Notes on Introduction to Real Analysis

Contents

1 Multivariable Differential Calculus

1.1 Linear Approximation

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$, where each component $f_i : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function.

Regardless of whether f is linear or non-linear, a natural question arises: can it be locally approximated by a linear function? And if so, how accurate is this approximation?

Ultimately, given a point $\mathbf{x}_0 \in \mathbb{R}^n$, our goal is to approximate f near $\mathbf{x}_0$ using its linearization:

$$f(\mathbf{x}) \approx f(\mathbf{x}_0) + Df(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$$

Definition 1.1 (Differentiability)
Let $E \subseteq \mathbb{R}^n$, $f : E \to \mathbb{R}^m$ be a function, and $\mathbf{x}_0$ be an interior point of E. Let $L : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. We say f is differentiable at $\mathbf{x}_0$ with derivative L if

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

Equivalently,

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{0}) - L(\mathbf{h})\|}{\|\mathbf{h}\|} = 0.$$

Therefore, multivariable differentiability is defined such that the relative error between the function and its linear approximation vanishes as the distance approaches zero.

Theorem 1.1

Let $E \subseteq \mathbb{R}$, $f : E \to \mathbb{R}$ be a function, $L \in \mathbb{R}$, and x_0 be an interior point of E. TFAE:

1. *f is differentiable at x_0 and $f'(x_0) = L$.*
2. $\lim_{x \to x_0} \frac{|f(x) - f(x_0) - L(x - x_0)|}{|x - x_0|} = 0$.

> *Proof.*
> For $x \neq x_0$, write
> $$\frac{f(x) - f(x_0)}{x - x_0} = L + \frac{f(x) - f(x_0) - L(x - x_0)}{x - x_0}.$$
> Rearrange L to the left, take the absolute value and limit $x \to x_0$, and we can show the limit on the both sides imply each other. The second expression is exactly the special case of definition (1.1).
>
> ■

Theorem 1.2 (Uniqueness of derivatives)

Let $\mathbf{x}_0 \in E$ be an interior point of E. Suppose $L_1, L_2 : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations such that f is differentiable at $\mathbf{x}_0$ with derivative L_1 and also differentiable at $\mathbf{x}_0$ with derivative L_2. Then $L_1 = L_2$.

> *Proof.*
> For $\mathbf{h} \neq \mathbf{0}$,
> $$\frac{\|L_1(\mathbf{h}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} \leq \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{0}) - L_1(\mathbf{h})\|}{\|\mathbf{h}\|} + \frac{\|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{0}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|}.$$
> By sandwich, $\lim_{\mathbf{h} \to \mathbf{0}} \frac{\|L_1(\mathbf{h}) - L_2(\mathbf{h})\|}{\|\mathbf{h}\|} = 0$.
> Let $\mathbf{h} = t\mathbf{v}, t \neq 0$, we have
> $$\lim_{\mathbf{v} \to \mathbf{0}} \frac{\|L_1(t\mathbf{v}) - L_2(t\mathbf{v})\|}{\|t\mathbf{v}\|} = \lim_{\mathbf{v} \to \mathbf{0}} \frac{|t|\|L_1(\mathbf{v}) - L_2(t\mathbf{v})\|}{|t|\|\mathbf{v}\|} \quad (L_1 \text{ and } L_2 \text{ are linear.})$$
> Hence, for $\mathbf{v} \neq \mathbf{0}$, $L_1(\mathbf{v}) = L_2(\mathbf{v})$, which is also true when $\mathbf{v} = \mathbf{0}$.
>
> ■

We may prove by contradiction that f is not differnetiable by the above theorem.

Theorem 1.3

If f is differentiable at $\mathbf{x}_0$, then f is continuous at $\mathbf{x}_0$.

Proof.
For $\mathbf{x} \neq \mathbf{x}_0$, write

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x}-\mathbf{x}_0)\| = \underbrace{\frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x}-\mathbf{x}_0)\|}{\|\mathbf{x}-\mathbf{x}_0\|}}_{\to 0 \text{ by assumption}} \underbrace{\|\mathbf{x}-\mathbf{x}_0\|}_{\to 0 \text{ as } \mathbf{x}\to\mathbf{x}_0}.$$

Next, use

$$\|f(\mathbf{x}) - f(\mathbf{x}_0)\| \leq \underbrace{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x}-\mathbf{x}_0)\|}_{\to 0 \text{ from the previous equality}} + \underbrace{\|L(\mathbf{x}-\mathbf{x}_0)\|}_{\to 0 \text{ as } \mathbf{x}\to\mathbf{x}_0}.$$

By Sandwich, $\lim_{\mathbf{x}\to\mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$.

■

Definition 1.2 (Directional Derivative)
Let $E \subseteq \mathbb{R}^n$, $f : E \to \mathbb{R}^m$ be a function, and $\mathbf{x}_0$ be an interior point of E. Fix $\mathbf{v} \in \mathbb{R}^n$, if the limit

$$D_{\mathbf{v}} f(\mathbf{x}_0) = \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t}$$

exists, we call this limit the directional derivative of f at $\mathbf{x}_0$ in the direction $\mathbf{v}$.

Definition 1.3 (Partial Derivative)
Let $E \subseteq \mathbb{R}^n$, $f : E \to \mathbb{R}^m$ be a function, and $\mathbf{x}_0$ be an interior point of E. Let $1 \leq j \leq n$, if the limit

$$\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = \lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{e}_j) - f(\mathbf{x}_0)}{t}$$

exists, we call this limit the partial derivative of f with respect to the variable x_j.

So partial derivative is just a special case of directional derivative.

Theorem 1.4
If f is differentiable at $\mathbf{x}_0$, then f is differentiable in the direction $\mathbf{v}$ at $\mathbf{x}_0$, and:

$$D_{\mathbf{v}} f(\mathbf{x}_0) = f'(\mathbf{x}_0) \cdot \mathbf{v}.$$

Proof.
For $\mathbf{h} \neq 0$, let $\mathbf{h} = t\mathbf{v}$, $t \neq 0$, write

$$\frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0) - L(t\mathbf{v})}{t} + \frac{L(t\mathbf{v})}{t}.$$

Rearrange $L(\mathbf{v})$ to the left, and take the Euclidean norm and $t \to 0$, the RHS converge to 0 by differentiablity ($\|\mathbf{v}\| \neq 0$). So we have

$$\lim_{t\to 0} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} = L(\mathbf{v}) = f'(\mathbf{x}_0) \cdot \mathbf{v}.$$

$\blacksquare$

We may also disprove the differentiability of f provided its directional derivative exists. The next question is: how to write the linear map $f'(\mathbf{x}_0)$ explicitly?

Theorem 1.5

Let $E \subseteq \mathbb{R}^n$, $f : E \to \mathbb{R}^m$ be a function, and $\mathbf{x}_0$ be an interior point of E. If all the partial derivatives of f exist on E and are continuous at $\mathbf{x}_0$, then f is differentiable at $\mathbf{x}_0$, and

$$f'(\mathbf{x}_0) \cdot \mathbf{v} = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(\mathbf{x}_0).$$

Proof.

Step 1: Setup

Let $L(\mathbf{v}) = \sum_{j=1}^{n} v_j \frac{\partial f}{\partial x_j}(\mathbf{x}_0)$ be our candidate for the derivative. Since $\mathbf{x}_0$ is an interior point of E, there exists $r > 0$ such that the ball $B(\mathbf{x}_0, r) \subseteq E$.

For $\epsilon > 0$, since each partial derivative $\frac{\partial f_i}{\partial x_j}$ is continuous at $\mathbf{x}_0$, there exists $\delta_{i,j} > 0$ such that $\|\mathbf{x} - \mathbf{x}_0\| < \delta_{i,j}$ implies

$$\left| \frac{\partial f_i}{\partial x_j}(\mathbf{x}) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0) \right| < \frac{\epsilon}{mn}$$

for all $1 \le i \le m$ and $1 \le j \le n$.

Let $\delta = \min\{r, \delta_{i,j} \mid 1 \le i \le m, 1 \le j \le n\}$. Now, fix $\mathbf{x} \in E$ such that $\|\mathbf{x} - \mathbf{x}_0\| < \delta$. Let $\mathbf{v} = \mathbf{x} - \mathbf{x}_0 = \sum_{j=1}^{n} v_j \mathbf{e}_j$. Note that $|v_j| \le \|\mathbf{x} - \mathbf{x}_0\| < \delta$. We construct a path from $\mathbf{x}_0$ to $\mathbf{x}$ along the coordinate axes. Define the sequence of points:

$$\begin{aligned}
\mathbf{x}^{(0)} &= \mathbf{x}_0 \\
\mathbf{x}^{(1)} &= \mathbf{x}_0 + v_1 \mathbf{e}_1 \\
\mathbf{x}^{(2)} &= \mathbf{x}_0 + v_1 \mathbf{e}_1 + v_2 \mathbf{e}_2 \\
&\vdots \\
\mathbf{x}^{(j)} &= \mathbf{x}^{(j-1)} + v_j \mathbf{e}_j \\
&\vdots \\
\mathbf{x}^{(n)} &= \mathbf{x}_0 + \sum_{j=1}^{n} v_j \mathbf{e}_j = \mathbf{x}
\end{aligned}$$

We can express the difference $f(\mathbf{x}) - f(\mathbf{x}_0)$ as a telescoping sum:

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \sum_{j=1}^{n} \left(f(\mathbf{x}^{(j)}) - f(\mathbf{x}^{(j-1)}) \right)$$

For each component i, consider $f_i(\mathbf{x}^{(j)}) - f_i(\mathbf{x}^{(j-1)})$. Let $\varphi_i^j(t) = f_i(\mathbf{x}^{(j-1)} + tv_j\mathbf{e}_j)$, then $f_i(\mathbf{x}^{(j)}) - f_i(\mathbf{x}^{(j-1)}) = \varphi_i^j(1) - \varphi_i^j(0)$. By the Mean Value Theorem, $\exists t_j \in (0,1)$ such that $\varphi_i^j(1) - \varphi_i^j(0) = (\varphi_i^j)'(t_j)$.

