MA225 Differentiation

Ian Rogers[*]

May 21, 2007

Contents

---

[*] Based on the lectures of Dr Vassili Gelfreich

1 Continuous functions on $\mathbb{R}^n$

1.1 Norms and distances on $\mathbb{R}^n$

Definition 1.1.1. $\| \ \|, \| \ \|'$ equivalent if $\exists\, \lambda_1, \lambda_2 > 0$ s.t. $\forall\, \mathbf{x} \in \mathbb{R}^n$

$$\|\mathbf{x}\| \leq \lambda_1 \|\mathbf{x}\|'$$
$$\|\mathbf{x}\|' \leq \lambda_2 \|\mathbf{x}\|$$

Theorem 1.1.2. *All norms on $\mathbb{R}^n$ equivalent.*

Definition 1.1.3 (Operator Norm)**.** $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$

$$\|T\|_{\text{op}} = \sup_{\|\mathbf{x}\|=1} \|T\mathbf{x}\|$$

Definition 1.1.4. *Open ball* $B(\mathbf{x}_0, \varepsilon) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\| < \varepsilon\}$

1.2 Continuity

Definition 1.2.1. $A \subseteq \mathbb{R}^n$. $f\colon A \to \mathbb{R}^p$ continuous at $\mathbf{x}_0 \in A$ if $\forall\, \varepsilon > 0\ \exists\, \delta > 0$ s.t. $(x \in A,\ \|\mathbf{x} - \mathbf{x}_0\| < \delta) \Rightarrow \|f(\mathbf{x}) - f(\mathbf{x}_0)\| < \varepsilon$.

Proposition 1.2.2. *$f\colon A \to \mathbb{R}^p$ is continuous at $\mathbf{x}_0 \in A$ iff every component of f is continuous at $\mathbf{x}_0$.*

Example 1.2.3. $f\colon \mathbb{R}^2 \to \mathbb{R}$

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

is:

- continuous function of $x\ \forall\, y$
- continuous function of $y\ \forall\, x$
- not continuous at $(0,0)$

1.3 Sequentially compact set

Definition 1.3.1.
- $U \subseteq \mathbb{R}^n$ *open* if $\forall\, \mathbf{x} \in U\ \exists\, \varepsilon > 0$ s.t. $B(\mathbf{x}, \varepsilon) \subset U$.
- $Z \subseteq \mathbb{R}^n$ *closed* if $\mathbb{R}^n \setminus Z$ is open.

Proposition 1.3.2. *$Z \subseteq \mathbb{R}^n$ closed iff every convergent sequence in Z has limit in Z.*

Proof. ($\Rightarrow$) Z closed. Consider convergent sequence $(\mathbf{x}_k) \subset Z$. $\mathbf{a} = \lim_{k\to\infty} \mathbf{x}_k \in \mathbb{R}^n$. $\forall\, \varepsilon > 0\ \exists\, K$ s.t. $\|\mathbf{x} - \mathbf{a}\| < \varepsilon \quad \forall\, k \geq K$. Suppose $\mathbf{a} \notin Z$. Then $\mathbf{a} \in \mathbb{R}^n \setminus Z$ open, so $\exists\, \varepsilon_0 > 0$ s.t. $B(\mathbf{a}, \varepsilon) \subset \mathbb{R}^n \setminus Z$. Contradiction from above.

($\Leftarrow$) Suppose Z not closed $\Rightarrow \mathbb{R}^n \setminus Z$ not open $\Rightarrow \exists\, \mathbf{a} \in \mathbb{R}^n \setminus Z$ s.t. $\forall\, \varepsilon > 0$ $B(\mathbf{a}, \varepsilon) \cap Z \neq \emptyset$.

$\forall\, k \quad B\left(\mathbf{a}, \frac{1}{k}\right) \cap Z \neq \emptyset$ take $\mathbf{x}_k$ in this intersection.

$(\mathbf{x}_k) \subset Z$ and $\lim_{k\to\infty} \mathbf{x}_k = \mathbf{a} \notin Z$. Contradiction. $\square$

Definition 1.3.3. Set X *sequentially compact* if every sequence in X has convergent subsequence (which converges to a point of X).

Proposition 1.3.4. *Every bounded sequence in $\mathbb{R}$ has convergent subsequence.*

Proof. (Lion Hunting) Sequence bounded $\Rightarrow \exists\, c, d \in \mathbb{R}$ s.t. $c \leq x_i \leq d\ \forall\, i$.

$$I_0 = [c, d] = \left[c, \frac{c+d}{2}\right] \cup \left[\frac{c+d}{2}, d\right]$$

Take x_{i_0} to be any element of (x_i).

$$I_1 = \left[c, \frac{c+d}{2}\right] \text{ or } \left[\frac{c+d}{2}, d\right]$$

and contains ∞-many x_i. Let $x_{i_1} \in I_1$ with $i_i > i_0$.

Repeat. Length $I_k = \frac{d-c}{2^k} \xrightarrow[k\to\infty]{} 0$ so x_i converges. $\square$

Theorem 1.3.5. *$X \subseteq \mathbb{R}^n$ sequentially compact iff X closed and bounded.*

Proof. ($\Rightarrow$) Take convergent sequence in X. Any subsequence converges to same limit, which is in X.

By proposition 1.3.2 X is closed. If X not bounded then $\forall K \in \mathbb{N}\ \exists \mathbf{x}_K \in X$ s.t. $\|\mathbf{x}_K\| > K$. Any subsequence not bounded. X sequentially compact $\Rightarrow \exists$ convergent subsequence of $(\mathbf{x}_k)$. Contradiction so X bounded.

($\Leftarrow$) X closed, bounded. Take $(\mathbf{x}_k) \subset X$. $(\mathbf{x}_k)$ bounded. $\mathbf{x}_k = (x_k[1], \ldots, x_k[n])$.

$x_k[j] \in \mathbb{R}$ $j = 1, \ldots, n$. By prop 1.3.4 $\exists$ subsequence s.t. $x_{k_i}[1]$ converges. Consider $(\mathbf{x}_{k_i})_{i=1}^{\infty}$.

From these $\exists$ subseqence s.t. 2nd component converges, …, so subsequence s.t. all components converges. Limit in X as X closed, so X sequentially compact. $\square$

1.4 Extreme Value Theorem

Proposition 1.4.1. *$X \subseteq \mathbb{R}^n$ sequentially compact and $f\colon X \to \mathbb{R}^p$ continuous then $f(X)$ sequentially compact.*

Proof. $\mathbf{y}_i \in f(X)$ sequence. $\forall\, i\ \exists\, \mathbf{x}_i \in X$ s.t. $\mathbf{y}_i = f(\mathbf{x}_i)$. X sequentially compact so $\exists\, \mathbf{x}_{i_k}$ convergent, $\lim_{k\to\infty} \mathbf{x}_{i_k} = \mathbf{a} \in X$.

