containing points not in f(I). Thus we should have $\nu(y) \geq 2$ for points in H' - f(I), contrary to what has just been proved.

10. <u>A second global theorem</u>. The extension of theorem 4 to allow the Jacobian to have isolated zeros is inadmissible when $n = 2$, and so we fail to cover the following standard theorem in conformal representation. Let $w = f(z)$ effect a (1,1) mapping of a simple closed curve C onto another, C', and be analytic in the finite region D bounded by C; then w maps the z-region D in a (1,1) manner onto the finite region D' bounded by C'. This is covered by theorem 4 if we assume that D' is simply-connected, the complement of D' + C' being an unbounded region, and furthermore that f'(z) has no zeros in D, the Jacobian being $|f'(z)|^2$. While nothing is lost by this last hypothesis, it seems desirable to dispense with it if possible.

Assuming the situation of theorem 3, that $f(\partial G)$ is the boundary of the range, the position may be put roughly as follows. If the inverse relationship $x = f^{-1}(y)$ is many-to-one in the regions concerned, then it is many-to-one on their boundaries. Restricting the relationship to be (1,1) at at any rate one point on the boundaries, we shall restrict it to be (1,1) in the interiors also. We have

THEOREM 5. Let the conditions of theorem 4' hold, except

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

that H' need not be simply-connected, and let in addition

(v) there be at least one point $x^{(o)} \in \partial G$, and a corresponding point $y^{(o)} \in f(\partial G)$ which is the unique map of it, in the sense that $x^{(k)} \in G$, $k = 1, 2, \ldots$, $x^{(k)} \to x^* \in \partial G$, $f(x^{(k)}) \to y^{(o)}$ as $k \to \infty$, imply that $x^* = x^{(o)}$,

(vi) in some region W, the intersection of G and some neighbourhood of $x^{(o)}$, $y = f(x)$ is univalent, so that $x, x' \in W$, $x \neq x'$ imply that $f(x) \neq f(x')$.

Then $y = f(x)$, $x \in G$ is uniquely soluble for $y \in H'$.

As in the case of theorem 4, it is only a question of rejecting the eventuality that $\nu(y) \geq 2$ in H'. By lemma 5, applied to $G^* = G - I$, $H^* = H' - f(I)$, which by lemma 7 are both regions, we have that $\nu(y)$ is constant in H^*. Since $f(I)$ is denumerable, there will be points of H^* arbitrarily close to $y^{(o)}$. Thus for a sequence $y^{(k)}$, $k = 1, 2, \ldots$, tending to $y^{(o)}$, there will be two solutions $x'^{(k)}$, $x''^{(k)}$ of $y^{(k)} = f(x)$. We consider the limiting behaviour of these x-sequences as $k \to \infty$. They cannot both tend to $x^{(o)}$, by condition (vi). Neither can they have a limit-point in G, since $y^{(o)}$ would then be in the interior of H'. Nor can they both have separate limit-points on ∂G, by (v). This completes the proof.

In the above, we may admit ∞ as a boundary point in either case.

We may replace condition (vi) of theorem 5 by more explicit conditions. We have

THEOREM 5'. The conclusion of theorem 5 holds if (vi) is replaced by

(vi') in the region W of (vi) y is uniformly differentiable and dy/dx has an inverse which is uniformly bounded.

This condition is essentially that of [2], p. 192. In postulating uniform differentiability, the formula for the total differential need only hold for points in W, and so in G.

What seems essentially this result was given by Jacobsthal [12]. Fundamentally, it seems to be a case of degrees of mappings of sets and their boundaries.

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

11. Two-dimensional cases. The following result approximates to the standard theorem on conformal mapping mentioned at the beginning of § 10, and is a little more general in some ways. Let C, C' be simple closed curves, bounding regions D, D', the complementary regions being unbounded. Let $f(z)$ be analytic in D, and let $f(z) \to C'$ as $z \to C$. Let $f(z) \to w_0 \in C'$ imply $z \to z_0 \in C$, and in the intersection of a neighbourhood of z_0 with D let $f'(z)$ be uniformly continuous and bounded from zero. Then $w = f(z)$ effects a (1,1) mapping of D onto D'.

A similar result holds for the more general two-dimensional mapping $y_1 = f_1(x_1, x_2)$, $y_2 = f_2(x_1, x_2)$. Assuming, for simplicity, continuous differentiability it will be sufficient that D' be simply-connected and that the Jacobian have no zeros, or alternatively that the boundary mapping be (1,1) at at least one point and that the Jacobian have, say, zeros without limit-point in D. It is not necessary to postulate both a (1,1) mapping of the boundary and that the Jacobian should have no zeros at all.

12. The case of a convex range. Since a convex region is simply-connected, we have

THEOREM 6. In any bounded and closed subset of the region G let y be uniformly differentiable, and let dy/dx have a uniformly bounded inverse. Let $f(\partial G)$ be the common boundary of regions H', H'', with $E_n = H' + H'' + f(\partial G)$, of which H' is convex. If then f(G) does not contain H'', in particular if H'' is unbounded and f(x) is uniformly bounded in G, then $y = f(x)$, $x \in G$ is uniquely soluble in H'.