Step 2: Continuity of f_i along the coordinate $\mathbf{x}_j$

Applying Lemma (1.1):

$$f_i(\mathbf{x}^{(j)}) - f_i(\mathbf{x}^{(j-1)}) = (\varphi_i^j)'(t_j) = \frac{\partial f_i}{\partial x_j}(\boldsymbol{\xi}_j)v_j,$$

where $\boldsymbol{\xi}_j = \mathbf{x}^{(j-1)} + t_j v_j \mathbf{e}_j$.

Note that

$$\begin{aligned}\|\mathbf{x}^{(j-1)} + t_j v_j \mathbf{e}_j - \mathbf{x}_0\| &= \|v_1\mathbf{e}_1 + v_2\mathbf{e}_2 + \cdots + v_{j-1}\mathbf{e}_{j-1} + t_j v_j \mathbf{e}_j\| \\ &\leq \left\|\sum_{i=1}^{n} v_i \mathbf{e}_i\right\| \quad (\because |t_j| < 1) \\ &= \|\mathbf{x} - \mathbf{x}_0\| \leq \delta,\end{aligned}$$

which implies

$$\begin{aligned}\left|f_i(\mathbf{x}^{(j)}) - f_i(\mathbf{x}^{(j-1)}) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)v_j\right| &= \left|\frac{\partial f_i}{\partial x_j}(\mathbf{x}^{(j-1)} + t_j v_j \mathbf{e}_j) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)\right| |v_j| \\ &< \frac{\epsilon}{mn}|v_j|.\end{aligned}$$

Step 3: Bound the error for each component f_i

$$\begin{aligned}\left|f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - \sum_{j=1}^{n} \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)v_j\right| &= \left|\sum_{j=1}^{n}\left(f_i(\mathbf{x}^{(j)}) - f_i(\mathbf{x}^{(j-1)}) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)v_j\right)\right| \\ &\leq \sum_{j=1}^{n}\left|\frac{\partial f_i}{\partial x_j}(\boldsymbol{\xi}_j) - \frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)\right| |v_j| \\ &< \sum_{j=1}^{n}\frac{\epsilon}{mn}|v_j| \\ &\leq \sum_{j=1}^{n}\frac{\epsilon}{mn}\|\mathbf{x} - \mathbf{x}_0\| = \frac{\epsilon}{m}\|\mathbf{x} - \mathbf{x}_0\|\end{aligned}$$

Step 4: Extend this to the vector-valued function f

$$\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)\| \leq \sum_{i=1}^{m}\left|f_i(\mathbf{x}) - f_i(\mathbf{x}_0) - \sum_{j=1}^{n}\frac{\partial f_i}{\partial x_j}(\mathbf{x}_0)v_j\right|$$

$$< \sum_{i=1}^{m} \frac{\epsilon}{m} \|\mathbf{x} - \mathbf{x}_0\| = \epsilon \|\mathbf{x} - \mathbf{x}_0\|$$

Dividing both sides by $\|\mathbf{x} - \mathbf{x}_0\|$, we get:

$$\frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} < \epsilon$$

Since this holds for any $\epsilon > 0$ when $\|\mathbf{x} - \mathbf{x}_0\|$ is sufficiently small, we conclude:

$$\lim_{\mathbf{x} \to \mathbf{x}_0} \frac{\|f(\mathbf{x}) - f(\mathbf{x}_0) - L(\mathbf{x} - \mathbf{x}_0)\|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0$$

This proves that f is differentiable at $\mathbf{x}_0$, and its derivative is given by L.

■

Lemma (Chain rule along a coordinate segment)
Let $E \subset \mathbb{R}^n$ and $f : E \to \mathbb{R}^m$. Fix $1 \le i \le m$ and $1 \le j \le n$. Define

$$\varphi_i^j(t) := f_i(\mathbf{z} + tv\mathbf{e}_j), \quad 0 \le t \le 1,$$

where $\mathbf{z} \in E$ and $v \in \mathbb{R}$. Assume that for some $t_0 \in (0, 1)$ there exists $\rho > 0$ such that

$$\mathbf{z} + (t_0 + h)v\mathbf{e}_j \in E \quad \textit{for all } |h| < \rho,$$

and that the partial derivative $\frac{\partial f_i}{\partial x_j}$ exists at the point $\mathbf{p} := \mathbf{z} + t_0 v\mathbf{e}_j$. Then φ_i^j is differentiable at t_0 and

$$(\varphi_i^j)'(t_0) = \frac{\partial f_i}{\partial x_j}(\mathbf{p})v.$$

Proof.
Let

$$\begin{aligned}(\varphi_i^j)'(t_0) = \lim_{h \to 0} \frac{\varphi_i^j(t_0 + h) - \varphi_i^j(t_0)}{h} &= \lim_{h \to 0} \frac{f_i(\mathbf{z} + (t_0 + h)v\mathbf{e}_j) - f_i(\mathbf{z} + t_0 v\mathbf{e}_j)}{t} \\ &= \lim_{h \to 0} \frac{f_i(\mathbf{p} + hv\mathbf{e}_j) - f_i(\mathbf{p})}{h}.\end{aligned}$$

If $v = 0$, $f_i(\mathbf{p} + hv\mathbf{e}_j) - \mathbf{p} = 0$. Now consider $v \neq 0$.
For arbitrarily small h, $\mathbf{z} + (t_0 + h)v\mathbf{e}_j = \mathbf{p} + hv\mathbf{e}_j \in B(\mathbf{p}, \rho) \subseteq E$. Set $hv = s$, the above quotient becomes:

$$\lim_{s \to 0} \frac{f_i(\mathbf{p} + s\mathbf{e}_j) - f_i(\mathbf{p})}{s} \cdot v = \frac{\partial f_i}{\partial x_j}(\mathbf{p})v,$$

which shows that φ_i^j is differentiable at t_0 and $(\varphi_i^j)'(t_0) = \frac{\partial f_i}{\partial x_j}(\mathbf{p})v$.

■

Remark. Let $E \subseteq \mathbb{R}^n$, $f : E \to \mathbb{R}^m$ be a function, and $\mathbf{x}_0$ be an interior point of E. If f is differentiable at $\mathbf{x}_0$, by defintion, there exists a linear map such that the relative error converge to zero near $\mathbf{x}_0$. Then we call this linear map (total) derivative of f at $\mathbf{x}_0$, denoted by

$$Df(\mathbf{x}_0) = f'(\mathbf{x}_0) = \begin{pmatrix} - & \nabla f_1(\mathbf{x}_0) & - \\ - & \nabla f_2(\mathbf{x}_0) & - \\ & \vdots & \\ - & \nabla f_m(\mathbf{x}_0) & - \end{pmatrix}_{m\times n}$$

$$= \begin{pmatrix} | & | & & | \\ \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \\ | & | & & | \end{pmatrix}_{m\times n}.$$

Hence, we have:

1. The partial derivative is simply the directional derivative along the standard basis vector $\mathbf{e}_j$:
$$\frac{\partial f}{\partial x_j}(\mathbf{x}_0) = D_{\mathbf{e}_j} f(\mathbf{x}_0) = f'(\mathbf{x}_0)\mathbf{e}_j$$
2. If f is differentiable, then for any direction vector $\mathbf{v} = \sum_{j=1}^n v_j \mathbf{e}_j$, the directional derivative can be expanded as:
$$D_{\mathbf{v}} f(\mathbf{x}_0) = f'(\mathbf{x}_0)\mathbf{v} = f'(\mathbf{x}_0)\left(\sum_{j=1}^n v_j \mathbf{e}_j\right) = \sum_{j=1}^n v_j f'(\mathbf{x}_0)\mathbf{e}_j = \sum_{j=1}^n v_j \frac{\partial f}{\partial x_j}(\mathbf{x}_0)$$
This can be naturally represented in matrix multiplication form (where the derivative is a $m \times n$ matrix):
$$\begin{pmatrix} | & | & & | \\ \frac{\partial f}{\partial x_1}(\mathbf{x}_0) & \frac{\partial f}{\partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial f}{\partial x_n}(\mathbf{x}_0) \\ | & | & & | \end{pmatrix}_{m\times n} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}_{n\times 1}.$$

