Calculus Review - Topology

Definition

Given $r > 0$ and $x \in \mathbb{R}^n$, the *open ball* of radius r and centre x is the set

$$B_r(x) = \{y \in \mathbb{R}^n : |x - y| < r\}.$$

The *closed ball* of radius r and centre x is the set

$$\bar{B}_r(x) = \{y \in \mathbb{R}^n : |x - y| \leq r\}.$$

The *sphere* of radius r and centre x is the set

$$\mathbb{S}_r^{n-1}(x) = \{y \in \mathbb{R}^n : |x - y| = r\}.$$

Definition

For $x, y \in \mathbb{R}^n$, the distance $|x - y|$ is defined to be

$$|x - y| = \sqrt{(x^1 - y^1)^2 + \cdots + (x^n - y^n)^2}$$

where $x = (x^1, \ldots, x^n)$ and $y = (y^1, \ldots, y^n)$.

- The open ball is the set of points of distance to x strictly less than r.
- The closed ball is the set of points of distance to x less than or equal to r.
- The sphere is the set of points of distance to x equal to r.

It is sometimes said that analysis is simply applications of the triangle inequality:

$$|x - y| \leq |x - z| + |z - y|.$$

Definition

A set $U \subset \mathbb{R}^n$ is said to be *open* provided for every $x \in U$, there exists an $r = r(x)$ such that

$$B_r(x) \subseteq U.$$

A set C is *closed* if it's complement,

$$\mathbb{R}^n \backslash C := \{y \in \mathbb{R}^n : y \notin C\}$$

is open.

- By this definition, open balls are open, closed balls are closed and spheres are closed.
- Given any point of an open set, we can always move *uniformly* a little in any direction and remain in the open set.

Definition

A set $S \subseteq \mathbb{R}^n$ is *bounded* if there exists an $x \in \mathbb{R}^n$ and an $r > 0$ such that $S \subseteq B_r(x)$.
A set $K \subseteq \mathbb{R}^n$ is *compact* if it is closed and bounded.

- $S \subseteq \mathbb{R}^n$ is bounded iff for every $x \in \mathbb{R}^n$ there exists an $r = r(x)$ such that $S \subseteq B_r(x)$.
- A set $K \subseteq \mathbb{R}^n$ is compact if and only if for every *open cover* $\{U_\alpha\}$, there exists a *finite subcover*.
  - An *open cover* is a collection of open sets $\{U_\alpha\}$ such that $K \subseteq \cup_\alpha U_\alpha$.
  - A *finite subcover* is a cover by finitely many $U_{\alpha_1}, \cdots, U_{\alpha_N}$
  - Equivalent for $\mathbb{R}^n$.

Calculus Review - Limits

The fundamental concept in calculus is that of limits.

Definition

A sequence $(x_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}^n$ *converges* to $x \in \mathbb{R}^n$ if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that $(x_n)_{n\geq N} \subseteq B_\epsilon(x)$. We write $\lim_{n\to\infty} x_n = x$.

The condition for convergence to x says that $|x - x_n| < \epsilon$ for $n \geq N$.

To prove a limit converges, we need to know what the limit should be, and then prove the sequence converges to that limit. But how can we know what the limit should be? The next idea allows us to avoid the issue entirely and prove a sequence is convergent using only properties of the sequence itself, and without guessing what the limit should be.

Definition

The sequence (x_n) is *Cauchy* if for every $\epsilon > 0$, there exists a $N \in \mathbb{N}$ such that $(x_m)_{m\geq N} \subseteq B_\epsilon(x_n)$ for every $n \geq N$.

The condition to be a Cauchy sequence says that $|x_n - x_m| < \epsilon$ for $m, n \geq N$.

How does this help us prove a sequence is convergent?

Theorem: Completeness

A sequence is convergent if and only if it is Cauchy.

We won't prove this here as it's a standard result in analysis. Note that the theorem does not give us any idea of what the limit should be, just that it exists.

Continuity Review - Continuity

From limits we can introduce continuity.

Definition: Sequential Continuity

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at $x \in \mathbb{R}^n$ if for *every* sequence (x_n) with $\lim_{n\to\infty} = x$ we have $\lim_{n\to\infty} f(x_n) = f(x)$.

The above definition is for limits of sequences, which are discrete objects. The continuous version is as follows.