$\mathbf{y}_{i_k} = f(\mathbf{x}_{i_k})$. f continuous $\Rightarrow \lim_{k\to\infty} f(\mathbf{x}_{i_k}) = f(\mathbf{a}) \in f(X)$. Hence $f(X)$ sequentially compact. $\square$

Theorem 1.4.2 (Extreme Value Theorem)**.** *If $X \subseteq \mathbb{R}^n$ sequentially compact and $f\colon X \to \mathbb{R}$ continuous, then f bounded and attains bounds.*

Proof. 1.4.1 $\Rightarrow$ $f(X)$ sequentially compact $\underset{1.3.5}{\Rightarrow}$ $f(X)$ closed, bounded. Hence $a = \sup f(X) \in \mathbb{R}$. $\forall\, n \in \mathbb{N}\ \exists\, y_n \in f(X)$ s.t. $y_n > a - \frac{1}{n}$.

y_n converges to a. $f(X)$ sequentially compact so $a \in f(X)$, hence $\exists\, \mathbf{x}_0 \in X$ s.t. $a = f(\mathbf{x}_0)$.

$$f(\mathbf{x}_0) = \sup f(X) \iff \forall\, \mathbf{x} \in X,\ f(\mathbf{x}) \leq f(\mathbf{x}_0)$$

☐

Definition 1.4.3. $\mathbb{S}^{n-1} = \{\mathbf{x} : \|\mathbf{x}\|_2 = 1\} \subset \mathbb{R}^n$ closed and bounded so sequentially compact.

2 Differentiation

2.1 Derivatives

2.1.1 Little o notation

$\varphi\colon \mathbb{R} \to \mathbb{R}$, $\varphi(h) = o(h)$ if $\lim_{h\to 0} \frac{\varphi(h)}{h} = 0$.

$\alpha\colon \mathbb{R}^n \to \mathbb{R}^p$, $\alpha(\mathbf{h}) = o(\mathbf{h})$ if $\lim_{h\to 0} \frac{\|\alpha(\mathbf{h})\|}{\|\mathbf{h}\|} = 0$.

Proposition 2.1.1. *$f\colon \mathbb{R} \to \mathbb{R}$ is diff at $x_0 \in \mathbb{R}$ iff $\exists\, a \in \mathbb{R}$ s.t.*

$$f(x_0 + h) = f(x_0) + ah + \varphi(x_0, h) \quad (\varphi(x_0, h) = o(h))$$

Proof. ($\Rightarrow$) f diff at $x_0 \Rightarrow a = \lim_{h\to 0} \frac{f(x_0+h)-f(x_0)}{h}$.

Let $\varphi(x_0, h) = f(x_0 + h) - f(x_0) - ah$.

$$\begin{aligned}\lim_{h\to 0} \frac{\varphi(x_0,h)}{h} &= \lim_{h\to 0} \frac{f(x_0+h) - f(x_0) - ah}{h} \\ &= \lim_{h\to 0} \frac{f(x_0+h) - f(x_0)}{h} - a \\ &= 0\end{aligned}$$

($\Leftarrow$)

$$\frac{f(x_0+h) - f(x_0)}{h} = \frac{ah - \varphi(x_0,h)}{h} = a + \underbrace{\frac{\varphi(x_0,h)}{h}}_{\xrightarrow[h\to 0]{} 0}$$

so limit exists. ☐

Definition 2.1.2 (Partial derivatives)**.** $f\colon \mathbb{R}^n \to \mathbb{R}$ $\mathbf{x} \in \mathbb{R}^n$.

$$\frac{\partial f}{\partial x_1}(\mathbf{x}) = \lim_{h\to 0} \frac{f(x_1+h, x_2, \dots, x_n) - f(\mathbf{x})}{h}$$

if this limit exists, so

$$\frac{\partial f}{\partial x_k}(\mathbf{x}) = \lim_{h\to 0} \frac{f(\mathbf{x}+\mathbf{e}_k h) - f(\mathbf{x})}{h}$$

Definition 2.1.3. $f\colon \mathbb{R}^n \to \mathbb{R}^p$, $\mathbf{x}_0 \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^n$. If $\lim_{t\to 0} \frac{f(\mathbf{x}_0+t\mathbf{v})-f(\mathbf{x}_0)}{t}$ exists is called *derivative* of f in *direction* $\mathbf{v}$ at $\mathbf{x}_0$.

Definition 2.1.4. $X \subseteq \mathbb{R}^n$ open. $f\colon X \to \mathbb{R}^p$ is *differentiable* at $\mathbf{x}_0 \in X$ if $\exists$ linear $d_{\mathbf{x}_0} f\colon \mathbb{R}^n \to \mathbb{R}^p$ (the *differential*) s.t.

$$f(\mathbf{x}_0+\mathbf{h}) = f(\mathbf{x}_0) + d_{\mathbf{x}_0} f(\mathbf{h}) + \alpha(\mathbf{x}_0, \mathbf{h}) \quad (\alpha(\mathbf{x}_0,\mathbf{h}) = o(\mathbf{h}))$$

Proposition 2.1.5. *$T\colon \mathbb{R}^n \to \mathbb{R}^p$ is linear. Then T is diff at each $\mathbf{x}_0 \in \mathbb{R}^n$ and $d_{\mathbf{x}_0} T = T$.*

Proof. Let $L = T$

$$\lim_{h\to 0} \frac{T(\mathbf{x}_0+\mathbf{h}) - T(\mathbf{x}_0) - L(\mathbf{h})}{\|\mathbf{h}\|} = \lim_{h\to 0} \frac{T(\mathbf{x}_0) - T(\mathbf{x}_0) + T(\mathbf{h}) - L(\mathbf{h})}{\|\mathbf{h}\|} = 0 \quad \square$$

Proposition 2.1.6 (Uniqueness of differential)**.** *If $d_{\mathbf{x}_0} f$ exists it is unique.*

Proof. Suppose $\exists\, L, \tilde{L}\colon \mathbb{R}^n \to \mathbb{R}^p$ linear s.t.

$$\begin{aligned} f(\mathbf{x}_0+\mathbf{h}) &= f(\mathbf{x}_0) + L(\mathbf{h}) + \alpha(\mathbf{x}_0,\mathbf{h}) \\ &= f(\mathbf{x}_0) + \tilde{L}(\mathbf{h}) + \tilde{\alpha}(\mathbf{x}_0,\mathbf{h}) \end{aligned}$$

where $a, \tilde{a} = o(\mathbf{h})$. $0 = L(\mathbf{h}) - \tilde{L}(\mathbf{h}) + \alpha(\mathbf{x}_0,\mathbf{h}) - \tilde{\alpha}(\mathbf{x}_0,\mathbf{h})$,
so $L(\mathbf{h}) - \tilde{L}(\mathbf{h}) = \tilde{\alpha}(\mathbf{x}_0,\mathbf{h}) - \alpha(\mathbf{x}_0,\mathbf{h})$.