Definition 1.4 (Frobenius Norm)
Let $A = (a_{ij})_{i,j} \in M_{m\times n}(\mathbb{R})$. We can regard A as a vector in $\mathbb{R}^{mn}$, and define its Frobenius norm by

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}.$$

Remark. The Frobenius norm can also be elegantly expressed using the trace of a matrix. For $A \in M_{m\times n}(\mathbb{R})$, consider the matrix product $AA^\top$. By the definition of matrix multiplication, the i-th diagonal entry of $AA^\top$ is given by:

$$(AA^\top)_{ii} = \sum_{j=1}^{n} a_{ij}(A^\top)_{ji} = \sum_{j=1}^{n} a_{ij}^2$$

Since the trace of a matrix is the sum of its diagonal entries, taking the trace of $AA^\top$ yields:

$$\mathrm{tr}(AA^\top) = \sum_{i=1}^{m}(AA^\top)_{ii} = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$$

Therefore, the Frobenius norm can be equivalently written as:

$$\|A\|_F = \sqrt{\mathrm{tr}(AA^\top)}.$$

Property 1.1

Let $A, B \in M_{n\times n}(\mathbb{R})$ and $c \in \mathbb{R}$. The Frobenius norm is defined as $\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\mathrm{tr}(AA^T)}$. It satisfies the following:

1. ***Positive Definiteness**: $\|A\|_F \geq 0$, and $\|A\|_F = 0 \iff A = \mathbf{0}$.*
2. ***Absolute Homogeneity**: $\|cA\|_F = |c|\|A\|_F$.*
3. ***Triangle Inequality**: $\|A + B\|_F \leq \|A\|_F + \|B\|_F$.*
4. ***Sub-multiplicativity**: $\|AB\|_F \leq \|A\|_F\|B\|_F$.*
5. ***Power Inequality**: $\|A^k\|_F \leq \|A\|_F^k$ for $k \in \mathbb{N}$.*
6. ***Continuity**: For any column vector $\mathbf{x} \in \mathbb{R}^n$, $\|A\mathbf{x}\| \leq \|A\|_F\|\mathbf{x}\|$. This implies that the linear map $\mathbf{x} \mapsto A\mathbf{x}$ is continuous on $\mathbb{R}^n$.*

Proof.
We will only prove the sub-multiplicativity, and the rest is left as an exercise. By the definition of matrix multiplication, the (i, j)-th entry of the product matrix AB is given by:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$$

Applying the Cauchy-Schwarz inequality to the vectors $(a_{i1}, \ldots, a_{in})$ and $(b_{1j}, \ldots, b_{nj})$, we obtain an upper bound for the square of this entry:

$$|(AB)_{ij}|^2 = \left|\sum_{k=1}^{n} a_{ik}b_{kj}\right|^2 \leq \left(\sum_{k=1}^{n} a_{ik}^2\right)\left(\sum_{k=1}^{n} b_{kj}^2\right)$$

Now, we compute the squared Frobenius norm of the matrix AB by summing over all entries ($1 \leq i \leq m$ and $1 \leq j \leq p$):

$$\begin{aligned}\|AB\|_F^2 &= \sum_{i=1}^{m}\sum_{j=1}^{p}(AB)_{ij}^2 \\ &\leq \sum_{i=1}^{m}\sum_{j=1}^{p}\left(\sum_{k=1}^{n} a_{ik}^2\right)\left(\sum_{k=1}^{n} b_{kj}^2\right)\end{aligned}$$

Notice that the sum over i only depends on the terms a_{ik}, and the sum over j only depends on the terms b_{kj}. Therefore, we can separate and factor the summations:

$$\begin{aligned}\|AB\|_F^2 &\leq \left(\sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik}^2\right)\left(\sum_{j=1}^{p}\sum_{k=1}^{n} b_{kj}^2\right) \\ &= \|A\|_F^2 \cdot \|B\|_F^2\end{aligned}$$

Finally, taking the square root of both sides, we get:

$$\|AB\|_F \leq \|A\|_F\|B\|_F$$

This completes the proof. ∎

Theorem 1.6 (Neumann Series)
Let $M \in \mathbb{R}^{n\times n}$. If $\|M\|_F < 1$, then $(I + M)$ is invertible, and its inverse is given by the convergent series:

$$(I + M)^{-1} = \sum_{i=0}^{\infty}(-1)^i M^i$$

Proof.
Let $S_k = \sum_{i=0}^{k}(-1)^i M^i$. Since $(\mathbb{R}^{n\times n}, \|\cdot\|_F)$ is a complete metric space, it suffices to show that $\{S_k\}$ is a Cauchy sequence.
For $k > m$:

$$\|S_k - S_m\|_F = \left\|\sum_{i=m+1}^{k}(-1)^i M^i\right\|_F \leq \sum_{i=m+1}^{k}\|M\|_F^i \leq \frac{\|M\|_F^{m+1}}{1 - \|M\|_F}$$

As $m \to \infty$, the right-hand side vanishes since $\|M\|_F < 1$. Thus, $\{S_k\}$ is Cauchy and converges to some matrix K.
To verify $K = (I + M)^{-1}$, observe:

$$(I + M)S_k = (I + M)\sum_{i=0}^{k}(-1)^i M^i = \sum_{i=0}^{k}(-1)^i M^i + \sum_{i=0}^{k}(-1)^i M^{i+1}$$

$$= \sum_{i=0}^{k}(-1)^i M^i + \sum_{j=1}^{k+1}(-1)^{j-1}M^j = \sum_{i=0}^{k}(-1)^i M^i - \sum_{j=1}^{k+1}(-1)^j M^j.$$

Canceling the telescoping terms, we obtain:

$$(I+M)S_k = I - (-1)^{k+1}M^{k+1}$$

Taking the limit as $k \to \infty$ (since $\|M\|_F < 1 \implies M^{k+1} \to \mathbf{0}$):

$$(I+M)K = I + \mathbf{0} = I$$

Therefore, $K = (I+M)^{-1}$. ∎

Theorem 1.7 (Chain Rule)
Let $E \subseteq \mathbb{R}^n$, $F \subseteq \mathbb{R}^m$, $f : E \to F$, $g : F \to \mathbb{R}^p$ and $\mathbb{x}_0 \in int(E)$. Suppose f is differentiable at $\mathbf{x}_0$, $f(\mathbf{x}_0)$ lies in the interior of E, and g is differentiable at $f(\mathbf{x}_0)$. Then $g \circ f$ is differentiable at $\mathbf{x}_0$, and

$$(g \circ f)'(\mathbf{x}_0) = g'(f(\mathbf{x}_0))f'(\mathbf{x}_0).$$

Proof.
Let $\mathbf{y}_0 = f(\mathbf{x}_0)$ and denote the derivatives by $A = f'(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^m$ and $B = g'(f(\mathbf{x}_0)) = g'(\mathbf{y}_0) : \mathbb{R}^m \to \mathbb{R}^p$.

Step 1: Define the error function η
Define $\eta : \mathbb{R}^m \to \mathbb{R}^p$ by:

$$\eta(\mathbf{k}) = \begin{cases} \frac{g(\mathbf{y}_0+\mathbf{k}) - g(\mathbf{y}_0) - B\mathbf{k}}{\|\mathbf{k}\|} & \text{if } \mathbf{k} \neq \mathbf{0} \\ \mathbf{0} & \text{if } \mathbf{k} = \mathbf{0} \end{cases}$$

Because g is differentiable at $\mathbf{y}_0$, $\lim_{\mathbf{k}\to\mathbf{0}} \eta(\mathbf{k}) = \mathbf{0}$, which implies η is continuous at $\mathbf{0}$. Moreover, the relation

$$g(\mathbf{y}_0 + \mathbf{k}) - g(\mathbf{y}_0) - B\mathbf{k} = \eta(\mathbf{k})\|\mathbf{k}\|$$

is true for all $\mathbf{k} \in \mathbb{R}^m$.

Step 2: Expand the composition difference
Take $\mathbf{h}$ such that $\mathbf{x}_0 + \mathbf{h} \in E$. Let $\mathbf{k} = f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$. Consider the difference:

$$\begin{aligned}(g \circ f)(\mathbf{x}_0 + \mathbf{h}) - (g \circ f)(\mathbf{x}_0) &= g(f(\mathbf{x}_0 + \mathbf{h})) - g(f(\mathbf{x}_0)) \\ &= g(\mathbf{y}_0 + \mathbf{k}) - g(\mathbf{y}_0) \\ &= B(f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0))\end{aligned}$$

$$+\eta(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0))\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\|$$

Subtracting $BA\mathbf{h}$ from both sides, we get:

$$\begin{aligned}&(g\circ f)(\mathbf{x}_0+\mathbf{h})-(g\circ f)(\mathbf{x}_0)-BA\mathbf{h}\\&=B(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h})+\eta(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0))\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\|\end{aligned}$$

Step 3: Estimate the first term
Note that:

$$\frac{\|B(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h})\|}{\|\mathbf{h}\|}\le\|B\|\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h}\|}{\|\mathbf{h}\|}$$

By the assumption that f is differentiable at $\mathbf{x}_0$, $\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h}\|}{\|\mathbf{h}\|}=0$. By the Squeeze Theorem, the first term converges to zero.

