UIC Math 549

Differentiable Manifolds—Vector Calculus Background

John Wood

Some sources and inspiration for this treatment are the advanced calculus or analysis books by Dieudonné, Loomis & Sternberg, and Spivak, and notes and books by Milnor.

1. The derivative

DEFINITION. Let $U \subset \mathbb{R}^m$ be an open set, $a \in U$, and $f : U \longrightarrow \mathbb{R}^n$. The map f is differentiable at a if there is a linear map $\lambda \in \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ with

$$\lim_{x \to a} \frac{|f(x) - f(a) - \lambda(x-a)|}{|x-a|} = 0.$$

LEMMA. If there is such a λ it is unique.

PROOF. Let λ and λ_1 both satisfy the definition. Then

$$|(\lambda - \lambda_1)(x-a)| \leq |f(x) - f(a) - \lambda(x-a)| + |-f(x) + f(a) + \lambda_1(x-a)|$$

hence $|(\lambda - \lambda_1)(x-a)|/|x-a| \to 0$ as $x \to a$. For $v \neq 0$, letting $x = a + v \in U$,

$$|(\lambda - \lambda_1)(v)|/|v| = |(\lambda - \lambda_1)(tv)|/|tv| \to 0 \text{ as } t \to 0.$$

Therefore $\lambda(v) = \lambda_1(v)$.

When f is differentiable at a this unique linear map is denoted $Df(a)$.

2. The case $m = n = 1$

Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and assume $f'(a)$ exists. Then

$$\frac{|f(x) - f(a) - f'(a)(x-a)|}{|x-a|} = \left| \frac{f(x) - f(a)}{x-a} - f'(a) \right| \to 0 \text{ as } x \to a$$

so $Df(a)(v) = f'(a)v$. The 1×1-matrix for the linear map $Df(a)$ has entry $f'(a)$.

3. The case $n = 1$ of real-valued functions, partial derivatives

PROPOSITION. If $f : U \longrightarrow \mathbb{R}$ is differentiable at $a \in U \subset \mathbb{R}^m$, then the partial derivatives of f exist at a and determine $Df(a)$.

Proof. Let $e_1, \ldots, e_m$ be the standard orthonormal basis for $\mathbb{R}^m$. Then

$$\lim_{t \to 0} \left| \frac{f(a+te_i) - f(a)}{t} - Df(a)(e_i) \right| = \lim_{t \to 0} \frac{|f(a+te_i) - f(a) - D(f)(a)(te_i)|}{|te_i|} = 0,$$

hence the partial derivative with respect to the ith variable exists:

$$\frac{\partial f}{\partial x_i}(a) = D_i f(a) = Df(a)(e_i) = \lim_{t \to 0} \frac{f(a+te_i) - f(a)}{t}.$$

If $v = \sum_i v_i e_i$, then $Df(a)v = \sum_i D_i f(a) v_i$.

More generally, the directional derivative is defined by

$$D_v f(a) = \lim_{t \to 0} \frac{f(a+tv) - f(a)}{t}.$$

This limit may exist, in some or all directions, even if f is not differentiable at a. The gradient of f at a is the vector $\operatorname{grad} f(a) = \sum_i D_i f(a) e_i$ and, if f is differentiable at a,

$$Df(a)v = D_v f(a) = \operatorname{grad} f(a) \cdot v$$

For f to be differentiable at a it is necessary, but not sufficient, for the partial derivatives to exist at a. It is even necessary, but not sufficient, for the directional derivative to exist at a for all v and to define a linear function. A sufficient condition for f to be differentiable is given by the following theorem, but this condition is not necessary.

Theorem. Let $f : U \longrightarrow \mathbb{R}$, U open in $\mathbb{R}^m$. Suppose the partial derivatives $D_i f$ are each continuous at $a \in U$. Then f is differentiable at a and $Df(a)v = \sum_i D_i f(a) v_i$.

Proof. Given $\varepsilon > 0$ there exists $\delta > 0$ such that

$$|x - a| < \delta \Rightarrow |D_i f(x) - D_i f(a)| < \varepsilon \text{ for all } i.$$

Let $\xi_i = (x_1, \ldots, x_i, a_{i+1}, \ldots, a_m)$; $\xi_0 = a$, $\xi_m = x$. Then $|\xi_i - a| < \delta$ and

$$f(x) - f(a) = \sum_{i=0}^{m} f(\xi_i) - f(\xi_{i-1}).$$

Let $\varphi_i(t) = f(\xi_{i-1} + te_i)$. Then

$$f(\xi_i) - f(\xi_{i-1}) = \varphi_i(x_i - a_i) - \varphi_i(0) = \varphi'(t_i)(x_i - a_i) = D_i f(\xi_{i-1} + t_i e_i)(x_i - a_i)$$

for some t_i with $0 < t_i < x_i - a_i$, by the mean value theorem in one variable. Now

$$\left| f(x) - f(a) - \sum D_i f(a)(x_i - a_i) \right| \le \sum |f(\xi_i) - f(\xi_{i-1}) - D_i f(a)(x_i - a_i)|$$

$$\le \sum |f(\xi_i) - f(\xi_{i-1}) - D_i f(\xi_{i-1} + t_i e_i)(x_i - a_i)| + \sum |\{D_i f(\xi_{i-1} + t_i e_i) - D_i f(a)\}(x_i - a_i)|$$

$$\le 0 + n\varepsilon |x - a|.$$

Hence $\dfrac{|f(x) - f(a) - \lambda(x - a)|}{|x - a|} \to 0$ as $x \to a$ where λ is the linear map defined by $\lambda(v) = \sum D_i f(a) v_i$. Therefore f is differentiable at a.

