

ON THE INVERSE FUNCTION THEOREM

F. H. CLARKE

The classical inverse function theorem gives conditions under which a C^1 function admits (locally) a C^1 inverse. The purpose of this article is to give conditions under which a Lipschitzian (not necessarily differentiable) function admits (locally) a Lipschitzian inverse. The classical result is a special case of the theorem.

1. Introduction. Let $f: R^n \rightarrow R^n$ satisfy a Lipschitz condition in a neighborhood of a point x_0 in R^n . Thus for some constant K , for all x and y near x_0 , we have

$$(1) \quad |f(x) - f(y)| \leq K |x - y|,$$

where $|\cdot|$ denotes the usual Euclidean norm. The usual $n \times n$ Jacobian matrix of partial derivatives, when it exists, is denoted $Jf(x)$. We topologize the vector space \mathcal{M} of $n \times n$ matrices with the norm

$$\|M\| = \max |m_{ij}|,$$

where

$$M = (m_{ij}), \quad 1 \leq i \leq n, \quad 1 \leq j \leq n.$$

DEFINITION 1. The *generalized Jacobian* of f at x_0 , denoted $\partial f(x_0)$, is the convex hull of all matrices M of the form

$$M = \lim_{i \rightarrow \infty} Jf(x_i),$$

where x_i converges to x_0 and f is differentiable at x_i for each i .

The above extends to vector-valued functions the notion of "generalized gradient" introduced by the author in [2]. It is a consequence of Rademacher's theorem that f is almost everywhere differentiable near x_0 . Furthermore, $Jf(x)$ is bounded near x_0 as a result of (1). These observations imply

PROPOSITION 1. $\partial f(x_0)$ is a nonempty compact convex subset of \mathcal{M} .

DEFINITION 2. $\partial f(x_0)$ is said to be of *maximal rank* if every M in $\partial f(x_0)$ is of maximal rank.

The following theorem, which is our main result, is proven in §2:

THEOREM 1. *If $\partial f(x_0)$ is of maximal rank, then there exist neighborhoods U and V of x_0 and $f(x_0)$ respectively, and a Lipschitzian function $g: V \rightarrow R^n$ such that*

- (a) $g(f(u)) = u$ for every $u \in U$
- (b) $f(g(v)) = v$ for every $v \in V$.

When f is C^1 , $\partial f(x_0)$ reduces to $Jf(x_0)$, and the function g above is necessarily C^1 as well. Thus we recover the classical theorem.

REMARK 1. There are instances of Lipschitz functions which satisfy the conclusions but not the hypotheses of the theorem. Note that it is not enough to assume that Jf is of maximal rank whenever it exists, as the function $|x|$ ($n = 1$) demonstrates. A simple example to which the theorem applies ($n = 2$) is the following: $f(x, y) = [|x| + y, 2x + |y|]$, near $(0, 0)$. We find

$$\partial f(0, 0) = \left\{ \begin{bmatrix} s & 1 \\ 2 & t \end{bmatrix} : -1 \leq s \leq 1, -1 \leq t \leq 1 \right\}.$$

2. Proof of the theorem. We let $\mathcal{M}(0, 1)$ denote the unit ball in \mathcal{M} .

LEMMA 1. *Let ϵ be a positive number. Then for all x sufficiently near x_0 ,*

$$\partial f(x) \subset \partial f(x_0) + \epsilon \mathcal{M}(0, 1).$$

This is a direct consequence of Definition 1, which is equivalent to the following: ∂f is the smallest convex-valued upper-semicontinuous set-valued mapping which contains Jf whenever the latter exists.

We suppose that $h: R^n \rightarrow R$ is a given C^1 function, and that f is Lipschitzian near x . We denote by Df the derivative of f when it exists.