Step 4: Estimate the second term
Since f is differentiable at $\mathbf{x}_0$, f is continuous at $\mathbf{x}_0$, meaning $\lim_{\mathbf{h}\to\mathbf{0}}(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0))=\mathbf{0}$. On the other hand, η is continuous at $\mathbf{0}$ (and $\lim_{\mathbf{k}\to\mathbf{0}}\eta(\mathbf{k})=\mathbf{0}$), it follows that $\lim_{\mathbf{h}\to\mathbf{0}}\eta(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0))=\mathbf{0}$
Next, consider:

$$\begin{aligned}\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\|}{\|\mathbf{h}\|}&=\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h}+A\mathbf{h}\|}{\|\mathbf{h}\|}\\&\le\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h}\|}{\|\mathbf{h}\|}+\frac{\|A\mathbf{h}\|}{\|\mathbf{h}\|}\\&\le\underbrace{\frac{\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)-A\mathbf{h}\|}{\|\mathbf{h}\|}}_{\to0\text{ by differentiability}}+\underbrace{\|A\|}_{fixed}\end{aligned}$$

By the Squeeze Theorem:

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|\eta(f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0))\|\|f(\mathbf{x}_0+\mathbf{h})-f(\mathbf{x}_0)\|}{\|\mathbf{h}\|}=0$$

Combining Step 3 and Step 4, we conclude:

$$\lim_{\mathbf{h}\to\mathbf{0}}\frac{\|(g\circ f)(\mathbf{x}_0+\mathbf{h})-(g\circ f)(\mathbf{x}_0)-BA\mathbf{h}\|}{\|\mathbf{h}\|}=0$$

This proves that $g\circ f$ is differentiable at $\mathbf{x}_0$ with derivative $BA=g'(f(\mathbf{x}_0))f'(\mathbf{x}_0)$.

■

Definition 1.5
Let $E\subseteq\mathbf{R}^n$ be open and $f:E\to\mathbf{R}^m$. We say f is twice contiuously differentiable on E,

denoted by $f \in C^2(E)$.

1. *f is continuously differentiable (its first order partial derivatives exist and are continuous).*
2. *For each $i = 1, 2, \cdots n$, the partial derivative $\frac{\partial f}{\partial x_i} : E \to \mathbf{R}^m$ is continously differentiable on E (its second order partial derivatives exist and are continous on E).*

Theorem 1.8 (Clairaut)
Let E be an open set and $f \in C^2(E)$. Then for all $1 \le i, j \le n$ and all $\mathbf{x}_0 \in E$:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0).$$

Proof.
Let $f = (f_1, f_2, \ldots, f_m)$ where $f_k : E \to \mathbb{R}$. Without loss of generality, we can assume $f : E \to \mathbb{R}$ and the point of interest is $\mathbf{x}_0 = \mathbf{0}$. (By some shifting, $F(\mathbf{x}) = f(\mathbf{x} + \mathbf{x}_0) \Rightarrow F(\mathbf{0}) = f(\mathbf{x}_0)$). We also assume $i \neq j$.
Since $f \in C^2(E)$ and $\mathbf{0} \in E$, given any $\epsilon > 0$, there exists $\delta_0 > 0$ such that:

$$\left| \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{0}) \right| \le \frac{\epsilon}{2} \quad \text{and} \quad \left| \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) - \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{0}) \right| \le \frac{\epsilon}{2}$$

if $\|\mathbf{x}\| < 2\delta_0$.

Step 1: The Second-Order Difference (DiD Approximation)
Fix $\delta < \delta_0$, and consider the rectangle. Define the second-order difference X:

$$X = f(\delta \mathbf{e}_i + \delta \mathbf{e}_j) - f(\delta \mathbf{e}_i) - f(\delta \mathbf{e}_j) + f(\mathbf{0})$$

We can group the terms in two different ways:

$$X = \big(f(\delta \mathbf{e}_i + \delta \mathbf{e}_j) - f(\delta \mathbf{e}_i)\big) - \big(f(\delta \mathbf{e}_j) - f(\mathbf{0})\big) \tag{1}$$
$$X = \big(f(\delta \mathbf{e}_i + \delta \mathbf{e}_j) - f(\delta \mathbf{e}_j)\big) - (f(\delta \mathbf{e}_i) - f(\mathbf{0})) \tag{2}$$

Step 2: Apply Mean Value Theorem for Grouping (1)
Let $\phi(t) = f(t\mathbf{e}_i + \delta \mathbf{e}_j) - f(t\mathbf{e}_i)$. Then from (1), we have:

$$X = \phi(\delta) - \phi(0)$$

By the MVT, $X = \delta \phi'(c_i)$ for some $c_i \in (0, \delta)$ (*).
Note that $\phi'(t) = \frac{\partial f}{\partial x_i}(t\mathbf{e}_i + \delta \mathbf{e}_j) - \frac{\partial f}{\partial x_i}(t\mathbf{e}_i)$. So, $\phi'(c_i) = \frac{\partial f}{\partial x_i}(c_i \mathbf{e}_i + \delta \mathbf{e}_j) - \frac{\partial f}{\partial x_i}(c_i \mathbf{e}_i)$.

Now, let $g(s) = \frac{\partial f}{\partial x_i}(c_i\mathbf{e}_i + s\mathbf{e}_j)$. Then $g'(s) = \frac{\partial^2 f}{\partial x_j \partial x_i}(c_i\mathbf{e}_i + s\mathbf{e}_j)$. We can rewrite (*) as:

$$\phi'(c_i) = g(\delta) - g(0)$$

By applying MVT again to $g(s)$, there exists $c_j \in (0, \delta)$ such that $g(\delta) - g(0) = \delta g'(c_j)$. Thus, $\phi'(c_i) = \delta \frac{\partial^2 f}{\partial x_j \partial x_i}(c_i\mathbf{e}_i + c_j\mathbf{e}_j) \quad$ (**).

From (*) and (**), we obtain:

$$X = \delta^2 \frac{\partial^2 f}{\partial x_j \partial x_i}(c_i\mathbf{e}_i + c_j\mathbf{e}_j) \quad \text{where } 0 < c_i < \delta,\ 0 < c_j < \delta.$$

Step 3: Symmetric Argument for Grouping (2)

Similarly, starting from (2) and taking derivatives first with respect to x_j and then x_i, MVT gives us:

$$X = \delta^2 \frac{\partial^2 f}{\partial x_i \partial x_j}(d_i\mathbf{e}_i + d_j\mathbf{e}_j) \quad \text{where } 0 < d_i < \delta,\ 0 < d_j < \delta.$$

Step 4: Bounding the Error using Continuity

Notice that the norm of the points we found is strictly bounded:

$$\|c_i\mathbf{e}_i + c_j\mathbf{e}_j\| = \sqrt{c_i^2 + c_j^2} < \sqrt{\delta^2 + \delta^2} = \sqrt{2}\delta < \sqrt{2}\delta_0 < 2\delta_0$$

Because these points fall within the $2\delta_0$ radius, our initial continuity conditions apply:

$$\left| \frac{\partial^2 f}{\partial x_j \partial x_i}(c_i\mathbf{e}_i + c_j\mathbf{e}_j) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{0}) \right| \le \frac{\epsilon}{2} \implies \left| \frac{X}{\delta^2} - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{0}) \right| \le \frac{\epsilon}{2}$$

Similarly,

$$\left| \frac{X}{\delta^2} - \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{0}) \right| \le \frac{\epsilon}{2}$$

By the triangle inequality:

$$\left| \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{0}) - \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{0}) \right| \le \epsilon \quad \forall \epsilon > 0$$

Hence, we conclude that:

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{0}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{0}).$$

■

Remark (Numerical Approximation via Second-Order Difference).

From the proof above, we know that by the continuity of $\frac{\partial^2 f}{\partial x_j \partial x_i}$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$,

$$\lim_{\delta \to 0} \frac{X}{\delta^2} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{0}) \quad \text{and} \quad \lim_{\delta \to 0} \frac{X}{\delta^2} = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{0})$$

Hence, if $f \in C^2(E)$, for any point $\mathbf{x}_0$, we can use numerical approximation by a second-order finite difference without computing the derivative directly:

$$\lim_{\delta \to 0} \frac{f(\mathbf{x}_0 + \delta \mathbf{e}_i + \delta \mathbf{e}_j) - f(\mathbf{x}_0 + \delta \mathbf{e}_i) - f(\mathbf{x}_0 + \delta \mathbf{e}_j) + f(\mathbf{x}_0)}{\delta^2} = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)$$

1.2 Inverse Function Theorem

Definition 1.6 (Metric Space)
The pair (X, d) is called a metric space if a set X endowed with the function $d : X \times X \to \mathbb{R}$ such that ...

Definition 1.7 (Convergence)
Given a sequence $\{x_n\}_{n=1}^{\infty}$ in X. We say the sequence converges to x if $\forall \varepsilon > 0\, \exists N \in \mathbb{N}$ s.t. for $n \geq N$, $d(x_n, x) < \varepsilon$.