As in the case of theorem 4, of which this is a special case, we could admit isolated zeros of the Jacobian if $n \geq 3$.

We cite this result separately, since it can be proved independently, by methods of successive approximation. We indicate the approach briefly.

The general idea is the following. Supposing we have an $x' \in G$, $f(x') = y' \in H'$. We wish to solve $y = f(x)$ for some $y \in H'$, and treat x' as a first approximation to the desired root x. Assuming the method is available, and that there are in fact several roots $x^{(k)}$, $k = 1, 2, \ldots$, of $y = f(x)$ for the given y, we may classify G into subsets G_k, so that if $x' \in G_k$ is the

https://doi.org/10.4153/CMB-1961-020-5 Published online by Cambridge University Press

first approximation, the method leads to the root $x^{(k)}$. The sets G_k will then be non-empty, disjoint and open, if the method satisfies rather desirable conditions. It is however impossible to dissect G into a number of disjoint open regions, since a line joining points in two regions would contain boundary points not in any of the G_k.

The argument could be based on one of the methods of successive approximations used to prove the local form of the implicit function theorem. In the treatment of Graves [13], this involves the principle that a contraction mapping has a fixed point. Such mappings have recently been considered by Edelstein [14].

Another approach would be to replace discrete successive approximations by continuous approximation, obtaining the desired x, starting from x', as the limit as $t \to \infty$ of the solution of $dx/dt = -(dy/dx)^{-1}(y - y')$, $x(0) = x'$. Equivalently, we could join y to y' by a straight line, and continue x by means of the local implicit function theorem along this line, starting with x' at the point y'.

13. Mappings of E_n. An interesting special case is

THEOREM 7. Let y be uniformly differentiable in any finite x-region, and in any such region let dy/dx have a uniformly bounded inverse. Let also $|y| \to \infty$ as $|x| \to \infty$, uniformly in all directions. Then to each y corresponds exactly one x.

As previously, if $n \geq 3$ we could allow the Jacobian to have isolated zeros, but not if $n = 2$.

14. A result on critical values. For a critical value of a scalar-valued function $F(x)$ we must of course have $\text{grad } F = 0$. Setting $y = \text{grad } F$, we have a mapping of the form $y = f(x)$, in which $\det(dy/dx)$ will be the determinant of second partial derivatives of F, that is to say the Hessian of F. Thus

THEOREM 8. Let the scalar-valued function $F(x)$ be continuously twice differentiable, and let its Hessian nowhere vanish. Let also $|\text{grad } F| \to \infty$ as $|x| \to \infty$, uniformly in all directions. Then there is a unique critical point x at which grad $F = 0$.

If in addition $F \to +\infty$, say, as $|x| \to \infty$, we derive that F has a unique minimum. For deeper results of this character we refer to [15]. Again, the Hessian could have isolated zeros if $n \geq 3$.

REFERENCES

1. T. W. Chaundy, Differential Calculus, (Oxford, 1935).

2. O. Haupt, G. Aumann, and C. Y. Pauc, Differential und Integralrechnung, Vol. II, (Berlin, 1950)

3. M. Reichbach, Some theorems on mappings onto, Pac. J. of Math. 10 (1960), 1397-1407.

4. G. S. Young, Extensions of Liouville's theorem to n dimensions, Math. Scand. 6 (1958), 289-292

5. E. J MacShane and T. A. Botts, Real Analysis, (Princeton, 1959).

6. H. Seifert and W. Threlfall, Topologie, (Leipzig, 1934).

7. H. Behnke and F. Sommer, Analytische Funktionen, (Berlin, 1955).

8. M. Nagumo, A theory of degree of mapping based on infinitesimal analysis, Amer. J. Math. 73 (1951), 485-496.

9. A. M. Rodnyanskiǐ, On differentiable mappings of regions, Doklady Akad. Nauk SSSR (N. S.) 72 (1950), 15-17.

10. Ch. -J. de la Vallée Poussin, Cours d'Analyse Infinitesimale, Vol. I, (7th edition, Paris, 1930).

11. C. Carathéodory and H. Rademacher, Über die Eineindeutigkeit im Kleinen und im Grossen stetiger Abbildungen von Gebieten, Arch. der Math. u. Phys. (3), 26 (1917), 1 - 9.

12. E. Jacobsthal, Über die eineindeutige Abbildungen zweier Bereiche aufeinander bei nichtverschwindender Funktionaldeterminante, Kong. Norsk. Vidensk. Selskab Forhandlinger, 13, no. 30 (1940), 123-126.

13. L. M. Graves, Theory of functions of real variables, (New York, 1946).

14. M. Edelstein, An extension of Banach's contraction principle, Proc. Amer. Math. Soc. 12 (1961), 7-10.

15. M. Morse, Calculus of Variations in the Large, (New York, 1934).

University of Toronto