Consider $\mathbf{0} \neq \mathbf{y} \in \mathbb{R}^n$. $\mathbf{h}_n = \frac{\mathbf{y}}{n}$.

$$\begin{aligned} L(\mathbf{y}) - \tilde{L}(\mathbf{y}) &= L(n\mathbf{h}_n) - \tilde{L}(n\mathbf{h}_n) \\ &= n(L(\mathbf{h}_n) - \tilde{L}(\mathbf{h}_m)) \\ &= n(\tilde{\alpha}(\mathbf{x}_0,\mathbf{h}_n) - \alpha(\mathbf{x}_0,\mathbf{h}_n)) \\ &= \frac{\tilde{\alpha}(\mathbf{x}_0,\mathbf{h}_n) - \alpha(\mathbf{x}_0,\mathbf{h}_n)}{\|\mathbf{h}_n\|}\|\mathbf{y}\| \\ &\xrightarrow[n\to\infty]{} 0 \end{aligned}$$

so $L(\mathbf{y}) = \tilde{L}(\mathbf{y})$. $\square$

2.2 Properties of differentiable functions

Throughout $X \subseteq \mathbb{R}^n$ open.

Proposition 2.2.1. *$f\colon X \to \mathbb{R}^p$ is diff at $\mathbf{x}_0 \in X$ iff each component of $f = (f_1, \ldots, f_p)$ is diff at $\mathbf{x}_0$.*

Proof.

$$\begin{gathered} f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + \alpha(\mathbf{x}_0, \mathbf{h}) \quad (\alpha(\mathbf{x}_0, \mathbf{h}) = o(\mathbf{h})) \\ \iff \\ f_i(\mathbf{x}_0 + \mathbf{h}) = f_i(\mathbf{x}_0) + L_i(\mathbf{h}) + \alpha_i(\mathbf{x}_0, \mathbf{h}) \quad \forall\, i \end{gathered}$$

where $\alpha_i(\mathbf{x}_0, \mathbf{h}) = o(\mathbf{h})$. ☐

Proposition 2.2.2. *If $f\colon X \to \mathbb{R}^p$ is diff at $\mathbf{x}_0 \in X$ then f is continuous at $\mathbf{x}_0$.*

Proof. Fix $\varepsilon > 0$

$$\begin{aligned} \|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\| &= \|L(\mathbf{h}) + \alpha(\mathbf{x}_0, \mathbf{h})\| \\ &\leq \|L(\mathbf{h})\| + \|\alpha(\mathbf{x}_0, \mathbf{h})\| \end{aligned}$$

$\alpha = o(\mathbf{h})$ so $\exists\, \delta_1$ s.t. $\|\mathbf{h}\| < \delta_1 \Rightarrow \frac{\|\alpha\|}{\|\mathbf{h}\|} < \frac{\varepsilon}{2}$. Assume $\delta_1 < 1$, so

$$\|\mathbf{h}\| < \delta_1 \Rightarrow \|\alpha(\mathbf{x}_0, \mathbf{h})\| < \frac{\varepsilon}{2}\|\mathbf{h}\| < \frac{\varepsilon}{2}$$

$\|L(\mathbf{h})\| \leq \|L\|_{\text{op}}\|\mathbf{h}\| < \frac{\varepsilon}{2}$ provided $\|\mathbf{h}\| < \delta_2$ where $\delta_2 < \frac{\varepsilon}{2\|L\|_{\text{op}}}$.

Let $\delta = \min\{\delta_1, \delta_2\}$. If $\|\mathbf{h}\| < \delta$ then

$$\begin{aligned} \|f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)\| &\leq \|L(\mathbf{h})\| + \|\alpha(\mathbf{x}_0, \mathbf{h})\| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$ ☐

Proposition 2.2.3. *If $f\colon X \to \mathbb{R}^p$ diff at $\mathbf{x}_0 \in X$ then $\exists$ derivative of f in direction $\mathbf{v}$ at $\mathbf{x}_0$ $\forall\, \mathbf{v} \in \mathbb{R}^n$. Moreover $D_{\mathbf{v}}f(\mathbf{x}_0) = d_{\mathbf{x}_0}f(\mathbf{v})$.*

Proof. $\mathbf{v} \in \mathbb{R}^n$. f diff $\iff$ $f(\mathbf{x}_0+\mathbf{h}) = f(\mathbf{x}_0)+L(\mathbf{h})+\alpha(\mathbf{x}_0, \mathbf{h})$ where $\alpha = o(\mathbf{h})$.

Let $\mathbf{h} = t\mathbf{v}$.

$$\begin{aligned} \frac{f(\mathbf{x}_0 + t\mathbf{v}) - f(\mathbf{x}_0)}{t} &= \frac{L(t\mathbf{v})}{t} + \frac{\alpha(\mathbf{x}_0, t\mathbf{v})}{t} \\ &= L(\mathbf{v}) + \frac{\alpha(\mathbf{x}_0, t\mathbf{v})}{t} \end{aligned}$$

as L linear.

$$\lim_{t \to 0} \frac{\alpha(\mathbf{x}_0, t\mathbf{v})}{t} = \|\mathbf{v}\| \lim_{t \to 0} \frac{\alpha(\mathbf{x}_0, t\mathbf{v})}{\|t\mathbf{v}\|} = 0$$

so $\lim_{t \to 0} \frac{f(\mathbf{x}_0+t\mathbf{v})-f(\mathbf{x}_0)}{t}$ exists, and equals $L(\mathbf{v})$. ☐

Corollary 2.2.4. *If $f\colon X \to \mathbb{R}$ diff at $\mathbf{x}_0 \in X$ then each of partial derivatives exist at $\mathbf{x}_0$. Moreover*

$$d_{\mathbf{x}_0} f(\mathbf{h}) = \langle \nabla f(\mathbf{x}_0), \mathbf{h} \rangle$$

Corollary 2.2.5. *If $f\colon X \to \mathbb{R}^p$ diff at $\mathbf{x}_0 \in X$ then*

$$d_{\mathbf{x}_0} f(\mathbf{h}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}_0) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{x}_0) & \dots & \frac{\partial f_p}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix}$$

Definition 2.2.6.

$$\left(\frac{\partial f_i}{\partial x_j}\right)_{i=1,\dots,p;\ j=1,\dots,n}$$

called *Jacobian matrix* of f.

2.3 Sufficient condition for differentiability

Example 2.3.1 (Function has all partial derivatives on $\mathbb{R}^2$ but not diff at $\mathbf{0}$)**.**

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

$\frac{\partial f}{\partial x_1} = \frac{y}{x^2+y^2} - \frac{2x^2y}{(x^2+y^2)^2}$.

$\frac{\partial f}{\partial x_1}(x,0) = 0 \quad \forall\, x$. Similarly for $\frac{\partial f}{\partial x_2}$, so $\exists$ partial derivative $\forall\, \mathbf{x} \in \mathbb{R}^2$.

Diff $\Rightarrow$ cts, but f not cts at $\mathbf{0}$.

Proposition 2.3.2. *Let $f\colon X \to \mathbb{R}$. If all 1st order partial derivatives of f exist and are cts in a neighbourhood of $\mathbf{x} \in X$, then f is diff at $\mathbf{x}$.*

Proof. $\exists\, \varepsilon > 0$ s.t. $\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}$ cts in $B(\mathbf{x}, \varepsilon)$. Take any $\mathbf{h} \in \mathbb{R}^n$, $\|\mathbf{h}\| < \varepsilon$.