Definition: ϵ-δ Continuity

Write

$$\lim_{x\to x_0} f(x) = y$$

provided for every $\epsilon > 0$, there exists a $\delta > 0$ such that $f(B_\delta(x_0)) \subseteq B_\epsilon(y)$.
Then f is continuous at x_0 if $\lim_{x\to x_0} f(x) = f(x_0)$.

There is yet another version of continuity. The previous definitions deal with continuity at a single point which is a local property. A global property would be for a function to be continuous at every point. There is another very useful way to define global continuity.

Definition: Topological Contitnuity

The function f is continuous (at every x_0) if $f^{-1}(V)$ is an open set for every open set $V \subseteq \mathbb{R}^m$.

Now we have three definitions. Fortunately we have the following theorem:

Theorem

A function is sequentially continuous at x_0 if and only if it is ϵ-δ continuous at x_0. A function is (sequentially/ϵ-δ) continuous at every point if and only if it is topologically continuous.

The theorem allows us to simply refer continuity at a point x_0 (with no mention of sequential or ϵ-δ). Similarly we may talk of a continuous function, meaning continuity at all points, or equivalently, meaning topologically continuity.

- The first definition requires that $f(x_n) \to f(x)$ for *every* sequence.
- The condition in the second definition that $f(B_\delta(x_0)) \subseteq B_\epsilon(y)$ is the same thing as $|f(x) - f(x_0)| < \epsilon$ whenever $|x - x_0| < \delta$.
- The second definition says that given *any tolerance* $\epsilon > 0$, there is an *adjustment* $\delta > 0$ so that provided we are sufficiently close to x_0 (i.e. $|x - x_0| < \delta$), then $f(x)$ is within the desired tolerance of $f(x_0)$ (i.e. $|f(x) - f(x_0)| < \epsilon$.
- The equivalence of the first and second definitions is a standard exercise in analysis using the *completeness* of the real numbers $\mathbb{R}$.
- The final definition is the general *topological* definition.
- The equivalence of the topological and ϵ-δ definitions follows by writing $U = \cup_{y\in U} B_{r(y)}(y)$ as a union of open balls and using properties of the pull-back f^{-1}.

Example: Cautionary Example

Let

$$f(x,y) = \begin{cases} \frac{x^2 y}{x^4+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$

Then f is **not** continuous at $(x,y) = (0,0)$.
Along every straight line through the origin $y = ax$, the limit is in fact 0:

$$\lim_{t\to 0} f(t,at) = \lim_{t\to 0} \frac{t^2 \cdot at}{t^4 + a^2 t^2} = \lim_{t\to 0} \frac{t^2}{t^2} \frac{at}{t^2 + a^2} = 0.$$

But along the curve $y = x^2$, we get something else:

$$\lim_{t\to 0} f(t,t^2) = \lim_{t\to 0} \frac{t^2 \cdot t^2}{t^4 + (t^2)^2} = \lim_{t\to 0} \frac{t^4}{t^4} \frac{1}{2} = \frac{1}{2}.$$

Calculus Review - Differentiability

Definition

The i'th *partial derivative* of a function $f : \mathbb{R}^n \to \mathbb{R}$ at $x = (x^1, \ldots, x^n)$ is

$$\begin{aligned} \partial_i f(x) &= \frac{\partial f}{\partial x^i}(x) \\ &= \lim_{h\to 0} \frac{f(x^1, \ldots, x^i + h, \ldots x^n) - f(x^1, \ldots, x^n)}{h} \end{aligned}$$

Partial derivatives are the usual derivatives in one variable holding all other variables fixed.

Definition

Let $X \in \mathbb{R}^n$. The *directional derivative* $\partial_X f(x)$ of f at x in the direction X is

$$\partial_X f(x) = \partial_t|_{t=0} f(x + tX) = \lim_{h\to 0} \frac{f(x + hX) - f(x)}{h}.$$

Note that the i'th partial derivative is the directional derivative in the e_i direction where $e_i = (0, \ldots, 0, 1, 0, \ldots 0)$ with the 1 in the i'th spot. That is,

$$\partial_i f = \partial_{e_i} f$$

Recall that Taylor's theorem with remainder states that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_{x_0}(x)$$

where

$$\lim_{x\to x_0} \frac{|R_{x_0}(x)|}{x - x_0} = 0.$$

We write $R_{x_0}(x) = o(|x - x_0|)$ as $x \to x_0$.