Definition 1.8 (Cauchy Sequence)
A sequence $\{x_n\}_{n=1}^{\infty}$ is said to be Cauchy if $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ s.t. for $m, n \geq N$, $d(x_n, x_m) < \varepsilon$.

Definition 1.9 (Completeness)
(X, d) is complete if every Cauchy sequence in X converges to a limit in X.

Theorem 1.9
Let E be a closed subset in X, and (X, d) is a complete metric space. Then (E, d) is also complete.

> *Proof.*
> Take any Cauchy sequence in E, which is also a Cauchy in X. By completeness of X, the sequence converges to some point in X. Since a closed set contains all limit point, the Cauchy sequence converges to a limit in E.
>
> ■

Definition 1.10 (Contraction)
Let (X, d) be a metric space and $T : X \to X$ be a map.

1. *We say T is a contraction if $d(T(x), T(y)) \leq d(x, y)\ \forall x, y \in X$.*

2. *T is a strict contraction if there exists a constant $0 < c < 1$ s.t. $d(T(x), T(y)) \leq cd(x, y)$.*

Remark. Under the standard setting, the term **contraction** simply refers to a **strict contraction**. Terence Tao use "contraction" for maps where $d(T(x), T(y)) \leq d(x, y)$ and explicitly write "strict contraction" when the constant is strictly less than 1.

Definition 1.11 (Fixed Point)
Let $T : X \to X$ be a map and $x^* \in X$. We say x^* is a fixed point of T if $T(x^*) = x^*$.

Remark. Suppose X is a subset of $\mathbb{R}$. The existence of a fixed point means that the graph of $y = f(x)$ intersect with $y = x$.

Theorem 1.10 (Contraction Mapping Theorem)
Let (X, d) be a metric space and $T : X \to X$ be a (strict) contraction. Then T has at most one fixed point. Moreover, if we assume X is non-empty and complete, then T has exactly one fixed point.

Proof.
Suppose x, y are distinct fixed point of T, consider $d(x, y) = d(T(x), T(y)) \leq cd(x, y)$ for some $c \in (0, 1)$. The inequality $(1 - c)d(x, y) \leq 0$ forces $x = y$.
Now we further assume X is complete. Take any point x_0, construct the sequence $\{x_n\}_{n=0}^{\infty}$, $x_{n+1} = T(x_n)\, \forall n \in \mathbb{N}$. Observe that $d(x_{n+1}, x_n) \leq c^n d(x_1, x_0)$ for $n \geq 1$. Suppose $n > m \geq 1$,

$$
\begin{aligned}
d(x_n, x_m) &\leq d(x_n, x_{n-1}) + d(x_{n-1}, x_{n-2}) + \cdots + d(x_{m+1}, x_m) \\
&\leq c^{n-1} d(x_1, x_0) + c^{n-2} d(x_1, x_0) + \cdots + c^m d(x_1, x_0) \\
&\leq \frac{c^m}{1 - c} d(x_1, x_0).
\end{aligned}
$$

Take $m \to \infty$, $c^m \to 0$, which shows $\{x_n\}_{n=1}^{\infty}$ is a Cauchy in X. Since X is complete, $\lim_{n\to\infty} x_n$ exists and converges to x^*. Finally,

$$
\begin{aligned}
d(x^*, T(x^*)) &\leq d(x^*, x_{n+1}) + d(x_{n+1}, T(x^*)) \\
&\leq d(x^*, x_{n+1}) + d(T(x_n), T(x^*)) \\
&\leq d(x^*, x_{n+1}) + cd(x_n, x^*).
\end{aligned}
$$

This shows $x^* = T(x^*)$.

∎

Definition 1.12 (Lipschitz continuous)
Given $(X, d_X), (Y, d_Y)$ and a map $f : X \to Y$. We say f is Lipschitz continuous if there

exists a constant $L \geq 0$ such that $d_Y(f(x_1), f(x_2)) \leq L d_X(x_1, x_2)$ for any $x_1, x_2 \in X$. Any such L is referred to as a Lipschitz constant for the function f, and f may also be referred to as $L-$Lipschitz.

Definition 1.13 (Uniform Continuity)
Let $f : X \to Y$ be a map btw (X, d_X) and (Y, d_Y). f is uniformly continous if $\forall \epsilon > 0 \, \exists \, \delta > 0$ s.t. for any $x, y \in X$ $d_X(x, y) < \delta \to d_Y(f(x), f(y)) < \epsilon$. This means δ is independent of the location of $x \in X$.

Theorem 1.11
If a function f is Lipschitz continuous, then

1. *f is continous.*
2. *f is uniformly continous.*

Proof.
Let $f : X \to Y$ be a $L-$Lipschitz continous map. Fix $\epsilon > 0$ and a point $a \in X$, choose $\frac{\varepsilon}{L}$ such that if $d(x, a) < \frac{\varepsilon}{L}$, one has $d(f(x), f(a)) \leq Ld(x, a) < L\frac{\varepsilon}{L} = \varepsilon$, which shows that f is contiuous at a. The argument for 2 is analogous. ∎

Corollary
A contraction $T : X \to X$ is a continuous function.

Proof.
A contraction with contraction ratio c is $c-$Lipschitz, by theorem (1.11), T is continous. ∎

Theorem 1.12
Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Then T is continuous.

Proof.
Let $\{e_1, e_2, \ldots, e_n\}$ be the standard basis for $\mathbb{R}^n$. Any vector $x \in \mathbb{R}^n$ can be expressed as $x = \sum_{i=1}^n x_i e_i$. By the linearity of T and the triangle inequality, we have:

$$\|T(x)\| = \left\| T\left(\sum_{i=1}^n x_i e_i\right) \right\| = \left\| \sum_{i=1}^n x_i T(e_i) \right\| \leq \sum_{i=1}^n |x_i| \|T(e_i)\|$$

In the standard Euclidean norm, the absolute value of any component is bounded by the norm of the vector, so $|x_i| \leq \|x\|$ for all $i = 1, \ldots, n$. Substituting this yields:

$$\|T(x)\| \leq \sum_{i=1}^n \|x\| \|T(e_i)\| = \|x\| \sum_{i=1}^n \|T(e_i)\|$$

Define $M = \sum_{i=1}^{n} \|T(e_i)\|$. Since $\{T(e_i)\}$ is a finite set of fixed vectors in $\mathbb{R}^m$, M is a finite non-negative constant independent of x. Thus, we obtain the boundedness condition $\|T(x)\| \le M\|x\|$.
For any $x, y \in \mathbb{R}^n$, the linearity of T implies:

$$\|T(x) - T(y)\| = \|T(x - y)\| \le M\|x - y\|$$

This demonstrates that T is M-Lipschitz continuous. By theorem (1.11), it follows that T is continuous. ■

Lemma
Let $B(\mathbf{0}, r)$ be an open ball in $\mathbb{R}^n$ and $g : B(\mathbf{0}, r) \to \mathbb{R}^n$ be a map with $g(\mathbf{0}) = \mathbf{0}$. If $\|g(\mathbf{x}) - g(\mathbf{y})\| \le \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|$ for $\mathbf{x}, \mathbf{y} \in B(\mathbf{0}, r)$, then the function $f : B(\mathbf{0}, r) \to \mathbb{R}^n$ defined by $f(\mathbf{x}) = \mathbf{x} + g(\mathbf{x})$ is one-to-one, and the image of $f(B(\mathbf{0}, r))$ contains $B(\mathbf{0}, \frac{r}{2})$.