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) &= f(x_1+h_1, x_2+h_2, \dots, x_n+h_n) - f(x_1, \dots, x_n) \\ &= f(x_1+h_1, x_2+h_2, \dots, x_n+h_n) - f(x_1, x_2+h_2, \dots, x_n+h_n) \\ &+ f(x_1, x_2+h_2, \dots, x_n+h_n) - f(x_1, x_2, x_3+h_3, \dots, x_n+h_n) \\ &+ f(x_1, x_2, x_3+h_3, \dots, x_n+h_n) - \dots \\ &+ f(x_1, x_2, \dots, x_{n-1}, x_n+h_n) - f(x_1, \dots, x_n) \end{aligned}$$

Apply MVT to each line above

$$\begin{aligned} &= \frac{\partial f}{\partial x_1}(\xi_1, x_2+h_2, \dots, x_n+h_n)h_1 \\ &+ \frac{\partial f}{\partial x_2}(x_1, \xi_2, x_3+h_3, \dots, x_n+h_n)h_2 \\ &+ \dots \\ &+ \frac{\partial f}{\partial x_n}(x_1, x_2, \dots, x_{n-1}, \xi_n)h_n \end{aligned}$$

for $\xi_i \in [x_i, x_i + h_i]$.

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle + \sum_{i=1}^{n} \alpha_i(\mathbf{x}, \mathbf{h})$$

where $\alpha_i(\mathbf{x}, \mathbf{h}) = \left(\frac{\partial f}{\partial x_i}(x_1, \dots, x_{i-1}, \xi_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n) - \frac{\partial f}{\partial x_i}(\mathbf{x}) \right) h_i$.

$$\begin{aligned} \frac{|\alpha_i(\mathbf{x}, \mathbf{h})|}{\|\mathbf{h}\|} &= \frac{\left| \frac{\partial f}{\partial x_i}(x_1, \dots, x_{i-1}, \xi_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n) - \frac{\partial f}{\partial x_i}(\mathbf{x}) \right| |h_i|}{\|\mathbf{h}\|} \\ &\le \left| \frac{\partial f}{\partial x_i}(x_1, \dots, x_{i-1}, \xi_i, x_{i+1} + h_{i+1}, \dots, x_n + h_n) - \frac{\partial f}{\partial x_i}(\mathbf{x}) \right| \quad [\text{as } \frac{|h_i|}{\|\mathbf{h}\|} \le 1] \\ &\xrightarrow[h \to 0]{} 0 \end{aligned}$$

as $\frac{\partial f}{\partial x_i}$ cts, so $\alpha_i(\mathbf{x}, \mathbf{h}) = o(\mathbf{h}) \quad \forall\, i$.

Let $d_{\mathbf{x}} f(\mathbf{h}) = \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle$ (linear). ☐

Corollary 2.3.3. *$f \colon X \to \mathbb{R}^p$. If all partial derivatives $\frac{\partial f_i}{\partial x_j}$ $i = 1, \dots, p;\ j = 1, \dots, n$ exist and are cts on X then*

1. *f diff at each $\mathbf{x} \in X$.*
2. *The map $X \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^p)$ $x \mapsto d_{\mathbf{x}} f$ is cts.*

2.4 Properties of the differential

2.4.1 Basic properties

Proposition 2.4.1 (Linearity)**.** *$f, g \colon X \to \mathbb{R}^p$, $a, b \in \mathbb{R}$. If f, g diff at $\mathbf{x} \in X$, then $af + bg$ also diff at $\mathbf{x}$ and*

$$\boxed{d_{\mathbf{x}}(af + bg) = a d_{\mathbf{x}} f + b d_{\mathbf{x}} g}$$

Proof. X open $\Rightarrow\ \exists\, \varepsilon > 0$ s.t. $B(\mathbf{x}, \varepsilon) \subset X$. Take $\mathbf{h} \in \mathbb{R}^n$ s.t. $\|\mathbf{h}\| < \varepsilon$.

$$\begin{aligned} af(\mathbf{x}+\mathbf{h}) + bg(\mathbf{x}+\mathbf{h}) &= a\left(f(\mathbf{x}) + d_{\mathbf{x}} f(\mathbf{h}) + \alpha_f(\mathbf{x}, \mathbf{h})\right) + b\left(g(\mathbf{x}) + d_{\mathbf{x}} g(\mathbf{h}) + \alpha_g(\mathbf{x}, \mathbf{h})\right) \\ &= \underbrace{af(\mathbf{x}) + bg(\mathbf{x})}_{\text{const}} + \underbrace{a d_{\mathbf{x}} f(\mathbf{h}) + b d_{\mathbf{x}} g(\mathbf{h})}_{\text{linear in } \mathbf{h}} + \underbrace{a\alpha_f + b\alpha_g}_{= o(\mathbf{h})} \end{aligned}$$

☐

Proposition 2.4.2 (Product rule)**.** *If $f, g \colon X \to \mathbb{R}^p$ diff at $\mathbf{x} \in X$ then fg diff at $\mathbf{x}$ and*

$$\boxed{d_{\mathbf{x}}(fg) = f(\mathbf{x}) d_{\mathbf{x}} g + g(\mathbf{x}) d_{\mathbf{x}} f}$$

Proposition 2.4.3 (Quotient rule)**.** *If $f, g\colon X \to \mathbb{R}$ (not $\mathbb{R}^p$ as no division here) diff at $\mathbf{x} \in X$ and $g(\mathbf{x}) \neq 0$ then $\frac{f}{g}$ diff at $\mathbf{x}$ with*

$$\boxed{d_{\mathbf{x}}\left(\frac{f}{g}\right) = \frac{1}{g^2(\mathbf{x})}\left(g(\mathbf{x})d_{\mathbf{x}}f - f(\mathbf{x})d_{\mathbf{x}}g\right)}$$

Theorem 2.4.4 (Chain rule)**.** *$X \subset \mathbb{R}^n$ open, $Y \subset \mathbb{R}^p$ open. If $f\colon X \to \mathbb{R}^p$ diff at $\mathbf{x} \in X$, $g\colon Y \to \mathbb{R}^q$ diff at $\mathbf{y} = f(\mathbf{x})$ then $g \circ f$ diff at $\mathbf{x} \in X$ and*

$$\boxed{d_{\mathbf{x}}(g \circ f) = d_{f(\mathbf{x})}g \circ d_{\mathbf{x}}f}$$

Proof. X, Y open so $\exists\, \delta > 0$ s.t. $B(\mathbf{x}, \delta) \subset X$. f cts so $f\left((\mathbf{x}, \delta)\right) \subset Y$.

Take $\mathbf{h} \in \mathbb{R}^n$, $\|\mathbf{h}\| < \delta$. Let $\mathbf{y} = f(\mathbf{x})$, $\Delta_{\mathbf{y}} = f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})$.

$$\begin{aligned} g(f(\mathbf{x}+\mathbf{h})) - g(f(\mathbf{x})) &= g(\mathbf{y} + \Delta_{\mathbf{y}}) - g(\mathbf{y}) \\ &\underset{g \text{ diff}}{=} d_{\mathbf{y}}g(\Delta_{\mathbf{y}}) + \alpha_g(\mathbf{y}, \Delta_{\mathbf{y}}) \qquad (1) \end{aligned}$$

$\alpha_g = o(\Delta_{\mathbf{y}})$.

$$\Delta_{\mathbf{y}} \underset{f \text{ diff}}{=} d_{\mathbf{x}}f(\mathbf{h}) + \alpha_f(\mathbf{x}, \mathbf{h}) \qquad (2)$$

$\alpha_f = o(\mathbf{h})$.