Definition

We say $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x_0 if there exists a linear map $L_{x_0} : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to x_0} \frac{|f(x) - f(x_0) - L_{x_0}(x - x_0)|}{|x - x_0|} = 0.$$

Thus if f is differentiable we have a higher dimensional version of Taylor's theorem. That is, there exists a linear map written $L_{x_0} = df_{x_0}$ such that

$$f(x) = f(x_0) + df_{x_0}(x - x_0) + \mathrm{o}(|x - x_0|)$$

as $x \to x_0$

Lemma

Let f be differentiable at x_0. Then for each $i = 1, \ldots, n$, $\partial_i f(x_0)$ exists and

$$\partial_i f(x_0) = df_{x_0}(e_i).$$

Proof

For $h \neq 0$, let $x = x_0 + he_i$.

$$0 = \lim_{h \to 0} \left| \frac{f(x_0 + he_i) - f(x_0)}{h} - \frac{df_{x_0}(he_i)}{h} \right|$$

hence

$$\partial_i f(x_0) = df_{x_0}(e_i)$$

Exercise

Show that if f is differentiable at x_0, then for any $X \in \mathbb{R}^n$, the directional derivative $df_{x_0}(X)$ is defined.

Example: Cautionary Example

Let
$$f(x,y) = \begin{cases} \frac{xy}{x^2+y^2}, & (x,y) \neq (0,0) \\ 0, & (x,y) = (0,0). \end{cases}$$
Then
$$\begin{aligned} \partial_x f(0,0) &= \lim_{h\to 0} \frac{f(h,0) - f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} \frac{h \cdot 0}{h^2 + 0^2} \\ &= 0. \end{aligned}$$
Likewise, $\partial_y f(0,0) = 0$ and hence both partial dertivatives exists at $(0,0)$. However, $\partial_{(1,1)} f(0,0)$ is not defined since
$$\begin{aligned} \partial_{(1,1)} f(0,0) &= \lim_{h\to 0} \frac{f(h,h) - f(0,0)}{h} \\ &= \lim_{h\to 0} \frac{f(h,h)}{h} \\ &= \lim_{h\to 0} \frac{1}{h} \frac{hh}{h^2 + h^2} \\ &= \lim_{h\to 0} \frac{1}{2h}. \end{aligned}$$
Thus f cannot be differentiable at $(0,0)$ since if it were differentiable, all directional derivatives would exist. Note that in this example, for $(x,y) \neq (0,0)$ we have
$$\partial_x f = \frac{y(y^2 - x^2)}{(x^2+y^2)^2}$$
and so
$$\partial_x f(0,y) = \frac{1}{y}$$
which is not continuous up to $y = 0$ even though $\partial_x f(0,0)$ is defined.

The issue here is that the partial derivatives, though defined everywhere, are not continuous. This kind of issue does not come up with for C^1 functions.

Definition

A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is C^1 (i.e. has continuous derivative) if f is differentiable at each x and moreover, the map
$$x \mapsto df_x$$
is continuous. This is equivalent to having *continuous* partial derivatives.

Note here that df_x is a linear map $\mathbb{R}^n \to \mathbb{R}^m$ and the set of all these is linearly isomorphic to the space $M_{n,m}$ of n by m matrices, which is itself linearly isomorphic to $\mathbb{R}^{nm}$ (index by i,j with $1 \leq i \leq n$

and $1 \leq j \leq m$).

Concretely we may realise df_x as the Jacobian matrix

$$(df_x)_{ij} = \partial_i f^j(x)$$

since $df_x(e_i) = \partial_i f(x) = (\partial_i f^1, \ldots, \partial_i f^n)$

Then $x \in \mathbb{R}^n \mapsto df_x \in \mathbb{R}^{nm}$ is a map between Euclidean spaces so we can ask if it's differentiable. We say f is C^2 if df is C^1 and more generally, f is C^k if $d^k f$ is continuous.

For our purposes, perhaps the most important of the basic results for differentiable functions is the chain rule.

Theorem: Chain Rule

The chain rule states that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at x_0 and $g : \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $f(x_0)$, then

$$d(f \circ h)_{x_0} = dh_{f(x_0)} \cdot df_{x_0}.$$

Exercise

Show that by the *chain rule*, given any curve γ such that $\gamma(0) = x$ and $\gamma'(0) = X$ we have

$$df_x \cdot X = \partial_t|_{t=0} f(\gamma(t)).$$

Inverse Function Theorem

From calculus we have the result:

Theorem: (1D) Inverse Function Theorem

Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth function with $f'(x_0) \neq 0$, there exists an interval I containing x_0 and an interval J containing $f(x_0)$ so that $f : I \to J$ is a diffeomorphism.