Proof.
We first show injection. Suppose $f(\mathbf{x}) = f(\mathbf{y})$, by construction, we have $\|\mathbf{x} - \mathbf{y}\| = \|g(\mathbf{x}) - g(\mathbf{y})\| \le \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|$. The inequality $\frac{1}{2}\|\mathbf{x} - \mathbf{y}\| \le 0$ forces $\|\mathbf{x} - \mathbf{y}\| = 0$. So f is one-to-one.
Next, we want to show $B(\mathbf{0}, \frac{r}{2}) \subseteq f(B(\mathbf{0}, r))$. It suffices to show $\forall\, \mathbf{y} \in B(\mathbf{0}, \frac{r}{2})$, $\exists\, \mathbf{x} \in B(\mathbf{0}, r)$ s.t. $f(\mathbf{x}) = \mathbf{y}$. Define $F(\mathbf{x}) = \mathbf{y} - g(\mathbf{x})$. If we can find the fixed point of F, then we equivalently find $\mathbf{y} = f(\mathbf{x})$. Fix $\mathbf{y} \in B(\mathbf{0}, \frac{r}{2})$. $\|\mathbf{y}\| < \frac{r}{2}$. $\exists\, \varepsilon > 0$ s.t. $\|\mathbf{y}\| \le \frac{r-\varepsilon}{2}$. Consider the map $F : \overline{B(\mathbf{0}, r - \varepsilon)} \to \overline{B(\mathbf{0}, r - \varepsilon)}$. Take $\mathbf{x} \in \overline{B(\mathbf{0}, r - \varepsilon)}$, $\|\mathbf{x}\| \le r - \varepsilon$. Then we have $\|F(\mathbf{x})\| \le \|\mathbf{y}\| + \|g(\mathbf{x})\| \le \frac{r-\varepsilon}{2} + \|g(\mathbf{x})\| \le \frac{r-\varepsilon}{2} + \frac{r-\varepsilon}{2} = r - \varepsilon$, where $\|g(\mathbf{x})\| = \|g(\mathbf{x}) - g(\mathbf{0})\| \le \frac{1}{2}\|\mathbf{x}\| \le \frac{1}{2}(r - \varepsilon)$. So $F(\mathbf{x}) \in \overline{B(\mathbf{0}, r - \varepsilon)}$. Given a fixed $\mathbf{y}$, $\|F(\mathbf{x}_1) - F(\mathbf{x}_2)\| = \|\mathbf{y} - g(\mathbf{x}_1) - (\mathbf{y} - g(\mathbf{x}_2))\| \le \|g(\mathbf{x}_2) - g(\mathbf{x}_1)\| \le \frac{1}{2}\|\mathbf{x}_2 - \mathbf{x}_1\|$. Hence, $F : \overline{B(\mathbf{0}, r - \varepsilon)} \to \overline{B(\mathbf{0}, r - \varepsilon)}$ is a contraction $\forall\, \mathbf{y} \in B(0, \frac{r}{2})$. Since $\overline{B(\mathbf{0}, r - \varepsilon)}$ is complete, $\exists\, \mathbf{x}^* \in \overline{B(\mathbf{0}, r - \varepsilon)}$ s.t. $\mathbf{x}^* = F(\mathbf{x}^*)$ This shows $F(\mathbf{x}^*) = \mathbf{y} - g(\mathbf{x}^*) = \mathbf{x}^*$, and hence $\mathbf{x}^* + g(\mathbf{x}^*) = \mathbf{y}$.

■

Theorem 1.13 (Inverse Function Theorem)
Let E be open in $\mathbb{R}^n$ and $f : E \to \mathbb{R}^n \in C^1(E)$. Suppose the linear transformation $f'(\mathbf{x}_0) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then there exists an open set $U \subseteq E$ containing $\mathbf{x}_0$ and an open set $V \subseteq \mathbb{R}^n$ containing $f(\mathbf{x}_0)$ such that $f : U \to V$ is a bijection. Moreover, the inverse map $f^{-1} : V \to U$ is also continuously differentiable, and

$$(f^{-1})'(f(\mathbf{x_0})) = (f'(\mathbf{x}_0))^{-1}.$$

Proof.
Let $A = f'(\mathbf{x}_0)$. Firstly, f is differentiable at $\mathbf{x}_0$, so:

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{\|f(\mathbf{x}_0+\mathbf{x}) - f(\mathbf{x}_0) - A\mathbf{x}\|}{\|\mathbf{x}\|} = 0.$$

Given A is invertible, as $\mathbf{x}\to\mathbf{0}$, we have:

$$\frac{\|A^{-1}\big(f(\mathbf{x}_0+\mathbf{x}) - f(\mathbf{x}_0) - A\mathbf{x}\big)\|}{\|\mathbf{x}\|} \le \|A^{-1}\| \frac{\|f(\mathbf{x}_0+\mathbf{x}) - f(\mathbf{x}_0) - A\mathbf{x}\|}{\|\mathbf{x}\|} \to 0.$$

Step 1: Construct the auxiliary function g and show it is a contraction
Let $g(\mathbf{x}) = A^{-1}\big(f(\mathbf{x}_0+\mathbf{x}) - f(\mathbf{x}_0)\big) - \mathbf{x}$. It is clear that $g(\mathbf{0}) = \mathbf{0}$. Substituting this into our limit condition yields $\lim_{\mathbf{x}\to\mathbf{0}} \frac{\|g(\mathbf{x})\|}{\|\mathbf{x}\|} = 0$. Rewrite the limit condition as:

$$\lim_{\mathbf{x}\to\mathbf{0}} \frac{\|g(\mathbf{x}) - g(\mathbf{0}) - g'(\mathbf{0})\mathbf{x}\|}{\|\mathbf{x}\|} = 0,$$

which shows that g is differentiable at $\mathbf{0}$ with $g'(\mathbf{0}) = \mathbf{0}$.
Since both f and A are C^1, g is also C^1. By definition, there exists an $r > 0$ such that for all $\mathbf{x} \in B(\mathbf{0}, r)$:

$$\left\|\frac{\partial g}{\partial x_j}(\mathbf{x}) - \frac{\partial g}{\partial x_j}(\mathbf{0})\right\| = \left\|\frac{\partial g}{\partial x_j}(\mathbf{x})\right\| \le \frac{1}{2n^2}.$$

For $\mathbf{x} \in B(\mathbf{0}, r)$ and $\mathbf{v} \in \mathbb{R}^n$,

$$\|D_{\mathbf{v}}g(\mathbf{x})\| = \left\|\sum_{j=1}^{n} v_j \frac{\partial g}{\partial x_j}(\mathbf{x})\right\| \le \sum_{j=1}^{n} |v_j| \left\|\frac{\partial g}{\partial x_j}(\mathbf{x})\right\| \le \sum_{j=1}^{n} \|\mathbf{v}\| \frac{1}{2n^2} = \frac{1}{2n}\|\mathbf{v}\| \le \frac{1}{2}\|\mathbf{v}\|.$$

For $\mathbf{x}, \mathbf{y} \in B(\mathbf{0}, r)$, consider:

$$\begin{aligned} g(\mathbf{y}) - g(\mathbf{x}) &= \int_0^1 \frac{d}{dt} g(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))dt \quad \text{(FTC(1))} \\ &= \int_0^1 Dg(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))(\mathbf{y}-\mathbf{x})dt \quad \text{(Chain rule)} \\ &= \int_0^1 D_{\mathbf{y}-\mathbf{x}} g(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))dt \quad \text{(Since } g \text{ is } C^1\text{)} \end{aligned}$$

Therefore,

$$\|g(\mathbf{y}) - g(\mathbf{x})\| = \left\|\int_0^1 D_{\mathbf{y}-\mathbf{x}} g(\mathbf{x} + t(\mathbf{y}-\mathbf{x}))dt\right\| \le \int_0^1 \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|dt \le \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|,$$

which means g is a contraction.

Step 2: Define Φ, establish local bijection, and show Φ^{-1} is continuous
Define $\Phi(\mathbf{x}) = g(\mathbf{x}) + \mathbf{x} = A^{-1}\big(f(\mathbf{x}_0+\mathbf{x}) - f(\mathbf{x}_0)\big)$, where $g(\mathbf{0}) = \mathbf{0}$ and g is

a contraction. Apply the lemma (1.2), Φ is one-to-one on $B(\mathbf{0},r)$ and $B(\mathbf{0},\frac{r}{2}) \subseteq \Phi(B(\mathbf{0},r))$. Set $W = \Phi^{-1}(B(\mathbf{0},\frac{r}{2}))$, then $\Phi|_W : W \to B(\mathbf{0},\frac{r}{2})$ is a bijection.
Clearly, Φ is continuous. We claim that Φ^{-1} is also continuous. For $\mathbf{x},\mathbf{y} \in B(\mathbf{0},r)$:

$$\|\Phi(\mathbf{x}) - \Phi(\mathbf{y})\| = \|(\mathbf{x}-\mathbf{y}) + (g(\mathbf{x}) - g(\mathbf{y}))\| \geq \|\mathbf{x}-\mathbf{y}\| - \|g(\mathbf{x}) - g(\mathbf{y})\| \geq \frac{1}{2}\|\mathbf{x}-\mathbf{y}\|.$$

Let $\mathbf{x} = \Phi^{-1}(\mathbf{u})$ and $\mathbf{y} = \Phi^{-1}(\mathbf{v})$, with $\mathbf{u},\mathbf{v} \in B(\mathbf{0},\frac{r}{2}) \subseteq \Phi(B(\mathbf{0},r))$, then:

$$\begin{aligned}\|\Phi(\Phi^{-1}(\mathbf{u})) - \Phi(\Phi^{-1}(\mathbf{v}))\| &\geq \frac{1}{2}\|\Phi^{-1}(\mathbf{u}) - \Phi^{-1}(\mathbf{v})\| \\ \|\Phi^{-1}(\mathbf{u}) - \Phi^{-1}(\mathbf{v})\| &\leq 2\|\mathbf{u}-\mathbf{v}\|.\end{aligned}$$

This shows that Φ^{-1} is Lipschitz continuous and hence continuous.

