IV. Smooth Functions

In many mathematical contexts one considers functions with properties that are stronger than continuity. For example, if X and Y are open in Euclidean spaces it is often necessary or desirable to consider functions with good differentiability properties. The latter often have important metric or topological consequences, and in this course we shall be interested in certain results of this type.

IV.1 : Linear approximations

(Edwards, §§ II.1, II.2)

For functions of one real variable, a function is continuous at a point if it has a derivative at that point, but there are standard examples of functions of two variables that have partial derivatives defined near a point but are not continuous there. On the other hand, a basic result in multivariable calculus shows that functions that have continuous partial derivatives near a point are necessarily continuous at the point in question.

We shall begin by establishing a version of this result that holds for functions of an arbitrary (finite) number of real variables. It will be convenient to adopt some notation first. The unit vector in $\mathbf{R}^n$ whose i^{th} coordinate is 1 and whose other coordinates are 0 will be denoted by $\mathbf{e}_i$. If U is an open set in $\mathbf{R}^n$ and $f : U \to \mathbf{R}^n$ is a (not necessarily continuous) function such that all first partial derivatives exist at some point $p \in U$, then the *gradient* $\nabla f(p)$ will denote the vector whose i^{th} coordinate is the partial derivative of f with respect to the i^{th} variable at p. We shall use $\langle u, v\rangle$ to denote the usual dot product of two vectors $u, v \in \mathbf{R}^n$.

PROPOSITION. *Let U be an open subset in $\mathbf{R}^n$, let $x \in \mathbf{R}^n$, and let $f : U \to \mathbf{R}^n$ be a (not necessarily continuous) function such that f has continuous partial derivatives on some open subset of U containing x. Then for all sufficiently small $h \neq 0$ in $\mathbf{R}^n$ one can define a function $\theta(h)$ such that*

$$f(x+h) - f(x) = \langle \nabla f(x), h\rangle + |h|\,\theta(h)$$

where $\lim_{h\to 0}\theta(h) = 0$.

Proof. Write $h = \sum_i t_i h_i$ for suitable real numbers t_i, take $\delta > 0$ so that f has continuous partial derivatives on $N_\delta(x)$, and assume that $0 < |h| < \delta$. Define h_i for $0 \leq i \leq n$ recursively by $h_0 = 0$ and $h_{i+1} = h_i + t_{i+1}\mathbf{e}_{i+1}$. Then $h_n = h$ and we have

$$f(x+h) - f(x) = \sum_i f(x+h_i) - f(x_{i-1})$$

and if we apply the ordinary Mean Value Theorem to each summand we see that the right hand side is equal to

$$\sum_i \frac{\partial}{\partial x_i} f\left(x + h_{i-1} + K_i(x)t_i\right) \cdot t_i$$

for some numbers $K_i(x) \in (0,1)$. The expression above may be further rewritten in the form

$$\langle \nabla f(x), h\rangle + \left(\sum_i \frac{\partial}{\partial x_i} f\left(x + h_{i-1} + K_i(x)t_i\right) - \frac{\partial}{\partial x_i} f(x)\right) \cdot t_i$$

and therefore an upper estimate for $|f(x+h) - f(x) - \langle \nabla f(x), h\rangle\ |$ is given by

$$\sum_i \left|\frac{\partial}{\partial x_i} f\left(x + h_{i-1} + K_i(x)t_i\right) - \frac{\partial}{\partial x_i} f(x)\right| |t_i|.$$

Since the partial derivatives of f are all continuous at x, for every $\varepsilon > 0$ there is a $\delta_1 > 0$ such that $\delta_1 < \delta$ and $|h| < \delta_1$ implies that each of the differences of partial derivatives has absolute value less than ε/n. Since $h \neq 0$, if we define $\theta(h)$ as in the statement of the conclusion, the preceding considerations imply show that

$$|\theta(h)| < \sum_i \frac{\varepsilon \cdot |t_i|}{n \cdot |h|}$$

and since

$$|t_i| \leq \left(\sum_i t_i^2\right)^{1/2}$$

it follows that $|\theta(h)| < \varepsilon$ when $0 < |h| < \delta_1$.■

COROLLARY. *If f satisfies the conditions of the proposition, then f is continuous at x.*■

The conclusion of the theorem indicates the right generalization of differentiability to functions of more than one variable:

Definition. Let U be an open subset in $\mathbf{R}^n$ and let $f : U \to \mathbf{R}^n$ be a function (with no further assumptions at this point). The function f is said to be *differentiable* at the point $x \in U$ if there is a linear transformation $L : \mathbf{R}^n \to \mathbf{R}^m$ such that for all sufficiently small vectors h we have

$$f(x+h) - f(x) = L(h) + |h|\,\theta(h)$$

where $\lim_{h\to 0} \theta(h) = 0$.

Immediate Consequence. *If $m = 1$ in the definition above and we restrict h so that $h = t\mathbf{e}_i$ for $|t|$ sufficiently small then if f is differentiable at x it follows that all first partial derivatives of f are defined at x and*

$$\frac{\partial f(x)}{\partial x_i} = L(\mathbf{e}_i)\ .$$

In particular, the preceding shows that there is at most one choice of L for which the differentiability criterion is true, at least if $m = 1$.■

Of course the proposition above implies that a function is differentiable if it has continuous partial derivatives.