Generalising to arbitrary dimensions:

Theorem: Inverse Function Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ a smooth function such that df_{x_0} is invertible at x_0. Then there is an open set U containing x_0 and an open set V containing $f(x_0)$ such that $f|_U : U \to V$ is a diffeomorphism. Moreover

$$df^{-1}_{f(x_0)} = (df_{x_0})^{-1}$$

Note that if f is a diffeomorphism, then $f^{-1} \circ f(x) = x$. That is, $f^{-1} \circ f = \mathrm{Id}_x$. Since $d\,\mathrm{Id}_x = \mathrm{Id}_n$, by the chain rule we have

$$\mathrm{Id}_n = d\,\mathrm{Id}_x = d(f^{-1} \circ f)_{x_0} = df^{-1}_{f(x_0)} \cdot df_{x_0}.$$

That is df_{x_0} is invertible and

$$(df_{x_0})^{-1} = df^{-1}_{f(x_0)}.$$

Thus $d(f^{-1})$ at $y_0 = f(x_0)$ is necessarily equal to $(df)^{-1}$ at x_0. In one dimension $df = f'$ and $d(f^{-1}) = 1/f'$.

The basic idea is that if df is invertible, then f is invertible to first order. Writing

$$f(x) = f(x_0) + df_{x_0} \cdot (x - x_0) + o(|x - x_0|).$$

Let us ignore the $o(|x - x_0|)$ term (after all, it's insignificant compared with everything else for x near x_0!) and assume

$$f(x) = f(x_0) + df_{x_0} \cdot (x - x_0).$$

Then we can rearrange to solve for x to get

$$x = x_0 + df_{x_0}^{-1}(f(x) - f(x_0)).$$

Writing $y = f(x)$ and $y_0 = f(x_0)$ we obtain the inverse,

$$f^{-1}(y) = f^{-1}(y_0) + df_{x_0}^{-1} \cdot (y - y_0).$$

The task then is to work out how to deal with the presense of the $o(|x - x_0|)$ term. The approach is to construct a suitable *contraction* map (i.e. a map that strictly shrinks distances - see more below). We will prove the inverse function theorem below after considering some consequences. As a prview, to prove the Inverse Function Theorem, given y we need a uniquely solution of $f(x) = y$. Define

$$T_y(x) = x - df_{x_0}^{-1}(f(x) - y).$$

Then we show that for suitable $r > 0$, T_y is a contraction map $\bar{B}_r(x_0) \to \bar{B}_r(x_0)$ and a cornerstone result in analysis (namely the Banach Fixed Point Theorem) implies that T_y posses a unique fixed point x_y^*. That is, there is a unique point $x_y^* \in \bar{B}_r(x_0)$ such that $T_y(x_y^*) = x_y^*$. Observe then that

$$\begin{aligned} T_y(x_y^*) = x_y^* &\Leftrightarrow df_{x_0}^{-1}(f(x_y^*) - y) \\ &\Leftrightarrow f(x_y^*) = y. \end{aligned}$$

The last equivalence follows from the assumption that df_{x_0} is invertible. Thus $f(x_y^*) = y)$ if and only if T_y has a fixed point x_y^*. By showing this fixed point is unique we then may unambiguously define

$$f^{-1}(y) = x_y^*.$$

Here is an example application of the Inverse Function Theorem.

Example

Consider

$$\begin{cases} x - y^2 & = a \\ x^2 + y + y^3 & = b \end{cases}$$

For $(a, b) = (0, 0)$: $(x, y) = (0, 0)$ is a solution.
Question: For what (a, b) is the system solvable?
To answer the question, let $F(x, y) = (x - y^2, x^2 + y - y^3)$. Then

$$dF = \begin{pmatrix} 1 & -2y \\ 2x & 1 - 3y^2 \end{pmatrix}$$

We have $dF_{(0,0)} = \mathrm{Id}$ hence by the IFT there is a neighbourhood of $(x, y) = (0, 0)$ for which F maps diffeomorphically onto a neighbourhood of $(a, b) = (0, 0)$. Therefore, for (a, b) in a neighbourhood of $(0, 0)$, there is a neighbourhood of $(0, 0)$ containing a unique solution of $F(x, y) = (a, b)$.
Note that given (a, b)), there is not generally a unique solution. In fact, even for $(a, b) = (0, 0)$ there is not a unique solution since if y is a real root of $y^3 + y^2 + 1$, then $F(y^2, y) = (0, 0)$. Such a root always exists since $y^3 + y^2 + 1$ is an odd-degree polynomial.