4. The derivative of linear and bilinear maps

LEMMA. If f is a linear map then $Df(a) = f$.

PROOF. Since f is linear, $f(x) - f(a) - f(x-a) = 0$.

LEMMA. If U, V, W are vector spaces and $\beta : U \times V \longrightarrow W$ is bilinear, then

$$D\beta(a,b)(u,v) = \beta(a,v) + \beta(u,b).$$

PROOF. Note that the map $\ell(a,b)$ defined by $\ell(a,b)(u,v) = \beta(a,v) + \beta(u,b)$ is linear from $U \times V \longrightarrow W$ and

$$\beta(a+u, b+v) - \beta(a,b) - \ell(a,b)(u,v) = \beta(u,v).$$

The norm $|(u,v)| = \sqrt{|u|^2 + |v|^2}$, and $|u||v| \le \max\{|u|^2, |v|^2\} \le |u|^2 + |v|^2$, hence

$$\beta(u,v) = |u||v|\beta(u/|u|, v/|v|) \le |(u,v)|^2 \beta(u/|u|, v/|v|), \text{ for } u \ne 0, v \ne 0.$$

Therefore $|\beta(u,v)|/|(u,v)| \to 0$ as $(u,v) \to (0,0)$.

Examples of bilinear maps $\beta : \mathbb{R}^\ell \times \mathbb{R}^m \longrightarrow \mathbb{R}^n$.

$$\begin{aligned} \ell = m = n = 1, &\qquad \beta(r,s) = rs \\ \ell = 1,\ m = n, &\qquad \beta(r,u) = ru, \\ \ell = m,\ n = 1, &\qquad \beta(u,v) = u \cdot v, \\ \ell = m = n = 3, &\qquad \beta(u,v) = u \times v. \end{aligned}$$

5. A norm on $\mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$

Let $e_1, \ldots, e_m$ be the standard orthonormal basis for $\mathbb{R}^m$ and $\overline{e}_1, \ldots, \overline{e}_n$ be the standard orthonormal basis for $\mathbb{R}^n$. Let $x = \sum_i x_i e_i \in \mathbb{R}^m$, so $x_i = x \cdot e_i$. Let $\ell \in \mathrm{Hom}(\mathbb{R}^m, \mathbb{R}^n)$ and set $\ell_i^j = \ell(e_i) \cdot \overline{e}_j$. Then $\ell(x) = \sum_i x_i \ell(e_i) = \sum_j \sum_i \ell_i^j x_i \overline{e}_j$.

PROPOSITION. If $|\ell_i^j| \le k$ for all i, j, then $|\ell(x)| \le \sqrt{mn}\, k|x|$.

PROOF. By Cauchy's inequality, $|\sum_i \ell_i^j x_i| \le \{\sum_i (\ell_i^j)^2\}^{1/2}|x| \le \sqrt{m}\, k|x|$. Then

$$|\ell(x)| = \left\{ \sum_j \left(\sum_i \ell_i^j x_i \right)^2 \right\}^{1/2} \le \sqrt{mn}\, k|x|.$$

The continuous real-valued function $|\ell(x)|$ is bounded on the compact unit sphere, $\{x : |x| = 1\} \subset \mathbb{R}^m$, and attains its bound.

DEFINITION. For a linear map ℓ, define $\|\ell\| = \sup\{|\ell(x)| : |x| = 1\}$.

COROLLARY. (i) $\quad |\ell(x)| \le \|\ell\|\, |x| \quad$ and $\quad$ (ii) $\quad \|\ell\| \le \sqrt{mn}\, k$.

6. Lipschitz continuity of differentiable functions

Proposition. If $f : U \longrightarrow \mathbb{R}^n$ where U is open in $\mathbb{R}^m$ and f is differentiable at a, then there exist $\delta > 0$ and $k > 0$ such that $|x - a| < \delta \Rightarrow |f(x) - f(a)| \leq k|x - a|$.

Proof. There is a linear map λ such that the function $\varphi(x) = f(x) - f(a) - \lambda(x - a)$ satisfies $|\varphi(x)|/|x - a| \to 0$ as $x \to a$. Therefore there is a $\delta > 0$ such that $|\varphi(x)| \leq |x - a|$ for $|x - a| < \delta$. Then $|f(x) - f(a)| = |\lambda(x - a) + \varphi(x)| \leq (\|\lambda\| + 1)|x - a|$ for $|x - a| < \delta$. Take $k = \|\lambda\| + 1$.