The differentiability of a function turns out to be determined completely by the differentiability of its coordinate functions:

PROPOSITION. *let U be open in $\mathbf{R}^n$, let $f : U \to \mathbf{R}^m$ be a function, and express f in coordinates as $f(x) = \sum_j y_j(x)\mathbf{e}_j$. Then f is differentiable at x if and only if each y_j is differentiable at x, and in this case the linear transformation L is given by*

$$L(u) = \sum_i \langle \nabla y_i(x), u\rangle \mathbf{e}_i \ .$$

It follows that there is at most one choice of L for an arbitrary value of m. If we write $u = \sum_j u_j \mathbf{e}_j$ then this yields the fundamental identity

$$L(u) = \sum_i \sum_j \frac{\partial y_i(x)}{\partial x_j} u_j \mathbf{e_i}$$

for the derivative linear transformation. In words, the (i,j) entry of the matrix representing L is the j^{th} partial derivative of the i^{th} coordinate function.

Proof of proposition. Suppose that f is differentiable at x. For i between 1 and m, let $\mathbf{P}_i$ be projection onto the i^{th} coordinate. If we apply $\mathbf{P_i}$ to the formula for $f(x+h) - f(x)$ we obtain the following relationship:

$$y_i(x+h) - y_i(x) = \mathbf{P_i}L(h) + |h|\mathbf{P_i}\theta(h)$$

The composite $\mathbf{P}_i L$ is linear because both factors are linear, and the relation

$$\left|\mathbf{P_i}\theta(h)\right| \leq \left|\theta(h)\right|$$

shows that $\lim_{h\to 0} \mathbf{P_i}\theta(h) = 0$, so that $y_i = \mathbf{P}_i f$ is differentiable at X and

$$\nabla y_i(x) = \mathbf{P}_i L$$

as required.

Now suppose that each y_i is differentiable at x, and write

$$f(x+h) - f(x) = \sum_i \left(y_i(x+h) - y_i(x)\right)\mathbf{e_i} =$$

$$\sum_i \langle \nabla y_i(x), h\rangle \mathbf{e_i} + \sum_i |h|\theta_i(h)\mathbf{e_i}$$

where $h = \sum_j t_j \mathbf{e}_j$ and $\lim_{h\to 0} \theta_i(h) = 0$ for all i. Choose $\delta > 0$ so that $N_\delta(x) \subset U$ and $0 < |h| < \delta$ implies $|\theta_i(h)| < \varepsilon/n$ for all i. Then $0 < |h| < \delta$ implies $|\theta(h)| < \varepsilon$, whosing that f is differentiable at x and the linear transformation L has the form described in the proposition.■

Smoothness classes of functions

If U and V are open sets in Euclidean spaces and $f : U \to V$ is a function, then we say that f is *(smooth of class)* $\mathbf{C}^1$ if Df exists everywhere and is continuous. For $r \geq 2$ we inductively define f to be *(smooth of class)* $\mathbf{C}^r$ if Df is (smooth of class) $\mathbf{C}^{r-1}$, and we say that f is *(smooth of class)* $\mathbf{C}^\infty$ if it is smooth of class $\mathbf{C}^r$ for all positive integers r. For the sake of notational uniformity we often say that every continuous function is of class $\mathbf{C}^0$.

It is elementary to check that a function is of class $\mathbf{C}^r$ if and only if all its coordinate functions are and that

$$\mathbf{C}^\infty \Longrightarrow \mathbf{C}^r \Longrightarrow \mathbf{C}^{r-1} \Longrightarrow \mathbf{C}^0$$

for all r. Every polynomial function is obviously of class $\mathbf{C}^\infty$, and for each r there are many examples of functions that are $\mathbf{C}^r$ but not $\mathbf{C}^{r+1}$. If $r = 0$ the absolute value function $f_0(x) = |x|$ is an obvious example, and inductively one can construct an example f_r which is $\mathbf{C}^r$ but not $\mathbf{C}^{r+1}$ by taking an antiderivative of f_{r-1}.

One important point in ordinary and multivariable courses is that standard algebraic operations on differentiable (or smooth) functions yield differentiable (or smooth) functions. In particular, this applies to addition, subtraction, multiplication, and division (provided the denominator is nonzero in this case). We shall use these facts without much further comment. The smoothness properties of composites of differentiable and smooth functions will be discussed reasonably soon.

Matrix operations

Plenty of examples of smooth functions can be found in multivariable calculus books, so we concentrate here on some basic examples that will be needed shortly.

PROPOSITION. *Addition addition of $m \times n$ matrices is a $\mathbf{C}^\infty$ map from*

$$\mathbf{R}^{2mn} \cong (\mathbf{M}(m,n))^2$$

to $\mathbf{R}^{mn} \cong \mathbf{M}(m,n)$, scalar multiplication of $m \times n$ matrices is a $\mathbf{C}^\infty$ map from

$$\mathbf{R}^{mn+1} \cong \mathbf{R} \times \mathbf{M}(m,n)$$

to $\mathbf{R}^{mn} \cong \mathbf{M}(m,n)$, and matrix multiplication from $\mathbf{M}(m,n) \times \mathbf{M}(n,p)$ to $\mathbf{M}(m,p)$ is also a $\mathbf{C}^\infty$ map.