LEMMA 2.

$$\partial(h \circ f)(x) \subset \nabla h(f(x)) \partial f(x).$$

Proof. According to [2, Proposition 1.11], we may express the generalized gradient of the Lipschitzian function $h \circ f$ as in Definition 1 (for $h \circ f$ replacing f) with the added condition that x_i lie in the complement of E , where E is any set of measure 0, without affecting the result. Let us choose for E the set of points (near x) where Df fails to exist. At any point y not in E , by the chain rule we have

$$\nabla(h \circ f)(y) = \nabla h(f(y))Jf(y).$$

Thus $\partial(h \circ f)(x)$ is the convex hull of a set of points each of which is of the form

$$\lim_{i \rightarrow \infty} \nabla h(f(y_i))Jf(y_i),$$

where y_i converges to x . But any such point belongs to

$$\nabla h(f(x))\partial f(x).$$

The result now follows from the fact that this last set is convex.

We let B denote the open unit ball in R^n .

LEMMA 3. *There are positive numbers r and δ with the following property: given any unit vector v in R^n , there is a unit vector w in R^n such that, whenever x lies in $x_0 + rB$ and M belongs to $\partial f(x)$, then*

$$(2) \quad w \cdot (Mv) \geq \delta.$$

Proof. (Note that we think of vectors as columns, and that the required vector w is to depend only on v , and not on x .) Let S denote the unit sphere in R^n .

The subset $\partial f(x_0)S$ of R^n is compact and does not contain 0, since $\partial f(x_0)$ is of maximal rank. Hence for some $\delta > 0$, $\partial f(x_0)S$ is distance 2δ from 0. For positive ϵ sufficiently small, GS is distance at least δ from 0, where $G = \partial f(x_0) + \epsilon\mathcal{M}(0, 1)$. By Lemma 1, it follows that for some positive r ,

$$(3) \quad x \in x_0 + rB \Rightarrow \partial f(x) \subset G.$$

We may suppose r chosen so that f satisfies (1) on $x_0 + r\bar{B}$.

Now let any unit vector v be given. It follows from the above that the convex set Gv is distance at least δ from 0. By the usual separation theorem for convex sets, there is a unit vector w such that

$$w \cdot (\gamma v) \geq \delta$$

for every γ in G . Relation (2) follows from this along with (3).

LEMMA 4. *If x and y lie in $x_0 + r\bar{B}$, then*

$$|f(x) - f(y)| \geq \delta |x - y|.$$

Proof. We may suppose $x \neq y$ and (in view of the continuity of f) $x, y \in x_0 + rB$. Set

$$v = (y - x)/|y - x|,$$

$$\lambda = |y - x|,$$

so that $y = x + \lambda v$.

Let π be the plane perpendicular to v and passing through x . The set P of points x' in $x_0 + rB$ where $Df(x)$ fails to exist is of measure 0, and hence by Fubini's theorem, for almost every x' in π , the ray

$$x' + tv, \quad t \geq 0$$

meets P in a set of 0 one-dimensional measure. Choose an x' with the above property and sufficiently close to x so that $x' + tv$ lies in $x_0 + rB$ for every t in $[0, \lambda]$. Then the function

$$t \rightarrow f(x' + tv)$$

is Lipschitzian for t in $[0, \lambda]$ and has a.e. on this interval the derivative $Jf(x' + tv)v$. Thus

$$f(x' + \lambda v) - f(x') = \int_0^\lambda Jf(x' + tv)v dt.$$

Let w correspond to v as in Lemma 3. We deduce

$$w \cdot [f(x' + \lambda v) - f(x')] = \int_0^\lambda w \cdot [Jf(x' + tv)v] dt \geq \int_0^\lambda \delta dt = \delta \lambda.$$

Recalling the definition of λ , we arrive at:

$$|f(x' + \lambda v) - f(x')| \geq \delta |x - y|.$$

This may be done for x' arbitrarily close to x . Since f is continuous, the lemma ensues.

LEMMA 5. $f(x_0 + rB)$ contains $f(x_0) + (r\delta/2)B$.