Implicit Function Theorem

Using the isomorphism, $\mathbb{R}^n \oplus \mathbb{R}^k \simeq \mathbb{R}^{n+k}$ we may write a point in $\mathbb{R}^{n+k}$ as (x, y) with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^k$. Then for a function $F = F(x, y) : \mathbb{R}^{n+k} \to \mathbb{R}^k$ we also split the differential into x, y parts:

$$dF = \begin{pmatrix} d_x F & d_y F \end{pmatrix}.$$

Theorem: Implicit Function Theorem

Let $F : \mathbb{R}^{n+k} \to \mathbb{R}^k$ be smooth with (x_0, y_0) such that $d_y F|_{(x_0,y_0)}$ is invertible. Then there is an open neighbourhood U of x_0 and a unique smooth function $g : U \to \mathbb{R}^n$ such that

$$F(x, g(x)) = F(x_0, y_0).$$

The Implicit Function Theorem is equivalent to the Inverse Function Theorem. Here we show how to derive the Implicit Function Theorem from the Inverse Function Theorem.

Proof

Define

$$\bar{F}(x,y) = (x, F(x,y)) \in \mathbb{R}^{n+k}$$

Then

$$d\bar{F} = \begin{pmatrix} \mathrm{Id}_n & 0 \\ * & d_y F \end{pmatrix}$$

is invertible at (x_0, y_0) since the assumption is that $d_y F$ is invertible at (x_0, y_0). Hence by the inverse function theorem, $\bar{F}$ is locally invertible.
Since $\bar{F}(x,y) = (x, F(x,y))$,

$$\bar{F}^{-1}(x,y) = (x, G(x,y))$$

for a smooth function $G : \mathbb{R}^{n+k} \to \mathbb{R}^k$. This follows by writing $\bar{F}^{-1} = (H, G)$, from which we claim that necessarily $H(x,y) = x$. By the definition of inverse functions,

$$\begin{aligned} (x,y) &= \bar{F} \circ \bar{F}^{-1}(x,y) \\ &= \bar{F}(H(x,y), G(x,y)) \\ &= (H(x,y), F(G(x,y))). \end{aligned}$$

Comparing the first component of the left and right hand sides we see that $x = H(x,y)$ as claimed.
Now let $c = F(x_0, y_0)$ and

$$g(x) = G(x,c)$$

from which it follows that

$$\begin{aligned} (x, F(x, g(x))) &= \bar{F}(x, g(x)) \\ &= \bar{F}(x, G(x,c)) \\ &= \bar{F} \circ \bar{F}^{-1}(x,c) \\ &= (x,c) \end{aligned}$$

and $F(x, g(x)) = c = F(x_0, y_0)$ as required.

Exercise

Assuming the Implicit Function Theorem is true, prove the Inverse Function Theorem.

We may interpret the Implicit Function Theorem as follows: consider the level set

$$F^{-1}(c) = \{(x,y) : F(x,y) = c\}.$$

If $d_y F$ is invertible for each $(x,y) \in F^{-1}(c)$, then the level set is locally the graph of a smooth function.

Exercise

Let $F : \mathbb{R}^{n+k} \to \mathbb{R}^k$ be a smooth function such that dF has rank k at $z_0 \in \mathbb{R}^{n+k}$. By permuting the indices, use the Implicit Function Theorem to show that for z in a neighbourhood of z_0, we may parametrise the level set $F(z) = F(z_0)$ as the graph of a smooth function $g : \mathbb{R}^n \to \mathbb{R}^k$.

Example

Let $F(x, y) = x^2 + y^2$
Here $n = k = 1$

$$dF = \begin{pmatrix} 2x & 2y \end{pmatrix}$$

For $x \neq \pm 1$

$$F(x, \sqrt{1 - x^2}) = 1$$

Immersions and Submersions

Here are some further statements equivalent to the Inverse Function Theorem, and hence also equivalent to the Implicit Function Theorem.

Exercise

Prove that the theorems below are equivalent to the Inverse Function Theorem. You may find it easier to prove equivalence with the Implicit Function Theorem which is equivalent anyway.

Definition

Let $F : \mathbb{R}^{n+k} \to \mathbb{R}^k$ be a smooth map. Then F is an *submersion* if dF is surjective.