This simply reflects the fact that that the entries of a matrix sum or product are given by addition and multiplication operations on the entries of the original matrices (or matrix and scalar).■

The next result is slightly less trivial but still not difficult.

PROPOSITION. *The set of invertible $n \times n$ matrices $GL(n,\mathbf{R})$ is an open subset of $\mathbf{R}^{n^2} \cong \mathbf{M}(n,n)$, and the map from $GL(n,\mathbf{R})$ to itself sending a matrix to its inverse is a $\mathbf{C}^\infty$ map.*

Proof. We shall prove this using coordinates; it is possible to prove the result without using coordinates, but the proof using coordinates is shorter.

Recall that the determinant of a square matrix is a polynomial function in the entries of the matrix and that a matrix is invertible if and only if its determinant is nonzero. The former implies that the determinant function is continuous (and in fact $\mathbf{C}^\infty$), while the second observation and the continuity of the determinant imply that the set of invertible matrices, which is equal to the set $\det^{-1}(\mathbf{R} - \{0\})$, is open. But Cramer's Rule implies that the entries of the inverse to a matrix are rational expression in the entries of the original matrix, and thus the entries of an inverse matrix are $\mathbf{C}^\infty$ functions of the entries of the original matrix. Therefore the matrix inverse is a $\mathbf{C}^\infty$ function from $GL(n,\mathbf{R})$ to itself.■

Matrix norms

Before discussing the metric and topological properties of $\mathbf{C}^r$ functions it is necessary to know a little about the corresponding properties of their linear approximations. The most basic property is a strong form of uniform continuity.

PROPOSITION. *If $L : \mathbf{R}^n \to \mathbf{R}^m$ is a linear transformation, then there is a constant $b > 0$ such that $|L(u)| \leq b|u|$ for all $U \in \mathbf{R}^n$.*

Proof. The easiest way to see this is to note that L is continuous, and therefore the restriction of the function $h(u) = |L(u)|$ to the (compact!) unit sphere in $\mathbf{R}^n$ assumes a maximum value, say c, so that $|u| = c$. Every vector u may be written as a product cv where c is nonnegative and $|v| = 1$. It then follows that

$$|L(u)| = |cL(v)| = c \cdot |L(v)| = |u| \cdot |L(v)| \leq b \cdot |u|$$

as required.∎

Notation. The maximum value b is called the *norm* of L and written $\| L \|$.

PROPOSITION. *The norm of a matrix (or linear transformation) makes the space of $m \times n$ matrices into a normed vector space.*

Proof. By definition the norm is nonnegative, and if it is zero then $L(v) = 0$ for all $v \in \mathbf{R}^n$ satisfying $|v| = 1$; it follows that $L(v) = 0$ for all v (why?). If a is a scalar and L is a linear transformation, then the maximum value $\| L \|$ of $|aL(v)|$ for $|v| = 1$ is simply $|a| \cdot \| L \|$. Finally, if L_1 and L_2 are linear transformations and v is a unit vector such that $|[L_1 + L_2](v)| = \| L_1 + L_2 \|$, then we have

$$\| L_1 + L_2 \| = |[L_1 + L_2](v)| \ \leq \ |L_1(v)| + |L_2(v)| \ \leq \| L_1 \| + \| L_2 \| \ .$$

Thus the norm as defined above satisfies the conditions for a normed vector space.∎

The matrix norm has the following additional useful property:

PROPOSITION. *If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then $\| AB \| \leq \| A \| \cdot \| B \|$.*

Proof. Let v be a unit vector in $\mathbf{R}^p$ at which the function $f(x) = ABx$ takes a maximum value. Then we have

$$|ABx| \ \ \leq \ \ \| A \| \cdot |Bx| \ \ \leq \ \ \| A \| \cdot \| B \|$$

as required.∎

Comparisons of norms

Although there are many different norms that can be defined on $\mathbf{R}^n$, the following result shows that they all yield the same open sets.

THEOREM. *Let $|...|$ denote the standard Euclidean norm on $\mathbf{R}^n$, and let $\| ... \|$ denotes some other norm. Then there are positive constants A and B such that*

$$\| x \| \leq \ A\,|x| \ , \quad |x| \ \leq \ B \,\| x \|$$

for all $x \in \mathbf{R}^n$. In particular, the identity maps of normed vector spaces from Euclidean space $(\mathbf{R}^n, |...|)$ to $(\mathbf{R}^n, \| ... \|)$ and vice versa are uniformly continuous.

Proof. Given a typical vector x, write it as $\sum_i x_i \mathbf{e}_i$.

Choose M such that $\| \mathbf{e}_i \| \leq M$ for all i. Then we have

$$\| x \| \leq \sum_i |x_i| \cdot \| \mathbf{e}_i \| \leq nM \sum_i |x_i|$$

and by the Cauchy-Schwarz-Buniakovsky Inequality the summation is less than or equal to $n^{1/2}|x|$. Therefore $\| x \| \leq n^{3/2} M |x|$.