Substitute (2) into (1):

$$g(f(\mathbf{x}+\mathbf{h})) - g(f(\mathbf{x})) = (d_{\mathbf{y}}g \circ d_{\mathbf{x}}f)(\mathbf{h}) + d_{\mathbf{y}}g(\alpha_f(\mathbf{x}, \mathbf{h})) + \alpha_g(\mathbf{y}, \Delta_{\mathbf{y}})$$

Let $d_{\mathbf{y}}g(\alpha_f(\mathbf{x}, \mathbf{h})) + \alpha_g(\mathbf{y}, \Delta_{\mathbf{y}}) =: \alpha_{g\circ f}(\mathbf{x}, \mathbf{h})$.

RTP $\alpha_{g\circ f} = o(\mathbf{h})$

$$\frac{\|d_{\mathbf{y}}g(\alpha_f(\mathbf{x}, \mathbf{h}))\|}{\|\mathbf{h}\|} \leq \underbrace{\|d_{\mathbf{y}}g\|_{\text{op}}}_{\text{const}} \underbrace{\frac{\|\alpha_f(\mathbf{x}, \mathbf{h})\|}{\|\mathbf{h}\|}}_{\xrightarrow[\mathbf{h}\to 0]{} 0}$$

so $d_{\mathbf{y}}g(\alpha_f(\mathbf{x}, \mathbf{h})) = o(\mathbf{h})$.

Now to show $\alpha_g(\mathbf{y}, \Delta_{\mathbf{y}}) = o(\mathbf{h})$.

$$\begin{aligned} \|\Delta_{\mathbf{y}}\| &= \|f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x})\| \\ &= \|d_{\mathbf{x}}f(\mathbf{h}) + \alpha_f(\mathbf{x}, \mathbf{h})\| \\ &\leq \|d_{\mathbf{x}}f(\mathbf{h})\| + \|\alpha_f(\mathbf{x}, \mathbf{h})\| \\ &\leq \|d_{\mathbf{x}}f\|_{\text{op}}\|\mathbf{h}\| + \|\alpha_f(\mathbf{x}, \mathbf{h})\| \end{aligned}$$

$\frac{\alpha_f(\mathbf{x}, \mathbf{h})}{\|\mathbf{h}\|} \xrightarrow[\mathbf{h}\to 0]{} 0$ so $\exists\, \varepsilon > 0$ s.t.

$$\frac{\|\alpha_f(\mathbf{x}, \mathbf{h})\|}{\|\mathbf{h}\|} < 1 \quad \forall\, \|\mathbf{h}\| < \varepsilon$$

For these $\mathbf{h}$,

$$\|\Delta_{\mathbf{y}}\| \leq \underbrace{(1 + d_{\mathbf{x}}f\|_{\text{op}})}_{:=c} \|\mathbf{h}\|$$

We have then that $\alpha_g(\mathbf{y}, \Delta_{\mathbf{y}}) = o(\Delta_{\mathbf{y}}) = o(c\mathbf{h}) = o(\mathbf{h})$. ☐

Proposition 2.4.5 (Differential of inverse function)**.** *$X, Y \subseteq \mathbb{R}^n$ open. $f\colon X \to Y$ invertible. If f diff at $\mathbf{x} \in X$ and f^{-1} diff at $\mathbf{y} = f(\mathbf{x}) \in Y$ then*

$$\boxed{d_{\mathbf{y}} f^{-1} = (d_{\mathbf{x}} f)^{-1}}$$

Proof. $f^{-1} \circ f = \iota$ linear, so $d_{\mathbf{x}}\iota = \iota$, so $d_{\mathbf{x}}(f^{-1} \circ f) = \iota$.

Chain rule $\Rightarrow d_{\mathbf{y}} f^{-1} \circ d_{\mathbf{x}} f = \iota \Rightarrow d_{\mathbf{x}} f$ invertible linear map, inverse $d_{\mathbf{y}} f^{-1}$. □

2.4.2 Notation

$f \in C^1(X, \mathbb{R})$ means f diff on X, takes scalar values, $\frac{\partial f}{\partial x_i}$ all cts.

$\forall\, \mathbf{x}, \mathbf{h} \in \mathbb{R}^n$, $[\mathbf{x}, \mathbf{x} + \mathbf{h}] = \{\mathbf{x} + t\mathbf{h} : t \in [0, 1]\}$.

Theorem 2.4.6 (MVT)**.** *$X \subseteq \mathbb{R}^n$ open, $[\mathbf{x}, \mathbf{x} + \mathbf{h}] \subset X$. If $f \in C^1(X, \mathbb{R})$ then $\exists\, \boldsymbol{\xi} \in [\mathbf{x}, \mathbf{x} + \mathbf{h}]$ s.t.*

$$\boxed{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = d_{\boldsymbol{\xi}} f(\mathbf{h})}$$

Corollary 2.4.7. *$f \in C^1(X, \mathbb{R})$ and $d_{\mathbf{x}} f = 0 \quad \forall \mathbf{x} \in X$ then f locally constant.*

Proof. $\forall \mathbf{x} \in X \quad \exists \varepsilon > 0$ s.t.$B(\mathbf{x}, \varepsilon) \subset X$. Then $\forall \|\mathbf{h}\| < \varepsilon \quad [\mathbf{x}, \mathbf{x} + \mathbf{h}] \subset B(\mathbf{x}, \varepsilon)$.

MVT (theorem 2.4.6) $\implies f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = d_{\boldsymbol{\xi}} f(\mathbf{h}) = 0$ so f constant on $B(\mathbf{x}, \varepsilon)$. □

Remark 2.4.8. f not necessarily constant on X, e.g. if X disconnected.

3 Implicit Function Theorem

3.1 Simple version of Implicit Function Theorem

Proposition 3.1.1. *If $F \in C^1(\mathbb{R}^2, \mathbb{R})$ and $F(x_0, y_0) = 0$, $\frac{\partial F}{\partial y}(x_0, y_0) \neq 0$, then $\exists \varepsilon_0, \delta_0 > 0$ and $g \in C^1\left(B(x_0, \delta_0), B(y_0, \varepsilon_0)\right)$ s.t. $\forall x \in B(x_0, \delta_0)$, $\forall y \in B(y_0, \varepsilon_0)$*

$$F(x, y) = 0 \iff y = g(x)$$

Proof. Assume $\frac{\partial F}{\partial y}(x_0, y_0) > 0$. $\frac{\partial F}{\partial y}$ cts $\implies \exists\, \varepsilon_0 > 0$ s.t. $\forall\, x, y \in B((x_0, y_0), \varepsilon_0)$, $\frac{\partial F}{\partial y}(x, y) > 0$. That is, $F(x_0, y)$ is monotone function of y and $F(x_0, y_0) = 0$.