The conclusion of the Proposition is called Lipschitz continuity at a; it implies that f is continuous at a.

7. The chain rule

Theorem. If $a \in U \subset \mathbb{R}^m$, $b \in V \subset \mathbb{R}^n$, $f : U \longrightarrow V$, $f(a) = b$, $g : V \longrightarrow \mathbb{R}^p$, f is differentiable at a, and g is differentiable at b; then $g \circ f$ is differentiable at a and

$$D(g \circ f)(a) = Dg(b) \circ Df(a).$$

Proof. (See Spivak, p. 19.) Let $\lambda = Df(a)$, $\mu = Dg(b)$ and set

$$\begin{aligned}\varphi(x) &= f(x) - f(a) - \lambda(x - a)\\ \psi(y) &= g(y) - g(b) - \mu(y - b)\\ \rho(x) &= g(f(x)) - g(b) - \mu(\lambda(x - a)).\end{aligned}$$

We have

(i) $$|\varphi(x)|/|x - a| \to 0 \text{ as } x \to a,$$

(ii) $$|\psi(y)|/|y - b| \to 0 \text{ as } y \to b.$$

From the definitions,

$$\begin{aligned}\rho(x) &= g(f(x)) - g(b) - \mu(f(x) - f(a) - \varphi(x))\\ &= [g(f(x)) - g(b) - \mu(f(x) - f(a))] + \mu(\varphi(x))\\ &= \psi(f(x)) + \mu(\varphi(x)).\end{aligned}$$

First $|\mu(\varphi(x))| \leq \|\mu\||\varphi(x)|$, so by (i) $|\mu(\varphi(x))|/|x - a| \to 0$ as $x \to a$.
Second, by Proposition 6, there are $k > 0, \delta > 0$ such that

$$|x - a| < \delta \Rightarrow |f(x) - f(a)| \leq k|x - a|.$$

By (ii), for any $\varepsilon > 0$ there is a $\delta_1 > 0$ such that

$$|f(x) - f(a)| < \delta_1 \Rightarrow |\psi(f(x))| < \varepsilon|f(x) - f(a)|.$$

So for $0 \neq |x - a| < \min\{\delta, \delta_1/k\}$ we have $|\psi(f(x))|/|x - a| < \varepsilon k$. Hence $|\rho(x)|/|x - a| \to 0$ as $x \to a$ which gives the result.

8. Sample computations

(a) Let $f(x) = x \cdot x = \beta \circ \Delta(x)$ where $\Delta(x) = (x, x)$ is linear and $\beta(x, y) = x \cdot y$. Then

$$Df(a)(u) = D\beta(\Delta(a)) \circ D\Delta(a)(u) = D\beta(a, a)(u, u) = \beta(a, u) + \beta(u, a).$$

Since β is symmetric, $Df(a)(u) = 2a \cdot u$ and $\operatorname{grad} f(a) = 2a$.

If $g(x) = |x - p| = \sqrt{f(x - p)}$,

$$Dg(a)(u) = \frac{1}{2\sqrt{f(a - p)}} Df(a - p)(u) = \frac{a - p}{|a - p|} \cdot u \text{ for } a \neq p.$$

So, for $x \neq p$, $\operatorname{grad} g(x) = \dfrac{x - p}{|x - p|}$, the unit vector at x pointing away from p.

(b) The derivative of a sum.

Lemma. Let f and $g : U \longrightarrow R^n$ be differentiable at $a \in U \subset \mathbb{R}^m$.

Define $(f, g) : U \longrightarrow \mathbb{R}^n \times \mathbb{R}^n$ by $(f, g)(x) = (f(x), g(x))$. Then

$$D(f, g)(a) = (Df, Dg)(a).$$

Proof. Let $\lambda = Df(a)$, $\varphi(x) = f(x) - f(a) - \lambda(x - a)$, $\mu = Dg(a)$, and $\psi(x) = g(x) - g(a) - \mu(x - a)$. Then $(\varphi, \psi)(x) = (f, g)(x) - (f, g)(a) - (\lambda, \mu)(x - a)$ and

$$\frac{|(\varphi, \psi)(x)|}{|x - a|} = \sqrt{\frac{|\varphi(x)|^2}{|x - a|^2} + \frac{|\psi(x)|^2}{|x - a|^2}} \to 0 \text{ as } x \to a.$$

Define the linear map $s : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ by $s(y_1, y_2) = y_1 + y_2$. Now $(f + g)(x) = f(x) + g(x) = s \circ (f, g)(x)$. Hence the derivative of a sum is the sum of the derivatives:

$$D(f + g) = Df + Dg.$$

(c) The set $M(n)$ of $n \times n$-matrices is an n^2-dimensional vector space under addition and scalar multiplication and a ring under matrix multiplication. Let $\beta(A, B) = AB$ and $t(A) = A^t$ be the transpose. The maps t and (I, t) are linear as maps of vector spaces where I is the identity linear map. On products t satisfies $t(AB) = t(B)t(A)$. Define $f : M(n) \longrightarrow M(n)$ by $f(A) = AA^t$, so $f = \beta \circ (I, t)$

Let $O(n) \subset M(n)$ be the orthogonal group, $O(n) = \{A : f(A) = I\}$. Thus $A \in O(n)$ means A is invertible and $A^t = A^{-1}$.