Step 3: Prove Φ^{-1} is differentiable at $\mathbf{0}$

We claim that the derivative of Φ^{-1} at $\mathbf{0}$ is the identity map id. Let $\mathbf{z} \in B(\mathbf{0},\frac{r}{2})\setminus\{\mathbf{0}\}$. Since $\Phi|_W$ is bijective, let $\mathbf{z} = \Phi(\mathbf{x}) = \mathbf{x}+g(\mathbf{x})$, which means $\mathbf{x} = \Phi^{-1}(\mathbf{z})$ and $\mathbf{x} \neq \mathbf{0}$. We want to show:

$$\lim_{\mathbf{z}\to\mathbf{0}} \frac{\|\Phi^{-1}(\mathbf{z}) - \Phi^{-1}(\mathbf{0}) - \mathrm{id}(\mathbf{z})\|}{\|\mathbf{z}\|} = 0.$$

Since $\Phi^{-1}(\mathbf{0}) = \mathbf{0}$, the expression inside the limit simplifies to:

$$\frac{\|\Phi^{-1}(\mathbf{z}) - \mathbf{z}\|}{\|\mathbf{z}\|} = \frac{\|\mathbf{x} - (\mathbf{x} + g(\mathbf{x}))\|}{\|\mathbf{z}\|} = \frac{\|-g(\mathbf{x})\|}{\|\mathbf{z}\|} = \frac{\|g(\mathbf{x})\|}{\|\mathbf{x}\|} \cdot \frac{\|\mathbf{x}\|}{\|\mathbf{z}\|}.$$

From the Lipschitz continuity in Step 2, taking $\mathbf{v} = \mathbf{0}$, we have $\|\Phi^{-1}(\mathbf{z})\| \leq 2\|\mathbf{z}\|$, which implies $\frac{\|\mathbf{x}\|}{\|\mathbf{z}\|} \leq 2$. Furthermore, as $\mathbf{z} \to \mathbf{0}$, sequential continuity dictates that $\mathbf{x} = \Phi^{-1}(\mathbf{z}) \to \mathbf{0}$. Using the Squeeze Theorem:

$$\lim_{\mathbf{z}\to\mathbf{0}} \frac{\|\Phi^{-1}(\mathbf{z}) - \mathbf{z}\|}{\|\mathbf{z}\|} = \lim_{\mathbf{x}\to\mathbf{0}} \left(\frac{\|g(\mathbf{x})\|}{\|\mathbf{x}\|} \cdot \frac{\|\mathbf{x}\|}{\|\mathbf{z}\|}\right) \leq \lim_{\mathbf{x}\to\mathbf{0}} \left(\frac{\|g(\mathbf{x})\|}{\|\mathbf{x}\|} \cdot 2\right) = 0 \cdot 2 = 0.$$

This proves that Φ^{-1} is differentiable at $\mathbf{0}$ with $(\Phi^{-1})'(\mathbf{0}) = \mathrm{id}$.

Step 4: Prove f^{-1} is differentiable at $f(\mathbf{x}_0)$

Recall the definition $\Phi(\mathbf{x}) = A^{-1}\big(f(\mathbf{x}_0 + \mathbf{x}) - f(\mathbf{x}_0)\big)$. Multiplying by A and rearranging yields:

$$f(\mathbf{x} + \mathbf{x}_0) = A\Phi(\mathbf{x}) + f(\mathbf{x}_0).$$

Since Φ is a local bijection, f is also one-to-one and onto near $\mathbf{x}_0$. Let $\mathbf{y} = f(\mathbf{x}+\mathbf{x}_0)$, then $\mathbf{y} = A\Phi(\mathbf{x}) + f(\mathbf{x}_0)$. We can express $\mathbf{x}$ as:

$$\mathbf{x} = \Phi^{-1}\big(A^{-1}(\mathbf{y} - f(\mathbf{x}_0))\big).$$

Thus, the inverse function f^{-1} is given by:

$$f^{-1}(\mathbf{y}) = \mathbf{x} + \mathbf{x}_0 = \Phi^{-1}\big(A^{-1}(\mathbf{y} - f(\mathbf{x}_0))\big) + \mathbf{x}_0.$$

We apply the Chain Rule to differentiate f^{-1} at $f(\mathbf{x}_0)$:

$$\begin{aligned}
(f^{-1})'(f(\mathbf{x}_0)) &= (\Phi^{-1})'\big(A^{-1}(f(\mathbf{x}_0) - f(\mathbf{x}_0))\big) \circ (A^{-1})'(f(\mathbf{x}_0) - f(\mathbf{x}_0)) \\
&= (\Phi^{-1})'(\mathbf{0}) \circ A^{-1}(\mathbf{0}) \\
&= \mathrm{id} \circ A^{-1} = A^{-1} = (f'(\mathbf{x}_0))^{-1}.
\end{aligned}$$

■

1.3 Implicit Function Theorem

Definition 1.14 (Graph)
Given a function $f : U \subseteq X \to Y$, the graph of f is defined as the following subset of Cartesian product space:

$$\mathcal{G}(f) = \{(x, f(x)) : x \in X\}.$$

Example 1.1
Let $f : U \subseteq \mathbb{R}^{n-1} \to \mathbb{R}$ be a function. Define the graph of f as a $(n-1)$–dimension subset in $\mathbb{R}^n$:

$$\{(x_1, \ldots, x_{n-1}, f(x_1, \ldots, x_{n-1})) \in \mathbb{R}^n \mid (x_1, \ldots, x_{n-1}) \in U\}.$$

Definition 1.15 (Level Set)
*Let $U \subseteq \mathbb{R}^n$ be an open set and $f : U \to \mathbb{R}$ be a real-valued function. For a given constant $c \in \mathbb{R}$, the set of all points $\mathbf{x} \in U$ that satisfy the equation $f(\mathbf{x}) = c$ forms a subset of U. This subset is formally called the **level set** of f at level c, denoted by:*

$$L_c = \{\mathbf{x} \in U \mid f(\mathbf{x}) = c\}.$$

Example 1.2

1. ***Level Curve ($n = 2$):*** *A 1D level set in a 2D space is defined by a 2-variable equation $f(x, y) = c$. We call this a **level curve / contour line / isoline** of f at level c.*

2. ***Level Surface ($n = 3$):*** *A 2D level set in a 3D space is defined by a 3-variable equation $f(x, y, z) = c$. We call this a **level surface / isosurface** of f at level c.*

3. ***Level Hypersurface ($n > 3$):*** *An $(n-1)$-dimensional level set in an n-dimensional space is defined by $f(x_1, \ldots, x_n) = c$. We simply call this a **level hypersurface**.*

Remark. Let $U \subseteq \mathbb{R}^3$ and a three-variable function $f : U \to \mathbb{R}$ defined by $w = f(x, y, z)$.

1. The graph of f is defined as $\{(x, y, z, f(x, y, z)) : (x, y, z) \in U\}$, which is a 3-dimensional subset in $\mathbb{R}^4$.

2. Set $w = c$. Given this constant c, we have a level surface of f at level c, which is a 2-dimensional subset $S = \{(x, y, z) \in U : f(x, y, z) = c\}$ in $\mathbb{R}^3$.

3. Locally, S can be expressed as the graph of a 2-variable function, $S' = \{(x, y, z(x, y)) : (x, y) \in \mathbb{R}^2\}$, which is a 2-dimensional subset in $\mathbb{R}^3$, where $S' \subseteq S$.

Remark. A related but distinct concept is the intersection of two surfaces. For example, suppose we have a double cone and a cutting plane, both of which are level surfaces defined by 3-variable equations. Their intersection forms a conic section (a quadratic curve). Intuitively, a system of 3 variables restricted by 2 independent equations reduces the degrees of freedom to 1 ($3 - 2 = 1$), resulting in a 1-dimensional curve. However, this is not the same concept as a "level curve." A standard level curve is defined by a single equation in a 2D space (e.g., $f(x, y) = c$), whereas this intersection is a 1D curve living in a 3D space, defined by a system of two equations.

Theorem 1.14 (Implicit Function Theorem)
Let E be an open set in $\mathbb{R}^n$, and $f : E \to \mathbb{R}$ is C^1. Suppose $y \in E$ such that $f(y) = 0$ and $f_{x_n} = \frac{\partial f}{\partial x_n} \neq 0$, then:

1. *There exists an open subset $U \subseteq \mathbb{R}^{n-1}$ containing $(y_1, y_2, \cdots, y_{n-1})$, and a map $g : U \to \mathbb{R}$ such that $g(y_1, y_2, \cdots, y_{n-1}) = y_n$.*

2. *The level set $\{x \in V : f(x) = 0\}$ is a graph of a function of g over U, i.e. $\{x \in V : f(x) = 0\} = \{(x_1, x_2, \cdots, x_{n-1}, g(x_1, x_2, \cdots, x_{n-1})) : (x_1, x_2, \cdots, x_{n-1}) \in U\}$.*

3. *For $j = 1, 2, \cdots, n-1$,*

$$\frac{\partial g}{\partial x_j}(y_1, y_2, \cdots, y_{n-1}) = -\frac{\frac{\partial f}{\partial x_j}(y)}{\frac{\partial f}{\partial x_n}(y)}.$$

Proof.

Step 1: Define a new function F and check its Jacobian
Define $F : E \to \mathbb{R}^n$ by $(x_1, x_2, \ldots, x_n) \mapsto (x_1, x_2, \ldots, x_{n-1}, f(x_1, x_2, \ldots, x_n))$.