Proof. Let y be any point in $f(x_0) + (r\delta/2)B$, and let the minimum of $|y - f(\cdot)|^2$ over $x_0 + r\bar{B}$ be attained at x . We claim x belongs to $x_0 + rB$. Otherwise, by Lemma 4 and the triangle inequality,

$$\begin{aligned}
r\delta/2 > |y - f(x_0)| &\geq |f(x) - f(x_0)| - |y - f(x)| \\
&\geq \delta |x - x_0| - |y - f(x)| \\
&\geq \delta r - |y - f(x_0)| > \delta r - \delta r/2 = r\delta/2,
\end{aligned}$$

which is a contradiction. Thus x yields a local minimum for the function $|y - f(\cdot)|^2$, and consequently (see [2, Corollary 1.10])

$$0 \in \partial |y - f(x)|^2.$$

We now use Lemma 2 to conclude that 0 belongs to the set

$$2(y - f(x))'\partial f(x).$$

But Lemma 3 implies that every matrix in $\partial f(x)$ is invertible, hence the above is possible only if $f(x) = y$.

We now set $V = f(x_0) + (r\delta/2)B$, and we define g on V as follows: $g(v)$ is the unique x in $x_0 + rB$ such that $f(x) = v$. We choose U as any neighborhood of x_0 satisfying $f(U) \subset V$. The theorem is now seen to follow, since Lemma 4 implies that g is Lipschitz with constant $1/\delta$.

REMARK 2. The essential result of Lemma 5 is that the image under f of $x_0 + rB$ is a neighborhood of $f(x_0)$. H. Halkin and the referee have pointed out that this follows from Lemma 4 and the invariance of domain theorem. We retain our present proof for its elementary nature.

REMARK 3. When f is “strongly differentiable”, ∂f reduces to the derivative. Thus Theorem 1 extends the known result in this case [1] [3] (in finite dimensions).

REMARK 4. Following the author’s introduction of generalized gradients, two further (and successively more general) extensions of the notion of derivative have been introduced. J. Warga [6] [7] develops the concept of “derivate container”; it can be shown that Theorem 1 is a special case of his “implicit point” theorem. H. Halkin [4] [5] obtains the closely related “interior mapping” theorem (and other results) via the notion of “screen”.

REMARK 5. The theorem remains true (with slight modifications in proof) if in Definition 1 we impose the constraint that the points x_i lie in the complement of a given set E of measure zero (let us call the resulting generalized Jacobian $\partial_E f$). Thus the maximality of rank need be verified

only for $\partial_E f(x_0)$ rather than $\partial f(x_0)$. When $f: R^n \rightarrow R$ is real-valued, it is known that $\partial_E f = \partial f$ ([2, Proposition 1.11]). It has been pointed out by H. Halkin and J. Warga that the question of whether this is true in the vector-valued case is more profound. We remark that this question seems related to a possible "Denjoy property" for vector derivatives.

REFERENCES

1. N. Bourbaki, *Variétés différentielles et analytiques*, Hermann, Paris.
2. F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., **205** (1975), 247-262.
3. A. Nijenhuis, *Strong derivatives and inverse mappings*, Amer. Math. Monthly, **81** (1974), 969-980.
4. H. Halkin, *Interior mapping theorem with set valued derivatives*, Journal d'Analyse Mathématique, to appear.
5. ———, *Mathematical programming without differentiability*, Proceedings of "MRC Symposium on the Calculus of Variations and Optimal Control" (September, 1975), D. L. Russell, Ed., Mathematics Research Center, University of Wisconsin, Academic Press, N.Y. (1976).
6. J. Warga, *Controllability and necessary conditions in unilateral problems without differentiability assumptions*, to appear.
7. ———, *Derivate containers, inverse functions and controllability*, Proceedings of "MRC Symposium on the Calculus of Variations and Optimal Control" (September, 1975), D. L. Russell, Ed., Mathematics Research Center, University of Wisconsin, Academic Press, N.Y. (1976).

Received December 3, 1975. Supported in part by National Research Council of Canada grant A 9082.

UNIVERSITY OF BRITISH COLUMBIA