The preceding paragraph implies that the function $f(x) = \| x \|$ is a continuous function on $\mathbf{R}^n$ with respect to the usual Euclidean metric. Let $c > 0$ be the minimum value of f on the unit sphere defined by $|x| = 1$. It then follows that $\| x \| \geq c \cdot |x|$ for all x and hence that

$$|x| \leq \frac{1}{c} \| x \|$$

for all $x \in \mathbf{R}^n$.■

COROLLARY. *The same conclusion holds if the Euclidean norm is replaced by a second arbitrary norm.*

Proof. Let α and β denote arbitrary norms on $\mathbf{R}^n$ and let $|...|$ denote the Euclidean norm. Then there are positive constants A_α, A_β, B_α B_β such that the following hold for all $x \in \mathbf{R}^n$:

$$\alpha(x) \leq A_\alpha |x| , \quad |x| \leq B_\alpha \, \alpha(x)$$

$$\beta(x) \leq A_\beta |x| , \quad |x| \leq B_\beta \, \beta(x)$$

These immediately imply $\alpha(x) \leq A_\alpha B_\beta \beta(x)$ and $\beta(x) \leq A_\beta B_\alpha \alpha(x)$.■

The following observation will be useful later.

PROPOSITION. *If A is an $n \times n$ matrix such that $\| A \| < 1$, then $I - A$ is invertible.*

Proof. Suppose that the conclusion is false, so that $I - A$ is not invertible. Then there is a nonzero vector $v \in \mathbf{R}^n$ such that $(I - A)v = 0$. The latter implies that $Ax = x$ for some x such that $|x| = 1$, which in turn implies that $\| A \| \geq 1$.■

Note. If $\| A \| < 1$, then the the previously stated inequalities for the matrix norm show that

$$\| A^k \| \leq \| A \|^k$$

and the latter implies that the inverse to $I - A$ may be computed using the geometric series:

$$(I - A)^{-1} = \sum_k A^k$$

IV.2 : Properties of smooth functions

(Edwards, § II.3)

In this section we shall establish generalizations of two important principles from elementary calculus. One is a version of the Chain Rule, and the other is a general form of a basic consequence of the Mean Value Theorem. As an application of these results we shall show that the restriction of a smooth map to a compact set satisfies a metric inequality called a *Lipschitz condition* that generalizes the strong form of uniform continuity which holds for linear maps of (finite-dimensional) Euclidean spaces.

The Chain Rule

Undergraduate multivariable calculus courses generally state one or more extensions of the ordinary Chain Rule

$$[g \circ f]'(x) = g'(f(x)) \cdot f'(x)$$

to functions of several real variables. Linear transformations provide a conceptually simple way of summarizing the various generalizations of this basic fact from single variable calculus:

CHAIN RULE. *Let U and V be open in $\mathbf{R}^n$ and $\mathbf{R}^m$ respectively, let $f : U \to V$ be a map that is differentiable at x, and let $g : V \to \mathbf{R}^p$ be differentiable at $f(x)$. Then $g \circ f$ is differentiable at x and*

$$D[g \circ f](x) = D(g)(f(x)) \circ Df(x) \ .$$

Proof. By the definition of differentiability for g at $y = f(x)$, for $|k|$ sufficiently small we have

$$g(y+k) - g(y) = [Dg(y)](k) + |k| \cdot \alpha(k)$$

where $\lim_{k \to 0} \alpha(k) = 0$. If we take k so that $y + k = f(x+h)$ for $|h|$ sufficiently small, then we have $k = f(x+h) - f(x)$, and by the differentiability of f at x we have

$$k = f(x+h) - f(x) = [Df(x)](h) + |h| \cdot \beta(h)$$

where $\lim_{h \to 0} \beta(h) = 0$. If we make this substitution into the first equation in the proof we obtain the relation

$$[g \circ f](x+h) - [g \circ f](x) = [Dg(f(x))]([Df(x)](h) + |h| \cdot \beta(h)) + |k| \alpha(f(x+h) - f(x))$$

and for $h \neq 0$ the right hand side may be rewritten as follows:

$$[Dg(f(x))]([Df(x)](h)) + [Dg(f(x))](|h| \cdot \beta(h)) + |h| \cdot \frac{|k|}{|h|}\, \alpha\big(f(x+h) - f(x)\big)$$

Let $\varepsilon > 0$ be given. We need to show there is a $\delta > 0$ such that $|h| < \delta$ implies the following inequalities:

$$\| \, Dg(f(x)) \, \| \cdot |\beta(h)| \quad < \quad \frac{\varepsilon}{2}$$

$$\frac{|k|}{|h|} < \| Df(x) \| + 1$$

$$\left|f(x+h) - f(x)\right| < \frac{\varepsilon}{2(\| Df(x) \| + 1)}$$

The first of these can be realized by the limit condition on β, and the third can be realized by the limit condition on α and the continuity of f at x. To deal with the second condition, note that

$$|k| = \left|f(x+h) - f(x)\right| = \left|[Df(x)](h) + |h|\beta(h)\right| \leq$$

$$\| Df(x) \| \cdot |h| \; + \; |h|\left|\beta(h)\right|$$

so that

$$\frac{|k|}{|h|} \leq \| Df(x) \| \; + \; \left|\beta(h)\right|$$

and thus we may realize the second inequality if we choose δ so that $0 < |h| < \delta$ implies

$$\left|\beta(h)\right| < \frac{\varepsilon}{2(\| Df(x) \| + 1)} .$$

If we combine all these we see that

$$[g \circ f](x+h) \; - \; [g \circ f](x) \; = \; [Dg\,(f(x))]\,([Df(x)](h)) \; + \; |h|\,\gamma(h)$$

where $\lim_{h \to 0} \gamma(h) = 0$.■

COROLLARY. *In the notation of the preceding result, if f is $\mathbf{C}^r$ on U and g is $\mathbf{C}^r$ on V, then $g \circ f$ is $\mathbf{C}^r$.*