Note that

$$\begin{aligned} dF \text{ surjective} &\Leftrightarrow dF \text{ has maximal rank} \\ &\Leftrightarrow \operatorname{rnk} dF = k = \dim \operatorname{coDom}(dF) \\ &\Leftrightarrow \dim \ker dF = n \end{aligned}$$

Definition

An projection of $\mathbb{R}^{n+k}$ onto $\mathbb{R}^k$ is a map of the form

$$\pi : x \in \mathbb{R}^{n+k} \mapsto (x^{n+1}, \ldots, x^{n+k}) \in \mathbb{R}^k$$

Note that $d\pi = \begin{pmatrix} \mathrm{Id}_n & 0_k \end{pmatrix}$ is surjective.

We may also change the order: eg. $\pi(x_1, x_2, x_3) = (x_2, x_3)$

Theorem

Let F be a submersion. Then F is locally a projection up to diffeomorphism.

There are diffeomorphisms

$$\begin{aligned} \varphi &: U \subseteq \mathbb{R}^{n+k} \to V \subseteq \mathbb{R}^{n+k} \\ \psi &: W \subseteq \mathbb{R}^{k} \to Z \subseteq \mathbb{R}^{k} \end{aligned}$$

such that $F|_U = \psi^{-1} \circ \pi \circ \varphi$

Dual to the notion of submersion is the notion of immersion.

Definition

Let $F : \mathbb{R}^n \to \mathbb{R}^{n+k}$ be a smooth map. Then F is an *immersion* if dF is injective.

$$\begin{aligned} dF \text{ injective} &\Leftrightarrow dF \text{ has maximal rank} \\ &\Leftrightarrow \operatorname{rnk} dF = n = \dim \operatorname{Dom}(dF) \\ &\Leftrightarrow \dim \ker dF = 0 \end{aligned}$$

Definition

An inclusion of $\mathbb{R}^n$ into $\mathbb{R}^{n+k}$ is a map of the form

$$\iota : x \in \mathbb{R}^n \mapsto (x, 0_k)$$

where $0_k = (0, \dots, 0) \in \mathbb{R}^k$.

Note that $d\iota = \begin{pmatrix} \mathrm{Id}_n \\ 0_k \end{pmatrix}$ is injective.

We may also change the order: eg. $\iota(x_1, x_2) = (0, x_1, x_2, 0)$

Theorem

Let F be an immersion. Then F is locally an inclusion up to diffeomorphism.

There are diffeomorphisms

$$\begin{aligned} \varphi &: U \subseteq \mathbb{R}^{n} \to V \subseteq \mathbb{R}^{n} \\ \psi &: W \subseteq \mathbb{R}^{n+k} \to Z \subseteq \mathbb{R}^{n+k} \end{aligned}$$

such that $F|_U = \psi^{-1} \circ \iota \circ \varphi$

Contractions

Definition

A map $T : \bar{B}_r(p) \to \bar{B}_r(p)$ is a *contraction map* if there exists a constant $0 \leq L < 1$ such that

$$|T(x) - T(y)| \leq L\,|x - y|\,.$$

A contraction map strictly decreases the distance between two points. The primary significane of the definition is the following:

Theorem: Banach fixed point theorem

Let T be a contraction map. Then there exists a unique *fixed point* $x^* \in B_r(p)$ of T. That is, there exists a unique point x^* such that $T(x^*) = x^*$.

Proof

We have

$$\begin{aligned}|x - y| &\leq |x - T(x)| + |T(x) - y| \\ &\leq |x - T(x)| + |T(x) - T(y)| + |T(y) - y| \\ &\leq |x - T(x)| + L\,|x - y| + |T(y) - y|\,.\end{aligned}$$

and hence

$$|x - y| \leq \frac{|x - T(x)| + |T(y) - y|}{1 - L}$$

Thus if $T(x) = x$ and $T(y) = y)$ then $x = y$ and hence fixed points are unique.
To prove existence, pick any x_0 and define $x_n = T^n(x_0) = \underbrace{T \circ \cdots \circ T}_{n \text{ times}}(x_0)$. Supposing first that the limit exists, then using $x_n = T(x_{n-1})$ we have

$$x_* = \lim_{n\to\infty} x_n = \lim_{n\to\infty} T(x_{n-1}) = T(\lim_{n\to\infty} x_{n-1}) = T(x^*)$$

Thus x_* is a fixed point and we just need to prove the limit exists. We do this by showing that $x_n = T^n(x_0)$ is a Cauchy sequence:

$$\begin{aligned}&|T^n(x_0) - T^m(x_0)| \\ &\quad\leq \frac{|T(T^n(x_0)) - T^n(x_0)| + |T(T^m(x_0)) - T^m(x_0)|}{1 - L} \\ &\quad= \frac{|T^n(T(x_0)) - T^n(x_0)| + |T^m(T(x_0) - T^m(x_0)|}{1 - L} \\ &\quad\leq \frac{L^n\,|T(x_0) - x_0| + L^m\,|T(x_0) - x_0|}{1 - L} \to 0\end{aligned}$$

as $m, n \to \infty$. Note here that we used $|T^n(x) - T^n(y)| \leq L^n\,|x - y|$ which follows by induction. Completness now ensures the limit exists and the proof is complete.