Since $f \in C^1(E)$, it follows that $F \in C^1(E)$. The Jacobian matrix of F at x is:

$$DF(x) = \begin{pmatrix} - & \nabla x_1 & - \\ - & \nabla x_2 & - \\ & \vdots & \\ - & \nabla f & - \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_1}(x) & \frac{\partial f}{\partial x_2}(x) & \cdots & \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

We are given that $\frac{\partial f}{\partial x_n}(y) \neq 0$. Therefore, the determinant evaluated at y is:

$$\det(DF(y)) = \frac{\partial f}{\partial x_n}(y) \neq 0,$$

which means $DF(y)$ is invertible.

Step 2: Apply the Inverse Function Theorem to construct local inverse

By the Inverse Function Theorem, there exists an open set $V \subseteq E$ and $W \subseteq \mathbb{R}^n$ such that $y \in V$, $F(y) \in W$, and $F : V \to W$ is a C^1 diffeomorphism. Write the inverse in coordinates as $F^{-1}(u) = (h_1(u), h_2(u), \ldots, h_n(u))$ for $u = (u_1, u_2, \ldots, u_n) \in W$. By construction, $F(F^{-1}(u)) = u$, which implies:

$$F(h_1(u), h_2(u), \ldots, h_n(u)) = (u_1, u_2, \ldots, u_n).$$

Using the definition of F, the left-hand side expands to:

$$(h_1(u), h_2(u), \ldots, h_{n-1}(u), f(h_1(u), \ldots, h_n(u))).$$

Equating the components element by element, we obtain:

$$h_1(u) = u_1, \quad h_2(u) = u_2, \quad \ldots, \quad h_{n-1}(u) = u_{n-1},$$

$$f(h_1(u), h_2(u), \ldots, h_n(u)) = u_n \implies f(u_1, u_2, \ldots, u_{n-1}, h_n(u)) = u_n \quad (*).$$

Thus, the inverse function takes the form:

$$F^{-1}(u_1, u_2, \ldots, u_n) = (u_1, u_2, \ldots, u_{n-1}, h_n(u_1, u_2, \ldots, u_n)).$$

Step 3: Restrict to $u_n = 0$ and define the implicit function g

If we restrict the domain to $u_n = 0$, by $(*)$, we have:

$$f(u_1, \ldots, u_{n-1}, h_n(u_1, \ldots, u_{n-1}, 0)) = 0.$$

Set an open set $U = \{x' = (x_1, x_2, \ldots, x_{n-1}) \in \mathbb{R}^{n-1} \mid (x', 0) \in W\}$. Since $F(y_1, \ldots, y_n) = (y_1, \ldots, y_{n-1}, f(y_1, \ldots, y_n)) = (y_1, \ldots, y_{n-1}, 0) \in W$, U is an open set containing $(y_1, y_2, \ldots, y_{n-1}) = y'$.

Define $g : U \longrightarrow \mathbb{R}$ by $g(x') = g(x_1, \dots, x_{n-1}) = h_n(x_1, \dots, x_{n-1}, 0)$. Note that

$$y = \underbrace{(y', y_n)}_{\in V} = F^{-1}F(y) = F^{-1}(y', 0) = (y', h_n(y', 0)) = (y', g(y)).$$

So $y_n = g(y')$.

Step 4: Show the zero set is locally the graph of g

Claim: $\{x \in V : f(x) = 0\} = \{(x', g(x')) : x' \in U\}$. Namely, we want to show the solution set of the equation $f(x) = 0$ (A) is exactly the graph of the function $g(x')$ (B). Recall $U = \{x' \in \mathbb{R}^{n-1} : (x', 0) \in W\}$.

($\subseteq$) Take $x = (x', x_n) \in A$, where $x \in V$ and $f(x) = 0$.
Apply F first, $F(x) = (x', f(x)) = (x', 0)$, where $F(x) \in W$, so $x' \in U$.
Apply F^{-1} sequentially, $\underbrace{F^{-1}F(x)}_{x} = F^{-1}(x', 0) = (x', h_n(x', 0)) = (x', g(x'))$.
So $x \in B$.

($\supseteq$) Take $x = (x', x_n) \in B$, where $x' \in U$ and $x = (x', g(x'))$.
Since $x' \in U$, $(x', 0) \in W$. Apply F^{-1} first, $F^{-1}(x', 0) = (x', h_n(x', 0)) = (x', g(x')) = x$, where $x \in V$.
Apply F sequentially, $\underbrace{FF^{-1}(x', 0)}_{(x',0)} = F(x) = (x', f(x))$. So $x \in A$.

Step 5: Obtain IFT formula

Now we have $(x_1, \dots, x_{n-1}) \in U \subseteq \mathbb{R}^{n-1}$ (where U is open), $f(x', g(x')) = 0$, and $y_n = g(y')$.
Since U is open, there exists $\delta > 0$ such that if $|t| < \delta$, then $y' + te_j \in U$, which implies:

$$f(y' + te_j, g(y' + te_j)) = 0$$

Let $\phi(t) = f(y' + te_j, g(y' + te_j))$. Thus, $\phi(t) = 0$ when $|t| < \delta$. Therefore, $\phi'(t) = 0$ for $|t| < \delta$.
Differentiating with respect to t on both sides by the chain rule:

$$\phi'(t) = \frac{\partial f}{\partial x_j}(y)\frac{dx_j}{dt}(t) + \frac{\partial f}{\partial x_n}(y) \cdot \frac{\partial g}{\partial x_j}(y')\frac{dx_j}{dt}(t).$$

Since $\frac{dx_j}{dt} = 1$, we evaluate at $t = 0$:

$$0 = \frac{\partial f}{\partial x_j}(y) + \underbrace{\frac{\partial f}{\partial x_n}(y)}_{\neq 0}\frac{\partial g}{\partial x_j}(y')$$

$$\therefore \frac{\partial g}{\partial x_j}(y') = -\frac{\frac{\partial f}{\partial x_j}(y)}{\frac{\partial f}{\partial x_n}(y)}$$

■

Definition 1.16 (Manifold)
*A subset $M \subseteq \mathbb{R}^n$ is called a k-dimensional manifold if, for every point $p \in M$, there exists a neighborhood V of p in M that is **diffeomorphic** to an open set U in $\mathbb{R}^k$.*

Remark. In advanced calculus (e.g., Spivak's *Calculus on Manifolds*), being a manifold is often established through equivalent conditions. For a subset $M \subseteq \mathbb{R}^n$ to be a k-dimensional manifold locally around a point p, the following three perspectives are mathematically equivalent:

1. **The Geometric View (Diffeomorphism):** Locally, M is diffeomorphic to an open set in $\mathbb{R}^k$.
2. **The Analytical View (Graph of a Function):** Locally, M can be written as the **graph** of a C^1 function (by expressing $n-k$ variables in terms of the other k variables).
3. **The Algebraic View (Level Set / Zero Set):** Locally, M is the solution set of $n-k$ equations where the Jacobian matrix has full rank.

Applications in Economics

1. Indifference Curves and Utility Functions
   Let $U(x, y)$ be a utility function representing a consumer's preferences over two goods. An **indifference curve** is fundamentally a 1D **level curve** defined by the equation $U(x, y) = c$, where c is a constant utility level. By the IFT, provided the marginal utility $\frac{\partial U}{\partial y} \neq 0$, we can locally express y as a smooth function of x (i.e., $y = g(x)$). Most importantly, the theorem gives us the exact slope of this curve, which economists call the **Marginal Rate of Substitution (MRS)**:

$$\frac{dy}{dx} = -\frac{\frac{\partial U}{\partial x}}{\frac{\partial U}{\partial y}} = -\frac{MU_x}{MU_y}$$

   The IFT guarantees we can compute this trade-off without ever needing to explicitly isolate y.

2. Isoquant Curves and Production Functions
   Similarly, in producer theory, a production function is given by $Q = F(K, L)$,

where K is capital and L is labor. An **isoquant** is the level curve $F(K, L) = c$. Applying the IFT, we can find the slope of the isoquant, known as the **Marginal Rate of Technical Substitution (MRTS)**:

$$\frac{dK}{dL} = -\frac{\frac{\partial F}{\partial L}}{\frac{\partial F}{\partial K}} = -\frac{MP_L}{MP_K}$$

This tells the firm exactly how much capital can be replaced by one unit of labor while maintaining the same output level c.

3. Comparative Statics

   Beyond 2D level curves, the generalized n-dimensional IFT is the foundation of **Comparative Statics** in macroeconomics and general equilibrium theory. Suppose an economic equilibrium is defined by a complex system of equations $F(\mathbf{y}, \mathbf{x}) = \mathbf{0}$, where $\mathbf{y}$ are endogenous variables (e.g., prices, GDP) and $\mathbf{x}$ are exogenous policy parameters (e.g., tax rates, money supply). Even if the system is too complex to explicitly solve for the equilibrium state $\mathbf{y}^* = g(\mathbf{x})$, the IFT guarantees that this relationship exists locally, provided the Jacobian matrix $D_{\mathbf{y}}F$ is invertible. Furthermore, it allows economists to calculate $\frac{\partial \mathbf{y}^*}{\partial \mathbf{x}}$, precisely predicting how a shift in government policy will ripple through the equilibrium variables.