Exercise. This is the computational part of a proof that $O(n)$ is a manifold of dimension $n(n - a)/2$. Show:

$f(A)$ is symmetric, $f(A) = t(f(A))$.

$Df(A)(M) = AM^t + MA^t$.

If $A \in O(n)$, then $Df(A)$ maps $M(n)$ onto the vector space of symmetric matrices. [Hint: Given a symmetric S, take $M = \frac{1}{2}SA$.]

9. Differentiability of maps to $\mathbb{R}^n$

The results of §3 extend to maps to $\mathbb{R}^n$.

PROPOSITION. If $f : U \longrightarrow \mathbb{R}^n$ is differentiable at $a \in U$ then the partial derivatives of the components $D_i f_j$ exist at a and are the entries in the matrix representing $Df(a)$. If all the partials are continuous at a then f is differentiable at a.

PROOF. (See Spivak, p. 21, and for notation §§3, 5.) Define the linear projection map $\pi_j : \mathbb{R}^n \longrightarrow \mathbb{R}$ by $\pi_j(y) = y \cdot \overline{e}_j$. The jth component of f is $f_j = \pi_j \circ f$, $f(x) = \sum_j f_j(x)\overline{e}_j$ and

$$Df_j(a) = D\pi_j(f(a)) \circ Df(a) = \pi_j \circ Df(a).$$

The partial derivatives $\frac{\partial f_j}{\partial x_i}(a) = D_i f_j(a) = Df_j(a)(e_i) = Df(a)(e_i) \cdot \overline{e}_j$.

If $u = \sum_i u_i e_i$, then $Df(a)u = \sum_j \sum_i D_i f_j(a) u_i \overline{e}_j$.

Introducing the Jacobian matrix we write $Df(a)u$ as a matrix product:

$$Df(a)u = \begin{pmatrix} Df_1(a)u \\ \vdots \\ Df_n(a)u \end{pmatrix} = \begin{pmatrix} D_1 f_1(a) & \dots & D_m f_1(a) \\ \vdots & & \vdots \\ D_1 f_n(a) & \dots & D_m f_n(a) \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}.$$

If all the partials are continuous at a, by §3 each $D_i f(a)$ exists and by §8(b) $Df(a)$ exists.

When $m = 1$, $f(t)$ is a path in $\mathbb{R}^n$ and we define the velocity vector $f'(t) = Df(t)(e_1)$.

10. Mean value theorems

PROPOSITION. If $U \subset \mathbb{R}^m$ is convex, $f : U \longrightarrow \mathbb{R}$ is differentiable, and $a, x \in U$, then $f(x) - f(a) = Df(\zeta)(x - a)$ where $\zeta = a + t_0(x - a)$ for some $0 < t_0 < 1$.

PROOF. Let $\varphi(t) = f(a + t(x - a))$. By the chain rule $\varphi'(t) = Df(a + t(x - a))(x - a)$. By the one-variable mean value theorem

$$f(x) - f(a) = \varphi(1) - \varphi(0) = \varphi'(t_0) = Df(\zeta)(x - a)$$

where $\zeta = a + t_0(x - a)$ for some $0 < t_0 < 1$.

COROLLARY. If $\|Df(\zeta)\| \le k$ for any $\zeta \in U$, then $|f(x) - f(a)| \le k|x - a|$.

This follows from the Proposition and Corollary §5(i).

The Proposition is not true in general for maps to $\mathbb{R}^n$, $n > 1$. For example let $f : \mathbb{R} \longrightarrow \mathbb{R}^3$ describe a helix about the vertical axis and take x vertically above a. Then $x - a$ points straight up while $Df(t)(u)$ never does. The following Theorem extends the result of the Corollary to maps to $\mathbb{R}^n$. It says f is Lipschitz continuous *on* U.

THEOREM. If $U \subset \mathbb{R}^m$ is convex, $f : U \longrightarrow \mathbb{R}^n$ is differentiable on U, $a, x \in U$, and $\left|\dfrac{\partial f_j}{\partial x_i}\right| \le \dfrac{k}{\sqrt{mn}}$ on U for all i, j, then $|f(x) - f(a)| \le k|x - a|$.

PROOF. By the Proposition $f_j(x) - f_j(a) = Df_j(\zeta_j)(x - a)$. By §5 applied to the real-valued function f_j, $\|Df_j(\zeta_j)\| \le \frac{k}{\sqrt{n}}$. By the Corollary, $|f_j(x) - f_j(a)| \le \frac{k}{\sqrt{n}}|x - a|$. Then $|f(x) - f(a)| \le k|x - a|$ as in §5.