Proof of Inverse Function Theorem

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ a smooth function such that df_{x_0} is invertible at x_0. Then there is an open set U containing x_0 and an open set V containing $f(x_0)$ such that $f|_U : U \to V$ is a diffeomorphism. Moreover

$$df^{-1}_{f(x_0)} = (df_{x_0})^{-1}$$

The strategy of proof is as follows: Define $T_y(x) = x - df^{-1}_{x_0}(f(x) - y)$. Then we have two steps.

Step 1 T_y is a contraction $T_y : \bar{B}_r(x_0) \to \bar{B}_r(x_0)$

Step 2 Prove f^{-1} is smooth.

We will actually break up each step into smaller lemmas.

Lemma

For each fixed y sufficiently close to $y_0 = F(x_0)$,

$$T_y(x) = x - df^{-1}_{x_0}(f(x) - y)$$

is a contraction map.

Proof

$$dT_{x_0} = d\,\mathrm{Id}_{x_0} - df^{-1}_{x_0} df_{x_0} = 0.$$

Continuity of dT gives an open neighbourhood U of x_0 such that $\|dT_{x_0}\| \leq 1/2$:

$$|dT_x \cdot X| \leq \frac{1}{2} |X| .$$

From $|dT_x \cdot X| \leq \frac{1}{2} |X|$ and the mean value inequality,

$$|T(x_1) - T(x_2)| \leq \frac{1}{2} |x_1 - x_2|$$

so that T is *contractive* for $x_1, x_2 \in U$.

In order to conclude that T has a unique fixed point, we need to verify that there is an $r > 0$ such that $T : \bar{B}_r(x_0) \to \bar{B}_r(x_0)$. Since $x_0 \in U$ and U is open, there exists an $r > 0$ such that $B_r(x_0) \subseteq U$.

Lemma

Let $y_0 = f(x_0)$ and $y \in B_s(y_0)$ with s any number satisfying

$$0 < s < \frac{1-L}{\|df_{x_0}^{-1}\|} r.$$

Then for any $y \in B_s(y_0)$, T_y maps $\bar{B}_r(x_0)$ to itself.

Proof

For $x \in B_r(x_0)$, recalling $T(x) = x - df_{x_0}^{-1}(f(x) - y)$ we have

$$\begin{aligned} |T(x) - x_0| &\leq |T(x) - T(x_0)| + |T(x_0) - x_0| \\ &\leq L\,|x - x_0| + \left|-df_{x_0}^{-1}(f(x_0) - y)\right| \\ &\leq L\,|x - x_0| + \|df_{x_0}^{-1}\|\,|y_0 - y| \\ &\leq rL + \|df_{x_0}^{-1}\| s \\ &\leq rL + (1-L)r = r. \end{aligned}$$

Let us state explicitly what the previous lemmas give us.

Lemma

For each $y \in \bar{B}_s(y_0)$, there exists a unique fixed point $x_y^* \in \bar{B}_r(x_0)$ for T_y and hence f restricted to $f^{-1}(B_s(y_0) \cap B_r(x_0)$ is invertible.

Proof

For each $y \in B_s(y_0)$, $T_y : \bar{B}_r(x_0) \to \bar{B}_r(x_0)$ is a contraction mapping hence has a unique fixed point x_y^*:

$$x_y^* = T_y(x_y^*) = x_y^* - df_{x_0}^{-1}(f(x_y^*) - y).$$

Cancelling x_y^* from both sides and using the fact that $df_{x_0}^{-1}$ is non-singular:

$$\begin{aligned} x_y^* = T_y(x_y^*) &\Leftrightarrow df_{x_0}^{-1}(f(x_y^*) - y) = 0 \\ &\Leftrightarrow f(x_y^*) = y \end{aligned}$$

Thus the unique fixed point x_y^* is also the unique solution of $f(x) = y)$ for $x \in B_r(x_0)$ provided $f(x_y^* \in B_s(y_0)$ and we may define

$$f^{-1}(y) = x_y^*.$$

Note we need to restrict the range of x to the open set $f^{-1}(B_s(y_0)) \cap B_r(x_0)$ so that f maps this set into $B_s(y_0)$.