Proof. Suppose that $r = 1$. Then the Chain Rule formula, the continuity of the derivatives of f and g and the continuity of f show that $D(g \circ f)$ is continuous.

Suppose now that $r \geq 2$ is an integer and we have shown the Corollary inductively for $\mathbf{C}^s$ functions for $1 \leq s \leq r-1$. Then the functions $Dg \circ f$ and Df are $\mathbf{C}^{r-1}$ by the induction hypothesis and the fact that f is $\mathbf{C}^r$, and the matrix product of these functions is also $\mathbf{C}^r$ because matrix multiplication is $\mathbf{C}^\infty$.■

Since the result is true for all finite r, it follows immediately that the conclusion is also true if $r = \infty$.

Example. Suppose that U is open in $\mathbf{R}^n$ and $f : U \to \mathbf{R}^m$ is $\mathbf{C}^r$ for some $r \geq 1$; let $a, x \in U$ be such that $N_{2|x-a|}(a) \subset U$. Then the function

$$g(t) = f\,(a + (x-a))$$

is a $\mathbf{C}^r$ function on some interval $(-\delta, 1+\delta)$ and

$$g'(t) = \left[Df\,(a + t(x-a))\,\right](x-a) .$$

The Mean Value Theorem for real-valued differentiable functions of one real variable does not generalize directly to other situations, but some of its important consequences involving derivatives and definite integrals can be extended. The following example is important in many contexts:

PROPOSITION. *Let U be open in $\mathbf{R}^n$, let $f : U \to \mathbf{R}^m$ be a $\mathbf{C}^1$ function, and suppose that $a \in U$ and $\delta > 0$ are such that $|x - a|, |y - a| \leq \delta$ implies $x, y \in U$. Then for all such x we have the following inequality:*

$$|f(x) - f(y)| \quad \leq \quad \max_{|z-a|\leq\delta} \| \, Df(z) \, \| \cdot |x - y|$$

Proof. Let $h = x - y$, and set $g(t) = f(y + th)$; since open disks are convex we know that $y + th$ also satisfies $|y + th - a| \leq \delta$. Then we have

$$f(x) - f(y) = \int_0^1 g'(t)\, dt$$

and therefore

$$\left|f(x) - f(y)\right| \ \leq \ \int_0^1 \left|g'(t)\right| dt \ .$$

As indicated before, by the Chain Rule we know that

$$g'(t) = \left[Df\left(y + t(x - y)\right)\right](x - y)$$

and therefore we have the estimate

$$\int_0^1 \left|g'(t)\right| dt \quad \leq \quad \max_{0\leq t\leq 1} \| \, Df\left(y + t(x - y)\right) \| \cdot \left|x - y\right|$$

which immediately yields the inequality in the proposition. ∎

Lipschitz conditions

The restriction of a smooth function (say of class $\mathbf{C}^r$) to a compact set satisfies a strong form of uniform continuity that generalizes the matrix inequality $|Ax| \leq \| \, A \, \| \cdot |x|$.

THEOREM. *Let U be open in $\mathbf{R}^n$, let $f : U \to \mathbf{R}^m$ be a $\mathbf{C}^1$ function, and let $K \subset U$ be compact. Then there is a constant $B > 0$ such that*

$$|f(u) - f(v)| \ \leq \ B\,|u - v|$$

for all $u, v \in K$ such that $u \neq v$.

The displayed inequality is called a *Lipschitz condition* for f. This strong form of uniform continuity associates to each $\varepsilon > 0$ a corresponding δ equal to ε/B. An example of a function not satisfying any Lipschitz condition is given by $h(x) = \sqrt{x}$ on the closed unit interval $[0, 1]$ (use the Mean Value Theorem and $\lim_{t\to 0^+} h'(t) = +\infty$). Incidentally, the inverse of this map is a homeomorphism that does satisfy a Lipschitz condition (*e.g.*, we can take $B = 2$).

The inequality

$$\left|u - v\right| \quad \geq \quad \left||u| - |v|\right|$$

for $u, v \in \mathbf{R}$ shows that $f(x) = |x|$ is a function that satisfies a Lipschitz condition but is not $\mathbf{C}^1$.

A *Lipschitz constant* for f on a set K (not necessarily compact) is a number $B > 0$ such that $|f(u) - f(v)| \leq B\,|u - v|$ for all $u, v \in K$ such that $u \neq v$. Note that Lipschitz constants are definitely nonunique; if B is a Lipschitz constant for f on a set K and $C > B$, then C is also a Lipschitz constant for f on a set K.