Then $F(x_0, y_0 + \varepsilon_0) > 0$, $F(x_0, y_0 - \varepsilon_0) < 0$. F cts so $\exists\, \delta_1 > 0$ s.t. $|x - x_0| < \delta_1 \implies F(x, y_0 + \varepsilon_0) > 0$, $F(x, y_0 - \varepsilon_0) < 0$.

Let $\delta_0 = \min\{\delta_1, \varepsilon_0\}$. Then $|x - x_0|\delta_0$, $|y - y_0| < \varepsilon_0 \implies$

$$\left.\begin{array}{l} \frac{\partial F}{\partial y}(x, y) > 0 \\ F(x, y_0 + \varepsilon_0) > 0 > F(x, y_0 - \varepsilon_0) \end{array}\right\} \implies$$

$\forall\, x \in B(x_0, \delta_0) \quad \exists!\, y \in B(y_0, \varepsilon_0)$ s.t. $F(x, y) = 0$. Let $g(x) := y$.

Show that g is continuous: Consider $x \in B(x_0, \delta_0)$. Take $\varepsilon > 0$. Let $y = g(x)$ whenever $F(x, y) = 0$.

$\frac{\partial F}{\partial y}(x, y) > 0 \implies F(x, y + \varepsilon) > 0 > F(x, y - \varepsilon)$.

F is cts $\implies \exists\, \delta > 0$ s.t.

$$|\tilde{x} - x| < \delta \implies F(\tilde{x}, y + \varepsilon) > 0 > F(\tilde{x}, y - \varepsilon).$$

Consequently $\exists\, \tilde{y} \in (y - \varepsilon, y + \varepsilon)$ s.t. $F(\tilde{x}, \tilde{y}) = 0$. By uniqueness $\tilde{y} = g(\tilde{x})$, so

$$g(x) - \varepsilon < g(\tilde{x}) < g(x) + \varepsilon \iff |g(\tilde{x}) - g(x)| < \varepsilon \implies$$

g is cts.

Show that g is C^1: $x, \tilde{x} \in B(x_0, \delta_0)$. Let $y = g(x), \tilde{y} = g(\tilde{x})$. $F(x, y) = F(\tilde{x}, \tilde{y}) = 0$. Hence $F(x, y) - F(\tilde{x}, \tilde{y}) = 0$ so by MVT (2.4.6) $\exists \boldsymbol{\xi} \in [(x, y), (\tilde{x}, \tilde{y})] = [(x, g(x)), (\tilde{x}, g(\tilde{x}))]$ s.t.

$$\frac{\partial F}{\partial x}(\boldsymbol{\xi})(\tilde{x} - x) + \frac{\partial F}{\partial y}(\boldsymbol{\xi})(\tilde{y} - y) = 0$$

Hence

$$\begin{aligned} g'(x) &= \lim_{\tilde{x} \to x} \left(\frac{g(\tilde{x}) - g(x)}{\tilde{x} - x} \right) \\ &= - \lim_{\tilde{x} - x} \left(\frac{\frac{\partial f}{\partial x}(\boldsymbol{\xi})}{\frac{\partial f}{\partial y}(\boldsymbol{\xi})} \right) \\ &= - \frac{\frac{\partial f}{\partial x}(x, g(x))}{\frac{\partial f}{\partial y}(x, g(x))} \end{aligned} \tag{3}$$

so g' exists and is cts since (3) is cts. □

3.2 Contraction Mapping Theorem

Definition 3.2.1. X a metric space. $f \colon X \to X$ is *contraction* if $\exists\, c < 1$ s.t.

$$d(f(x), f(y)) < c d(x, y) \quad \forall\, x \neq y$$

Theorem 3.2.2 (Contraction mapping)**.** *If X complete metric space, $f \colon X \to X$ a contraction, then* $\boxed{\exists!\, \eta \in X \text{ s.t. } f(\eta) = \eta}$.

Sketch proof. Show uniqueness: Suppose $\exists\, \eta_1 = f(\eta_1) \neq \eta_2 = f(\eta_2)$

$$d(\eta_1, \eta_2) = d(f(\eta_1), f(\eta_2)) < c d(\eta_1, \eta_2)$$

Contradiction.

Show existence: Take $x_0 \in X$. $x_{n+1} = f(x_n)$. f contraction gives $(x_n)_{n=1}^{\infty}$ Cauchy. f complete $\implies \exists\, \eta = \lim_{n \to \infty} x_n$.

Contractions Lipschitz, so cts, so can swap lim, f.

$$\eta = \lim_{n \to \infty} x_{n+1} = \lim_{n \to \infty} f(x_n) = f\left(\lim_{n \to \infty} x_n \right) = f(\eta)$$ □

3.2.1 Families of contractions

$H\colon S \times Y \to Y$, S, Y metric spaces.

$H(s,y) =: h_s(y)$. For fixed $s \in S$ have $h_s\colon Y \to Y$.

Call s a parameter, h_s a family of maps.

Definition 3.2.3. If $\forall\, s \in S$, $h_s\colon Y \to Y$ is contraction with single contraction constant $0 \le c < 1$

- h_s a *uniform family of contractions.*
- c *uniform contraction constant.*

Corollary 3.2.4. *$B = B(b,r) \subset Y$, $h_s\colon Y \to Y$ uniform family of contractions with unif contraction const $c < 1$.*

- $d(b, h_s(b)) < (1-c)r \quad \forall\, s \in S$.
- $\forall\, y \in B$ $s \mapsto h_s(y)$ *is cts.*

Then

1. $h_s(B) \subset B \quad \forall\, s \in S$
2. $\forall\, s \in S$ $\exists!\, \eta_s = h_s(\eta_s) \in B$
3. *map* $S \to B$, $s \mapsto \eta_s$ *cts.*

Proof. 1. $z \in B(b,r)$.

$$d(h_s(z), h_s(b)) \le c(d(z,b)) \le cr$$

$$d(b, h_s(b)) < (1-c)r \Rightarrow d(h_s(z), b) < cr + (1-c)r = r$$

2. Follows from CMT (theorem 3.2.2).

3. $s, s_0 \in S \implies \eta_s = h_s(\eta_s),\ \eta_{s_0} = h_{s_0}(\eta_{s_0}) \implies$

$$\begin{aligned} d(\eta_{s_0}, \eta_s) &= d(h_s(\eta_s), h_{s_0}(\eta_{s_0})) \\ &\le d(h_s(\eta_s), h_{s_0}(\eta_s)) + d(h_{s_0}(\eta_s) + h_{s_0}(\eta_{s_0})) \\ &\le d(h_s(\eta_s) + h_{s_0}(\eta_s)) + cd(\eta_s, \eta_{s_0}) \\ &\implies \\ d(\eta_s, \eta_{s_0}) &\le \frac{1}{1-c} d(h_{s_0}(\eta_s), \eta_s) \underset{s \to s_0}{\longrightarrow} 0 \end{aligned}$$

$s \mapsto h_s(y)$ cts $\implies h_s(\eta_{s_0}) \to h_{s_0}(\eta_{s_0}) = \eta_{s_0} \implies d(\eta_{s_0}, \eta_s) \underset{s \to s_0}{\longrightarrow} 0$ □

3.3 Lipschitz continuous

Definition 3.3.1. $X \subseteq \mathbb{R}^n$, $f\colon X \to \mathbb{R}^p$ *Lipschitz cts* if $\exists M > 0$ s.t. $\forall \mathbf{x}, \mathbf{y} \in X$

$$\|f(\mathbf{x}) - f(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\|$$

M Lipschitz constant.