Now we move onto step 2, which we also prove by a series of lemmas.

Lemma

The (local) inverse f^{-1} is continuous.

Proof

$$\begin{aligned} \left|x_1 - x_2 - df_{x_0}^{-1}(f(x_1) - f(x_2))\right| &= |T(x_1) - T(x_2)| \\ &\leq L\,|x_1 - x_2|\,. \end{aligned}$$

By the *reverse triangle inequality*

$$|x_1 - x_2| - \left|df_{x_0}^{-1}(f(x_1) - f(x_2))\right| \leq L\,|x_1 - x_2|\,.$$

Thus

$$\begin{aligned} |x_1 - x_2| &\leq \frac{\left|df_{x_0}^{-1}(f(x_1) - f(x_2))\right|}{1 - L} \\ &\leq \frac{\|df_{x_0}^{-1}\|}{1 - L}\,|f(x_1) - f(x_2)|\,. \end{aligned}$$

Letting $y_i = f(x_i)$ so that $x_i = f^{-1}(y_i)$ gives continuity (even Lipschitz):

$$\left|f^{-1}(y_1) - f^{-1}(y_2)\right| \leq \frac{\|df_{x_0}^{-1}\|}{1 - L}\,|y_1 - y_2|\,.$$

Lemma

The (local) inverse f^{-1} is differentiable.

Proof

Pick any arbitrary $y \in B_s(y_0)$ and any h such that $y + h \in B_s(y_0)$, say $h \in B_\epsilon(0)$ so that $y + h \in B_\epsilon(y) \subseteq B_s(y_0)$.
Let $x = f^{-1}(y)$ and define

$$R = f^{-1}(y+h) - f^{-1}(y) - df_x^{-1} \cdot h.$$

We need to show that

$$\lim_{h \to 0} \frac{|R|}{|h|} = 0.$$

Let $k = f^{-1}(y+h) - f^{-1}(y)$ so that $h = f(x+k) - f(x)$. Then

$$\begin{aligned} R &= f^{-1}(y+h) - f^{-1}(y) - df_x^{-1} \cdot h \\ &= k - df_x^{-1}(f(x+k) - f(x)) \\ &= k - df_x^{-1}(df_x k + \mathrm{o}(k)) \\ &= -df_x^{-1}(\mathrm{o}(k)). \end{aligned}$$

Since f^{-1} is Lipschitz, with constant M say, we have

$$|k| = \left| f^{-1}(y+h) - f^{-1}(y) \right| \leq M \left| y + h - y \right| = M \left| h \right|.$$

$$\frac{|R|}{|h|} \leq \|df_x^{-1}\| \frac{\mathrm{o}(k)}{|h|} \leq M \|df_x^{-1}\| \frac{\mathrm{o}(k)}{|k|}.$$

The right hand side goes to zero as $h \to 0$ since $|k| \leq M\,|h|$ implies $k \to 0$ and then by definition of $\mathrm{o}(k)$.

Lemma

The (local) inverse, f^{-1} is smooth.

Proof

We have shown the existence of a differentiable local inverse f^{-1} to f with differential

$$d(f^{-1})_y = (df_x)^{-1}$$

where $x = f^{-1}(y)$.
Now, by Cramers's rule, given an invertible matrix A, the inverse is

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A$$

where the $\operatorname{adj} A$ is the *adjugate matrix* formed from cofactors of A - that is the determinants of the minors of A. As a function then, $A \mapsto A^{-1}$ we see that the components are rational functions of the entries of A (since determinants are polynomials in the entries of A). The inverse function Inv is smooth function from the open set of non-singular matrices ($\det A \neq 0$) to itself.
Then since $x \mapsto df_x$ is smooth,

$$y \mapsto df_y^{-1} = (df_{f^{-1}(y)})^{-1} = \mathrm{Inv} \circ df \circ f^{-1}(y)$$

is a composition of C^0 functions hence C^0. Then f^{-1} is C^1.
Now

$$df_y^{-1} = \mathrm{Inv} \circ df \circ f^{-1}(y)$$

and df^{-1} is the composition of C^1 functions hence is also C^1.
That is f^{-1} is C^2. By induction, f^{-1} is C^k for any k and hence smooth.