Proof of theorem. For each $x \in K$ there is a $\delta(x) > 0$ such that $N_{2\delta(x)}(x) \subset U$. By compactness there are finitely many points $x_1, \cdots, x_q$ such that the sets $N_{\delta(x_i)}(x_i)$ cover K. Let B_i be the maximum of $\| Df \|$ for $|y - x_i| \leq \delta(x_i)$. If $B_i = 0$ for all i then $Df = 0$ on an open set containing K and therefore f is constant on K, so that the conclusion of the theorem is trivial. Therefore we shall assume some $B_i > 0$ for the rest of the proof.

By the Mean Value Estimate we know that $y, z \in N_{\delta(x_i)}(x_i)$ implies that $|f(y) - f(z)| \leq B_i|y - z|$.

Let $\eta > 0$ be a Lebesgue number for the open covering of K determined by the sets $N_{\delta(x_i)}(x_i)$, and let $M \subset K \times K$ be the set of all points $(u, v) \in K \times K$ such that $|u - v| \geq \eta/2$. The function $\Delta(u, v) = |u - v|$ is continuous on $K \times K$, and consequently it follows that M is a closed and thus compact subset of $K \times K$. Consider the continuous real-valued function on M defined by

$$h(u, v) = \frac{|f(u) - f(v)|}{|u - v|} \; .$$

Since the denominator is positive on M, this is a continuous function and therefore attains a maximum value A.

Let B be the maximum of the numbers $A, B_1, \cdots, B_k$, and suppose that $(u, v) \in K \times K$. If $(u, v) \in M$, then by the preceding paragraph we have $|f(u) - f(v)| \leq A \cdot |u - v|$. On the other hand, if $(u, v) \notin M$, then $|u - v| < \eta/2$ and thus there some i such that $u, v \in N_{\delta(x_i)}(x_i)$. By the Mean Value Estimate we know that $|f(u) - f(v)| \leq B_i \cdot |u - v|$ in this case. Therefore $|f(u) - f(v)| \leq B \cdot |u - v|$ for all u and v.∎

IV.3 : Inverse Function Theorem

(Edwards, §§ III.2, III.3)

We have already mentioned the general question of recognizing when a 1–1 onto continuous function from one space to another has a continuous inverse. There are also many situations where it is useful to know simply whether a **local inverse** exists. For real valued functions on an interval, the Intermediate Value Property from elementary calculus implies that local inverses exist for functions that are strictly increasing or strictly decreasing (we have not actually proved this yet, however). Since the latter happens if the function has a derivative that is everywhere positive or negative close to a given point, one can use the derivative to recognize very quickly whether local inverses exist in many cases, and in these cases one can even compute the derivative of the inverse function using the standard formula:

$$g = f^{-1} \implies g'(y) = \frac{1}{f'(g(y))}$$

Of course this formula requires that the derivative of f is not zero at the points under consideration.

There is a far-reaching generalization of the single variable inverse function theorem for functions of several real variables. It is covered in many but not all courses on multivariable calculus or undergraduate courses on the theory of functions of a real variable, but even when it is covered the treatment is sometimes incomplete (for example, only worked out for functions of two or at most three variables).

INVERSE FUNCTION THEOREM. *Let U be open in $\mathbf{R}^n$, let $a \in U$, and let $f : U \to \mathbf{R}^n$ be a $\mathbf{C}^r$ map (where $1 \leq r \leq \infty$) such that $Df(a)$ is invertible. Then there is an open set W containing a such that the following hold:*

(i) The restriction of f to W is $1-1$ and its image is an open subset V.

(ii) There is a $\mathbf{C}^r$ inverse map $g : V \to U_0$ such that $g(f(x)) = x$ on U_0.

Proof. The proof given here uses the Contraction Lemma.

It is convenient to reduce the proof to the special case where $a = f(a) = 0$ and $Df(a) = I$. Suppose we know the result in that case. Let $A = Df(a)$, and define f_1 by the formula

$$f_1(x) = A^{-1}\left(f(x+a) - f(a)\right) \ .$$

Then $f_1(0) = 0$ and by the Chain Rule we have $Df_1(0) = I$. Then assuming the conclusion of the theorem is known for f_1, we take W_1, V_1, g_1 as in that conclusion. If we take $W = a + W_1$, $V = AV_1 + f(a)$, and

$$g(z) = g_1\left(A^{-1}\left(y - f(a)\right)\right) + a$$

then it follows immediately that the function $f(y) = Af_1(y-a) + f(a)$ satisfies the conclusions of the theorem.

Since f has a continuous derivative, there is a $\delta > 0$ such that $|x| \leq \delta$ implies $\| Df(x) - I \| < \frac{1}{2}$. For each y on the closed disk D of radius $\delta/2$ about the origin, define a map T on D by the formula $T(x) = x + y - f(x)$, and observe that $T(x) = x$ if and only if $y = f(x)$.

We want to apply the Contraction Lemma to T. The first step is to show that T maps D to itself. Let φ be the function $\varphi(x) = x - f(x)$; then $\varphi(0) = 0$ and $D\varphi = I - Df$, and consequently by the Mean Value Estimate we have that $|\varphi(x)| \leq |x|/2$ if $|x| \leq \delta$. Since $T(x) = y + \varphi(x)$ and $|y| < \delta/2$ it follows that $|T(x)| \leq \delta$ and hence $T(D) \subset D$.