Proposition 3.3.2. *$X \subseteq \mathbb{R}^n$ convex, $f \in C^1(X, \mathbb{R}^p)$. If $M = \sup_{\mathbf{x}\in X} \|d_{\mathbf{x}} f\|_{op} < \infty$ then f is Lipschitz with Lipschitz constant M.*

Proof. $g\colon [0,1] \to \mathbb{R}^p$, $g(t) = f((1-t)\mathbf{x} + t\mathbf{y})$. X convex so line belongs to X, so g diff.

$$\begin{aligned} f(\mathbf{y}) - f(\mathbf{x}) &= g(1) - g(0) \\ &= \int_0^1 g'(t)dt \\ &\implies \\ \|f(\mathbf{y}) - f(\mathbf{x})\| &= \left\| \int_0^1 g'(t)dt \right\| \\ &\leq \int_0^1 \|g'(t)\|dt \\ &\leq \sup_{t\in[0,1]} \|g'(t)\| \end{aligned}$$

Chain rule: $g'(t) = d_{(1-t)\mathbf{x}+t\mathbf{y}} f(\mathbf{y} - \mathbf{x}) \implies$

$$\begin{aligned} \|g'(t)\| &= \|d_{(1-t)\mathbf{x}+t\mathbf{y}} f(\mathbf{y} - \mathbf{x})\| \\ &\leq \|d_{(1-t)\mathbf{x}+t\mathbf{y}} f\|_{\text{op}} \|\mathbf{y} - \mathbf{x}\| \\ &\leq M\|\mathbf{y} - \mathbf{x}\| \end{aligned}$$

□

3.4 Implicit Function Theorem

Theorem 3.4.1 (Implicit Function Thm)**.** *$X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{R}^p$ open, $F\colon X \times Y \to \mathbb{R}^p$ is C^1, $\mathbf{a} \in X, \mathbf{b} \in Y$. $\mathbf{c} := F(\mathbf{a}, \mathbf{b})$, $f_{\mathbf{x}}(\mathbf{y}) := F(\mathbf{x}, \mathbf{y})$.*
If $d_{\mathbf{b}} f_{\mathbf{a}}$ is invertible linear map then $\exists\ r_1, r_2 > 0$ and C^1 $g\colon B(\mathbf{a}, r_1) \to B(\mathbf{b}, r_2)$ s.t.

$$F(\mathbf{x}, \mathbf{y}) = \mathbf{c} \iff \mathbf{y} = g(\mathbf{x}) \quad \forall\, \mathbf{x} \in B(\mathbf{a}, r_1),\ \forall\, \mathbf{y} \in B(\mathbf{b}, r_2)$$

3.5 Inverse Function Theorem

Theorem 3.5.1 (Inverse Function Thm)**.** *$X \subseteq \mathbb{R}^n$ open, $f\colon X \to \mathbb{R}^n$, $f \in C^1$, $\mathbf{x}_0 \in X$, $d_{\mathbf{x}_0} f$ invertible linear.*
Then $\exists$ neighbourhood U of $\mathbf{x}_0$ and V of $f(\mathbf{x}_0)$ s.t. $f\colon U \to V$ invertible and $f^{-1}\colon V \to U$ is C^1.

Proof. Let $F(\mathbf{x},\mathbf{y}) = f(\mathbf{y}) - \mathbf{x}, \mathbf{a} = f(\mathbf{x}_0), \mathbf{b} = \mathbf{x}_0$. $f \in C^1 \Rightarrow F \in C^1$. $F(\mathbf{a},\mathbf{b}) = \mathbf{0}$.

Let $f_{\mathbf{x}}(\mathbf{y}) = F(\mathbf{x},\mathbf{y}) = f(\mathbf{y}) - \mathbf{x}$ so $d_{\mathbf{y}} f_{\mathbf{x}} = d_{\mathbf{b}} f$ is invertible.

By ImFT (theorem 3.4.1) $\exists r_1, r_2 > 0, g \in C^1\left(B(\mathbf{a},r_1), B(\mathbf{b},r_2)\right)$ s.t. $\forall\, \mathbf{x} \in B(\mathbf{a},r_1)$, $\forall\, \mathbf{y} \in B(\mathbf{b},r_2)$; $F(\mathbf{x},\mathbf{y}) = \mathbf{0} \iff \mathbf{y} = g(\mathbf{x})$.

$F(\mathbf{x},\mathbf{y}) = \mathbf{0} \iff \mathbf{x} = f(\mathbf{y})$ so $g = f^{-1}$. Let $V = B(f(\mathbf{x}_0), r_1)$, $U = g(V)$. □

Remark 3.5.2. $f \in C^1(X, \mathbb{R}^n)$, $X \subseteq \mathbb{R}^n$ open. If $d_{\mathbf{x}_0} f$ *not* invertible at $\mathbf{x}_0 \in X$ then $\nexists\, C^1 \ni f^{-1}$ near $\mathbf{x}_0$.

Proof. Suppose $\exists g \in C^1, g \circ f(\mathbf{x}) = \mathbf{x}$. Differentiating: $d_{f(\mathbf{x}_0)} g \circ d_{\mathbf{x}} f = \iota \Rightarrow d_{\mathbf{x}_0} f$ invertible. Contradiction. □

Definition 3.5.3. $f\colon X \to Y$ called *diffeomorphism* if

- $f \in C^1$
- $\exists\, f^{-1}\colon Y \to X$
- $f^{-1} \in C^1$

3.6 Corollaries from Implicit and Inverse Function Theorems

Proposition 3.6.1. *$X \subseteq \mathbb{R}^n$ open. Let $f\colon X \to \mathbb{R}$ be C^1 s.t. $\nabla f(\mathbf{a}) \neq 0$ for some $\mathbf{a} \in X$. Let $c = f(\mathbf{a})$. Then $\exists\, U, V \subset \mathbb{R}^n$ open s.t. $\mathbf{a} \in U, \mathbf{0} \in V$ and $\exists$ diffeomorphism $g\colon V \to U$ s.t. $g(\mathbf{0}) = \mathbf{a}$ and $(f \circ g)(\mathbf{y}) = y_1 + c\ \forall\, \mathbf{y} \in V$.*

Proof. $\exists\, k$ s.t. $\frac{\partial f}{\partial x_k}(\mathbf{a}) \neq 0$. W.l.o.g. assume $k = 1$.