10a. Alternate proof of the mean value theorem

In §10 we used the one-variable mean value theorem. The following proof gives both the Corollary and Theorem above without assuming the one-variable theorem and does not depend on bounds on the partial derivatives. See Loomis & Sternberg, p. 148, or Dieudonné, p. 153.

THEOREM. Let $f : [a,b] \longrightarrow \mathbb{R}^n$ be continuous on $[a,b]$ and differentiable on (a,b). Assume $|f'(t)| \leq k$ for $a < t < b$, where (see §9) $f'(t) = D_1 f(t)(e_1)$. Then

$$|f(b) - f(a)| \leq k(b-a).$$

PROOF. Fix $\varepsilon > 0$. Let $A = \{x \in [a,b] : |f(x) - f(a)| \leq (k+\varepsilon)(x-a) + \varepsilon\}$.

(1) Since f is continuous at a there is a $\delta > 0$ such that

$$|f(x) - f(a)| \leq \varepsilon \text{ for } a \leq x < a + \delta$$

so $x \in A$ for some $x > a$.

(2) Set $\ell = \sup A$. Either $\ell \in A$ or for any $\delta > 0$ there is a t with $\ell - \delta < t \leq \ell$ and $t \in A$. But then, by the continuity of f at ℓ, $\ell \in A$.

(3) If $\ell < b$ then $f'(\ell)$ exists and $|f'(\ell)| \leq k$. Hence there is a $\delta > 0$ such that

$$\ell \leq t < \ell + \delta \Rightarrow |f(t) - f(\ell)| \leq (k+\varepsilon)(t-\ell).$$

Then

$$\begin{aligned} |f(t) - f(a)| &\leq |f(t) - f(\ell)| + |f(\ell) - f(a)| \\ &\leq (k+\varepsilon)(t-\ell) + (k+\varepsilon)(\ell - a) + \varepsilon \\ &= (k+\varepsilon)(t-a) + \varepsilon. \end{aligned}$$

and hence $t \in A$ for some $t > \ell$, a contradiction. Therefore $\ell = b$ and, as in (2), $b \in A$.

Since $\varepsilon > 0$ is arbitrary, $|f(b) - f(a)| \leq k(b-a)$.

COROLLARY. Let $U \subset \mathbb{R}^m$ be convex, $a, b \in U$, $f : U \longrightarrow \mathbb{R}^n$ be differentiable, and assume $\|Df(x)\| \leq k$ for $x \in U$. Then

$$|f(b) - f(a)| \leq k|b-a|.$$

PROOF. Define $c : \mathbb{R} \longrightarrow \mathbb{R}^n$ by $c(t) = tb + (1-t)a$. Then $c'(t) = b - a$ and $f \circ c(1) - f \circ c(0) = f(b) - f(a)$. For $0 \leq t \leq 1$, $c(t) \in U$ and $D(f \circ c)(t)(e_1) = Df(c(t))(b-a)$, so $|(f \circ c)'(t)| \leq \|Df(c(t))\|\,|b-a| \leq k|b-a|$. The result follows from the Theorem.

11. The inverse function theorem

Definition. A function $f : U \longrightarrow \mathbb{R}^n$ is said to be of class C^1 if the partial derivatives exist and are continuous everywhere on U, f is of class C^k if the partial derivatives of orders k and less are continuous, and f is C^∞ if it is C^k for all positive integers k.

Theorem. Given $a \in U \subset \mathbb{R}^n$, U open, and a C^1 function $f : U \longrightarrow \mathbb{R}^n$ with $f(a) = b$ such that $Df(a)$ is invertible, there are neighborhoods V of a, $V \subset U$, and W of b and a unique C^1 map $g : W \longrightarrow V$ such that the restriction $f|V$ and g are inverses. The derivative of g is $Dg(y) = Df(g(y))^{-1}$. Further, if f is C^k $(1 \le k \le \infty)$ then g is also.

Plan. The map g will need to satisfy $g(b) = a$. Let $g_0(y) = a$ be a first approximation to g. Since $Df(a)$ is invertible, the linear approximation to f, $y = f(x) \sim f(a) + Df(a)(x-a)$, can be solved for x. Let $g_1(y)$ be this solution: $g_1(y) = a + Df(a)^{-1}(y-b)$. We will define iteratively a sequence of functions $\{g_n\}$ converging to the local inverse of f.

Proof. (1) Define $F(x,y) = x + Df(a)^{-1}(y - f(x))$ on $U \times \mathbb{R}^n$. Let $D_1F(a,b)$ denote the derivative of the function $x \mapsto F(x,b)$ at $x = a$. Then

$$\begin{aligned} F(a,b) &= a + Df(a)^{-1}(b - f(a)) = a, \\ D_1F(x,y) &= I - Df(a)^{-1} \circ Df(x), \text{ and} \\ D_1F(a,y) &= I - Df(a)^{-1} \circ Df(a) = 0. \end{aligned}$$

$D_1F(x,y)$ does not depend on y and is the zero map for $x = a$. Hence for x near a, $Df(x)$ is invertible and the entries in matrix $D_1F(x,y)$ are small. Choose $k > 0$ so that:

(i) $$\overline{B_k(a)} \subset U \text{ and } Df(x) \text{ is invertible for } x \in \overline{B_k(a)}, \text{ and}$$
$$\|D_1F(x,y)\| \le \frac{1}{2} \text{ for } x \in \overline{B_k(a)}. \quad \text{Then}$$

(ii) $$x, \xi \in \overline{B_k(a)} \Rightarrow |F(x,y) - F(\xi,y)| \le \frac{1}{2}|x - \xi|$$

using the mean value theorem for the function $x \mapsto F(x,y)$. Since

$$|F(a,y) - a| = |Df(a)^{-1}(y-b)| \le \|Df(a)^{-1}\| \, |y-b|,$$

if we set $\delta = \dfrac{k}{2\|Df(a)^{-1}\|}$ we have:

(iii) $$y \in B_\delta(b) \Rightarrow F(a,y) \in B_{k/2}(a)$$

and the same implication for the closed balls.

(2) Let $\mathcal{F}$ be the set of continuous functions $h : \overline{B_\delta(b)} \longrightarrow \overline{B_k(a)}$ such that $h(b) = a$. For $h \in \mathcal{F}$ define $Th(y) = F(h(y), y)$. Then $Th(b) = F(a,b) = a$. For $y \in \overline{B_\delta(b)}$,

$$\begin{aligned} |Th(y) - a| &= |F(h(y),y) - a| \\ &\le |F(h(y),y) - F(a,y)| + |F(a,y) - a| \\ &\le \frac{1}{2}|h(y) - a| + \frac{k}{2} \le k \quad \text{by (ii) and (iii).} \end{aligned}$$

Hence $Th(y) \in \overline{B_k(a)}$ so $Th \in \mathcal{F}$ and $T : \mathcal{F} \longrightarrow \mathcal{F}$. The same argument, using the open version of (iii), shows $y \in B_\delta(b) \Rightarrow T\gamma(y) \in B_k(a)$.

(3) T has a fixed point.

Define a sequence of functions in $\mathcal{F}$ by $g_0(y) = a$ and $g_{n+1}(y) = Tg_n(y) = F(g_n(y), y)$. Note that g_1 is as defined in the plan. To shorten notation, temporarily fix y and set $x_n = g_n(y)$. We have $x_0 = a$, $x_1 = F(a, y)$, and by (iii) $|x_1 - x_0| \le k/2$.

$$|x_{n+1} - x_n| = |F(x_n, y) - F(x_{n-1}, y)| \le \frac{1}{2}|x_n - x_{n-1}| \le \cdots \le \frac{1}{2^n}|x_1 - x_0| \le \frac{k}{2^{n+1}},$$

$$|x_m - x_n| \le |x_m - x_{m-1}| + \cdots + |x_{n+1} - x_n| \le \left(\frac{1}{2^m} + \cdots + \frac{1}{2^{n+1}}\right)k < \frac{k}{2^n},$$

for $n < m$. Therefore $\{x_n\}$ is a Cauchy sequence.

Let $x = \lim x_n$. Since each $x_n \in B_k(a)$, $x \in \overline{B_k(a)}$. Define the map

$$g : \overline{B_\delta(b)} \longrightarrow \overline{B_k(a)} \quad \text{by} \quad g(y) = x = \lim_{n\to\infty} g_n(y).$$

Since $|g(y) - g_n(y)| \le \dfrac{k}{2^n}$, the sequence $\{g_n\}$ converges uniformly on $\overline{B_\delta(b)}$, so g is continuous and $g \in \mathcal{F}$. Since F is continuous, $Tg = g$:

$$g(y) = \lim g_n(y) = \lim F(g_n(y), y) = F(\lim g_n(y), y) = F(g(y), y) = Tg(y).$$

(4) g is a unique local inverse of f.

Set $W = B_\delta(b)$ and $V = B_k(a) \cap f^{-1}(W) \subset U$. V and W are neighborhoods of a and b respectively. If $y \in W$, by (3) $Tg(y) = g(y)$ and by the definition of Tg, $g(y) = g(y) + Df(a)^{-1}(y - f(g(y)))$. Hence $f(g(y)) = y$. Then by (2), $g(y) \in V$, $g : W \longrightarrow V$, and $f \circ g = 1_W$.

If $x, \xi \in V$ and $f(x) = f(\xi) = y \in W$, then $F(x, y) = x$, and $F(\xi, y) = \xi$. By (ii) $|x - \xi| \le \frac{1}{2}|x - \xi|$, hence $x = \xi$. Therefore f is one-to-one on V. If $x \in V$, let $y = f(x) \in W$ and let $\xi = g(f(x)) \in V$. Now $f(\xi) = f(g \circ f(x)) = f \circ g(f(x)) = f(x)$. Therefore $x = \xi$, $g(f(x)) = x$, and $g \circ f = 1_V$.