We now need to estimate $|T(x_1) - T(x_0)|$ in terms of $|x_1 - x_0|$. Since T and φ differ by a constant it follows from the Mean Value Estimate that

$$\big|T(x_1) - T(x_0)\big| = \big|\varphi(x_1) - \varphi(x_0)\big| \leq \frac{1}{2}|x_1 - x_0|$$

and therefore the Contraction Lemma implies the existence of a unique point $x \in D$ such that $T(x) = x$, which is equivalent to $f(x) = y$.

Let $g : N_{\delta/2}(0) \to D$ be the inverse map sending a point y to the unique x such that $f(x) = y$. We claim that g is continuous. The first step is to show that $|x_0|,\ |x_1| \leq \delta/2$ implies $|f(x_1) - f(x_0)| \geq \frac{1}{2}|x_1 - x_0|$. To see this, use the identity $f(x) = x - \varphi(x)$ and use the equation and inequalities

$$|f(x_1) - f(x_0)| \geq |x_1 - x_0| - \big|\varphi(x_1) - \varphi(x_0)\big| \geq |x_1 - x_0| - \frac{1}{2}|x_1 - x_0| = \frac{1}{2}|f(x_1) - f(x_0)| \ .$$

If we set $y_i = f(x_i)$ so that $x_i = g(y_i)$ then we have

$$|x_1 - x_0| = |g(y_1) - g(y_0)| \ \leq \ 2 \cdot |y_1 - y_0|$$

and thus g is uniformly continuous.

Let U_0 be the image of g; we claim that U_0 is open. Suppose that $x \in U_0$, so that $f(x) = y$ where $|y| < \delta/2$. Then one can find some $\eta > 0$ so that $|z - x| < \eta$ implies $|f(z)| < \delta/2$ (why?) and the identity $g(f(z)) = z$ then implies that $z \in \mathrm{Image}(g)$. Thus we may take $V = N_{\delta/2}(0)$ and $U_0 = g(V)$.

Finally, we need to show that g is a $\mathbf{C}^r$ function if f is a $\mathbf{C}^r$ function. Given $y \in V$ and k such that $y + k \in V$, write $y = f(x)$ and $y + k = f(x + h)$. Since $\| Df(x) - I \| < 1$ it follows that $Df(x)$ is invertible. Let L be its inverse. Then we have

$$g(y + k) - g(y) - L(k) = h - L(k) = -L\left(f(x + h) - f(x) - Df(x)h\right)$$

and the right hand side is equal to

$$L\left(|h| \cdot \theta(h)\right)$$

where $\lim_{|h| \to 0} \theta(h) = 0$. Since $h = g(y + k) - g(y)$ we know that $|h| \leq 2|k|$ and therefore we also have

$$\lim_{|k| \to 0} \frac{1}{|k|} \cdot L\left(|h| \cdot \theta(h)\right) = 0$$

(where $h = g(y + k) - g(y)$ as above), which shows that g is differentiable at y and satisfies a familiar looking formula:

$$Dg(y) = \left(Df\left(g(y)\right)\right)^{-1}$$

Since the entries of an inverse matrix are rational expressions in the inverse of the original matrix, the continuity of g and the $\mathbf{C}^1$ property of f imply that g is also $\mathbf{C}^1$.

If f is a $\mathbf{C}^r$ function, one can now prove that g is a $\mathbf{C}^s$ function for all $s \leq r$ inductively as follows: Suppose we know that f is $\mathbf{C}^r$ and g is $\mathbf{C}^s$ for $1 \leq s < r$. By the formula for the derivative of g we know that Dg is formed by the composite of g, Df and matrix inversion. We know that g is $\mathbf{C}^s$, that Df is too because f is $\mathbf{C}^{s+1}$ (recall that $s + 1 \leq r$), and that inversion is $\mathbf{C}^\infty$ because its entries are given by rational functions, and therefore it follows that $D[g \circ f]$ is also $\mathbf{C}^s$, which means that $g \circ f$ is $\mathbf{C}^{s+1}$.■

COROLLARY. *Let U and V be open in $\mathbf{R}^n$, and let $f : U \to V$ be $1-1$ onto and $\mathbf{C}^r$ where $1 \leq r \leq \infty$. Then f^{-1} is also $\mathbf{C}^r$.*■

A similar result holds when $r = 0$, but the proof requires entirely different methods which come from algebraic topology.

The Inverse Function Theorem also has the following purely topological consequence for $\mathbf{C}^1$ mappings:

COROLLARY. *Let $f : U \to \mathbf{R}^n$ be a $\mathbf{C}^1$ function ($r \geq 1$), where U is open in $\mathbf{R}^n$, and assume that $Df(x)$ is invertible for all $x \in U$. Then f is open.*■

There is also an extension of this result to the $\mathbf{C}^0$ case provided f is locally 1–1 (this is Brouwer's *Invariance of Domain* Theorem); the proof again requires methods from algebraic topology.