Let h be another inverse of f with $h(b) = a$. Let both h and g be defined on $W_1 \subset W$, and set $V_1 = B_k(a) \cap f^{-1}(W_1) \subset V$. For $y \in W_1$, let $x = g(y)$, and $\xi = h(y)$. Since g and h are right inverses of f, $f(x) = f(\xi)$. Since f is 1-1, $x = \xi$ and hence $g = h$ on W_1.

(5) g is Lipschitz continuous.

Let $g(y) = x$, $g(\eta) = \xi$ for $y, \eta \in B_\delta(b)$. Since $g = Tg$, $x = F(x, y)$ and $\xi = F(\xi, \eta)$. Then

$$\begin{aligned} |x - \xi| &= |F(x, y) - F(\xi, \eta)| \\ &\le |F(x, y) - F(\xi, y)| + |F(\xi, y) - F(\xi, \eta)| \\ &\le \frac{1}{2}|x - \xi| + |Df(a)^{-1}(y - \eta)| \end{aligned}$$

Therefore $\frac{1}{2}|x - \xi| \le \|Df(a)^{-1}\|\,|y - \eta|$ and hence $|g(y) - g(\eta)| \le 2\|Df(a)^{-1}\|\,|y - \eta|$.

(6) g is differentiable.

Since f is C^1 and, by (i) $Df(\xi)$ is invertible for $\xi \in \overline{B_k(a)}$, we can choose κ so that

$$\|Df(\xi)^{-1}\| \leq \kappa \text{ for } \xi \in \overline{B_k(a)}.$$

Let

$$\varphi(x) = f(x) - f(\xi) - Df(\xi)(x-\xi).$$

Then $|\varphi(x)|/|x-\xi| \to 0$ as $x \to \xi,$ so for any $\varepsilon > 0, |\varphi(x)| \leq \varepsilon|x-\xi|$ for x near ξ.
Let

$$\begin{aligned}\psi(y) &= g(y) - g(\eta) - Df(\xi)^{-1}(y-\eta) \\ &= g(y) - g(\eta) - Df(\xi)^{-1}\{\varphi(x) + Df(\xi)(x-\xi)\} \\ &= g(y) - g(\eta) - (x-\xi) - Df(\xi)^{-1}(\varphi(x)) \\ &= -Df(\xi)^{-1}(\varphi(x)).\end{aligned}$$

Then

$$\begin{aligned}|\psi(y)| &\leq \kappa|\varphi(x)| \leq \kappa\varepsilon|x-\xi| \text{ for } x \text{ near } \xi, \\ &\leq 2\kappa^2\varepsilon|y-\eta| \text{ for } y \text{ near } \eta \text{ by (5).}\end{aligned}$$

Hence $|\psi(y)|/|y-\eta| \to 0$ as $y \to \eta$. Therefore g is differentiable at η and $Dg(\eta) = Df(g(\eta))^{-1}$.

(7) If f is C^k so is g.

We can write Dg as the composition $Dg = i \circ Df \circ g$ where $i(A) = A^{-1}$ is matrix inversion.

$$B_\delta(b) \xrightarrow{g} U \xrightarrow{Df} G\ell(n) \xrightarrow{i} G\ell(n),$$

where g is continuous, f is C^k so that Df is C^{k-1}, and i is C^∞ by Cramer's rule. Since g is continuous, the composition, Dg is continuous, so g is C^1. Now if g is C^j for any $j < k$, then similarly, Dg is C^j, and g is C^{j+1}. By induction g is C^k, for $1 \leq k \leq \infty$.

This completes the proof of the inverse function theorem.

12. Applications of the inverse function theorem

Implicit Function Theorem. Let $(a,b) \in \mathbb{R}^k \times \mathbb{R}^n$. Let f be a C^1 function from a neighborhood of (a,b) to $\mathbb{R}^n$ with $f(a,b) = c$. Let $D_2f(a,b)$, the derivative of the function $y \mapsto f(a,y)$, be invertible.

Then there are neighborhoods $a \in U \subset \mathbb{R}^k$, $(a,b) \in V \subset \mathbb{R}^k \times \mathbb{R}^n$, and $c \in W \subset \mathbb{R}^n$ and a C^1 function $g : U \longrightarrow \mathbb{R}^n$ such that $f(V) \subset W$ and

$$\begin{gathered}(x,y) \in V \text{ and } f(x,y) = c \iff x \in U \text{ and } y = g(x), \\ Dg(x) = -D_2f(x,g(x))^{-1} \circ D_1f(x,g(x)).\end{gathered}$$

Further there is a C^1 diffeomorphism $G : U \times W \longrightarrow V$ such that, defining

$$g_w(x) = \pi_2 \circ G(x, w), \qquad \text{we have} \qquad f(x, y) = w \iff y = g_w(x).$$

The function $\varphi_w : U \longrightarrow V$ define by $\varphi_w(x) = G(x, w)$ parameterizes the level surface

$$f^{-1}(w) = \{(x, y) \in V : f(x, y) = w\}.$$

PROOF. Define F on the domain of f with values in $\mathbb{R}^k \times \mathbb{R}^n$ by $F(x, y) = (x, f(x, y))$. Then $F(a, b) = (a, c)$ and the Jacobian matrix of $DF(x, y)$ is

$$\begin{pmatrix} I & 0 \\ L & M \end{pmatrix}$$

where

$$L = D_1 f = \frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_k)} \qquad \text{and} \qquad M = D_2 f = \frac{\partial(f_1, \ldots, f_n)}{\partial(y_1, \ldots, y_n)}.$$

Since $M(a, b)$ is invertible, $DF(a, b)$ is invertible.