Proof. Let W be open in U and let $x \in W$. Then the Inverse Function Theorem implies that there is an open subset $W_0(x) \subset W$ containing x such that f maps $W_0(x)$ onto an open subset $V(x)$ in $\mathbf{R}^n$. Therefore it follows that

$$f(W) \;=\; \bigcup_x f\left(W_0(x)\right) \;=\; \bigcup_x V_x$$

which is open in $\mathbf{R}^n$.

Examples. Consider the complex exponential mapping f from $\mathbf{R}^2$ to itself sending (x, y) to $(e^x \cos y, e^x \sin y)$. The derivative of this map is invertible at every point but the map is not 1–1 because every nonzero point in $\mathbf{R}^2$ is the image of infinitely many points; specifically, for every (x, y) and integer k we have $f(x, y + 2k\pi) = f(x, y)$.

Another example of this type is the complex square mapping $f : \mathbf{R}^2 - \{0\} \to \mathbf{R}^2$ sending (x, y) to $(x^2 - y^2, 2xy)$, which has the property that $f(-x, -y) = f(x, y)$ for all (x, y); note that if we write $z = u + iv$, then $f(z) = z^2$.

Final Remark. Given a $\mathbf{C}^1$ function $f : U \to \mathbf{R}^n$ with U open in $\mathbf{R}^n$, if we write the coordinate functions of f as $y_1, \cdots y_n$ then $\det Df(p)$ is just the classical *Jacobian function*

$$\frac{\partial(y_1, \cdots, y_n)}{\partial(x_1, \cdots, x_n)}(p)$$

and with this terminology the condition on $Df(p)$ in the Inverse Function Theorem may be rephrased to state that the Jacobian at p is nonzero.

The Implicit Function Theorem

There is a close relation between the Inverse Function Theorem and the standard Implicit Function Theorem from ordinary and multivariable calculus. In its simplest form the Implicit Function Theorem states that **locally** one can solve an equation $F(x, y) = 0$ uniquely for y in terms of x; more precisely, if $F(a, b) = 0$ and the second partial derivative of F is nonzero at (a, b), then on some open interval $(a-\delta, a+\delta)$ there is a unique function $f(x)$ such that $y = f(x) \iff F(x,y) = 0$ (hence $f(a) = b$) and

$$\frac{df}{dx} \quad = \quad - \frac{\left(\frac{\partial F}{\partial x}\right)}{\left(\frac{\partial F}{\partial y}\right)} \quad .$$

Here is a general version of this result:

IMPLICIT FUNCTION THEOREM. *Let U and V be open in $\mathbf{R}^n$ and $\mathbf{R}^m$ respecitvely, and let $f : U \times V \to \mathbf{R}^m$ be a smooth function such that for some $(x,y) \in U \times V$ we have $f(x,y) = 0$ and the partial derivative of f with respect to the last m coordinates is invertible. Then there is an open neighborhood U_0 of x and a smooth function $g : U_0 \to V$ such that $g(x) = y$ and for all $u \in U_0$ we have $f(u,v) = 0$ if and only if $v = g(u)$.*

For the sake of completeness we note that the partial derivative of f with respect to the last m coordinates is the derivative of the function $f^*(v) = f(x,v)$, and that smooth means smooth of class $\mathbf{C}^r$ for some r such that $1 \leq r \leq \infty$.

Proof. Define $h : U \times V \to \mathbf{R}^m \times \mathbf{R}^n$ by $h(u,v) = (f(u,v), u)$. Then the hypotheses imply that $Dh(x,y)$ is invertible, and therefore by the Inverse Function Theorem there is a local inverse

$$k : \text{Int } (\varepsilon D^m) \times U_0 \longrightarrow U \times V \ .$$

Since the second coordinate of $h(u,v)$ is u, it follows that the first coordinate of the inverse $k(z,w)$ is w so that we may write $k(z,w) = (w,\ Q(z,w)\,)$ for some smooth function Q.

On one hand we have $g\,(\,k(z,w)\,) = (z,w)$, but on the other hand we also have

$$g\,(k(z,w)\,) \quad = \quad g\,(w, Q(z,w)) \quad = \quad \big(f\,(w, Q(z,w))\,, w\big).$$

In particular, this means that

$$z \quad = \quad f\big(u, Q(z,u)\big)$$

for all z and u. If we take $g(u) = Q(0,u)$ it follows that $y = g(x)$ and $f(u,v) = 0$ if and only if $v = g(u)$.■

One can use the Chain Rule to calculuate $Dg(u)$ as follows: If $\varphi(u) = f(u, g(u))$ and $\mathbf{p} \in \mathbf{R}^n$, then the Chain Rule yields the formula

$$[D\varphi(u)](\mathbf{p}) \quad = \quad [D_1 f\,(u, g(u)\,)]\,(\mathbf{p}) \ + \ [D_2 f\,(u, g(u)\,)]\,(\,[Dg(u)](\mathbf{p})\,)$$

where D_1 and D_2 refer to partial derivatives with respect to the first and last sets of variables. Since $\varphi = 0$ it also follows that $[D\varphi(u)](\mathbf{p}) = 0$. Furthermore, if u and v are sufficiently close to x and y then the second partial derivative is invertible. Therefore one obtains the formula

$$Dg \quad = \quad - \, (D_2 f)^{-1} \circ D_1 f$$

which generalizes the formula in elementary multivariable calculus.■