The inverse function theorem gives a map G which we may assume is defined on a product neighborhood $U \times W \subset \mathbb{R}^k \times \mathbb{R}^n$ of (a, c). Let $V = G(U \times W)$. Then $F|V$ and $G|U \times W$ are inverses. If $(x, y) \in V$ and $F(x, y) = (x, f(x, y)) = (x, w) \in U \times W$, then $G(x, w) = (x, y)$ and $f(x, y) = w$. Define $g_w(x) = \pi_2 \circ G(x, w) = y$. Then $f(x, g_w(x)) = f(x, y) = w$. For the case $f(x, y) = c$, take $g = g_c$.

Since F has a C^1 inverse on V, it follows that DF is invertible on V and, from the form of its Jacobian matrix, that the matrix $M(x, y)$ of $D_2 f(x, y)$ is also invertible. As a composition, $g_w(x)$ is differentiable. Differentiating $f(x, g_w(x)) = w$ with respect to x using the chain rule we get

$$D_1 f(x, g_w(x)) + D_2 f(x, g_w(x)) \circ D g_w(x) = 0, \quad \text{hence}$$
$$D g_w(x) = -D_2 f(x, g_w(x))^{-1} \circ D_1 f(x, g_w(x)).$$

Notice that V is not a product, the slice $\{y \in \mathbb{R}^n : (x, y) \in V\}$ depends on x.

PROPOSITION 1. Let $p \in \mathbb{R}^m$ and let f be a C^1 map on a neighborhood of p to $\mathbb{R}^n$, $m \geq n$, with $Df(p)$ surjective. Then there is a neighborhood $p \in V \subset \mathbb{R}^m$ and a diffeomorphism $h : U \longrightarrow V$, U open in $\mathbb{R}^m$, such that $f \circ h(x_1, \ldots, x_m) = (x_{m-n+1}, \ldots, x_m)$ or $f \circ h = \pi_2$.

PROOF. Let $m = k + n$. Since $Df(p)$ is surjective we can reorder the variables, *i.e.* the coordinates of $\mathbb{R}^m$, $x_1, \ldots, x_m$, so that the Jacobian matrix of derivatives with respect to the last n variables is invertible. Then the implicit function theorem applies: the map $F(x) = (x_1, \ldots, x_k, f(x))$ restricted to a neighborhood V of a has an inverse $h : U \longrightarrow V$. Then $F \circ h(z) = z$ and $f \circ h = \pi_2 \circ F \circ h = \pi_2$.

PROPOSITION 2. Let $a \in U \subset \mathbb{R}^m$ be open and $f : U \longrightarrow \mathbb{R}^n$ be a C^1 map, $m \leq n$, with $Df(a)$ injective. Then there are neighborhoods $a \in U_1 \subset U$, $V \subset \mathbb{R}^n$ with $f(U_1) \subset V$, and $b \in W \subset \mathbb{R}^n$ and a diffeomorphism $h : V \longrightarrow W$ such that $h \circ f(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0)$.

PROOF. The Jacobian matrix of $Df(a)$ has an invertible $m \times m$ submatrix A. We may permute the coordinate functions, $f_1, \ldots, f_n$, *i.e.* the coordinates in the range $\mathbb{R}^n$, so that the first m rows of the Jacobian of f are an invertible matrix A.

Define $F : U \times \mathbb{R}^{n-m} \longrightarrow \mathbb{R}^n$ by

$$F(x_1, \ldots, x_n) = f(x_1, \ldots, x_m) + (0, \ldots, 0, x_{m+1}, \ldots, x_n)$$

Then $F(a, 0) = f(a) + 0 = b$ and

$$DF(a, 0) = \begin{pmatrix} A & 0 \\ B & I \end{pmatrix}$$

which is invertible. By the inverse function theorem there are neighborhoods $(a, 0) \in V \subset U \times \mathbb{R}^{n-m}$ and $b \in W \subset \mathbb{R}^n$ and a map $h : W \longrightarrow V$ inverse to $F|V : V \longrightarrow W$.

Set $i(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0)$, so $F \circ i = f$. Let $U_1 = i^{-1}(V)$. On U_1

$$h \circ f = h \circ F \circ i = i.$$

Think of (h, W) as a new coordinate chart for $\mathbb{R}^n$ with respect to which the map f has the simplest possible form: $h \circ f = i$.

It follows that $f|U_1$ is a homeomorphism onto its image in the induced topology. That is $\mathcal{O}$ is open in U_1 if and only if $f(\mathcal{O})$ is the intersection with $f(U_1)$ of an open set in $\mathbb{R}^n$.