Define $\tilde{g}(\mathbf{x}) = (f(\mathbf{x}) - f(\mathbf{a}), x_2 - a_2, \ldots, x_n - a_n)$. $\tilde{g} \in C^1(X, \mathbb{R}^n)$, $\tilde{g}(\mathbf{a}) = \mathbf{0}$.

$$\text{Jacobian}\,(\tilde{g}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

$$\det\left(\text{Jac}\,(\tilde{g}) \Big|_{\mathbf{x}=\mathbf{a}} \right) = \frac{\partial f}{\partial x_1}(\mathbf{a}) \neq 0$$

so Jac $(\tilde{g})$ invertible. Then by InFT (theorem 3.5.1) $\exists\, U, V \subset \mathbb{R}^n$ open s.t. $\mathbf{a} \in U$, $\mathbf{0} \in V$, $\tilde{g}\colon U \to V$ diffeomorphism.

Let $g = \tilde{g}^{-1}\colon V \to U$, $g \in C^1$, $g(\mathbf{0}) = \mathbf{a}$.

Then $(f \circ g)(\mathbf{y}) = f(\mathbf{x}) = y_1 + c$. □

4 Applications

4.1 Derivatives of higher order

Theorem 4.1.1. *If* $f\colon X \to \mathbb{R}$ *(*$X \subseteq \mathbb{R}^n$ *open) has cts partial derivatives* $\frac{\partial^2 f}{\partial x_i \partial x_j}$ *and* $\frac{\partial^2 f}{\partial x_j \partial x_i}$ *for some* $1 \le i, j \le n$ *then*

$$\boxed{\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) \quad \forall\, \mathbf{x} \in X}$$

Proof. IDEA: Define F, φ, find t_1, t_2 by MVT (thm 2.4.6) applied to φ, $\tilde{t}_1, \tilde{t}_2$. Take limits as $h_1, h_2 \to 0$.

Sufficient to consider $n = 2, i = 1, j = 2$. Take $\mathbf{x} \in X$. $\exists\, r > 0$ s.t. $B(\mathbf{x}, r) \subset X$ as X open. If $\|\mathbf{h}\| < r$ then $\mathbf{x} + \mathbf{h} \in X$.

$$F(h_1, h_2) = f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) - f(x_1, x_2 + h_2) + f(x_1, x_2)$$

Define $\varphi(t) = f(x_1 + th_1, x_2 + h_2) - f(x_1 + th_1, x_2)$. Then

$$\begin{aligned} F(h_1, h_2) &= \varphi(1) - \varphi(0) \\ &\underset{\text{MVT}}{=} \varphi'(t_1) \\ &= \frac{\partial f}{\partial x_1}(x_1 + t_1 h_1, x_2 + h_2) h_1 - \frac{\partial f}{\partial x_1}(x_1 + t_1 h_1, x_2) h_1 \\ &\underset{\text{MVT}}{=} \frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1 + t_1 h_1, x_2 + t_2 h_2) h_1 h_2 \end{aligned}$$

Similarly $\exists\, \tilde{t}_1, \tilde{t}_2 \in [0, 1]$ s.t.

$$F(h_1, h_2) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 + \tilde{t}_1 h_1, x_2 + \tilde{t}_2 h_2) h_1 h_2$$

If $h_1, h_2 \neq 0$

$$\frac{\partial^2 f}{\partial x_2 \partial x_1}(x_1 + t_1 h_1, x_2 + t_2 h_2) = \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_1 + \tilde{t}_1 h_1, x_2 + \tilde{t}_2 h_2)$$

These partial derivatives cts so taking limits as $h_1, h_2 \to 0$ gives result. ☐

Definition 4.1.2. $f \in C^k(X, \mathbb{R})$ if $f\colon X \to \mathbb{R}$ has partial derivatives of all orders $\le k$ and these are all cts.

Corollary 4.1.3. *If* $f \in C^k$ *then every partial derivative up to order* k *not depend on order of differentiation.*

4.2 Taylor formula

4.2.1 1D case

Let I open $\subseteq \mathbb{R}$, $\varphi \in C^k(I,\mathbb{R})$. Assume $0 \in I, t \in I$ s.t. $[0,t] \subset I$. Then

$$\varphi(t) = \varphi(0) + \varphi'(0)t + \frac{\varphi''(0)t^2}{2!} + \cdots + \frac{\varphi^{(k-1)}(0)t^{k-1}}{(k-1)!} + r_k(t)$$

Lagrange form of remainder

$\exists\, \tau \in [0,t]$ s.t. $r_k = \frac{\varphi^{(k)}(\tau)t^k}{k!}$ so $r_k(t) = O(t^k)$.

Peano form

$\varphi^{(k)}(t) - \varphi^{(k)}(0) \underset{t\to 0}{\longrightarrow} 0$ [as $\tau \in [0,t]$]. Hence

$$r_k(t) = \frac{\varphi^{(k)}(0)t^k}{k!} + o(t^k)$$

4.2.2 Formula for function of several variables

Theorem 4.2.1. *$x \subseteq \mathbb{R}^n$ open, $f \in C^k(X,\mathbb{R})$, $\mathbf{x} \in X, \mathbf{h} \in \mathbb{R}^n$ s.t. $[\mathbf{x}, \mathbf{x}+\mathbf{h}] \subset X$. Then*

$$\boxed{f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + D_{\mathbf{h}}f(\mathbf{x}) + \frac{1}{2!}D^2_{\mathbf{h}}f(\mathbf{x}) + \cdots + \frac{1}{k!}D^k_{\mathbf{h}}f(\mathbf{x}) + o\left(\|\mathbf{h}\|^k\right)}$$

where

$$D^m_{\mathbf{h}}f(\mathbf{x}) = \sum_{j_m=1}^{n}\sum_{j_{m-1}=1}^{n}\cdots\sum_{j_1=1}^{n} h_{j_m}h_{j_{m-1}}\dots h_{j_1}\frac{\partial^m f}{\partial x_{j_m}\partial x_{j_{m-1}}\dots\partial x_{j_1}}(\mathbf{x})$$

Proof. Consider $\varphi(t) = f(\mathbf{x}+t\mathbf{h})$. $f \in C^k(X,\mathbb{R})$ so $\varphi \in C^k([0,1],\mathbb{R})$.

$$\varphi^{(j)}(t) = D^j_{\mathbf{h}}f\Big|_{\mathbf{x}+t\mathbf{h}} \tag{4}$$

1D formula $\Longrightarrow \exists\, \tau \in [0,1]$ s.t.

$$\varphi(1) = \varphi(0) + \varphi'(0) + \cdots + \frac{\varphi^{(k-1)}(0)}{(k-1)!} + \frac{\varphi^{(k)}(\tau)}{k!} \tag{5}$$

Substitute (5) into (4):

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + D_{\mathbf{h}}f(\mathbf{x}) + \frac{1}{2!}D^2_{\mathbf{h}}f(\mathbf{x}) + \cdots + \frac{1}{(k-1)!}D^{k-1}_{\mathbf{h}}f(\mathbf{x}) + \frac{1}{k!}D^k_{\mathbf{h}}f\Big|_{\mathbf{x}+\tau\mathbf{h}}$$

$$
\begin{aligned}
D^k_{\mathbf{h}} f\Big|_{\mathbf{x}+\tau\mathbf{h}} &= \sum_{j_m=1}^{n} \cdots \sum_{j_1=1}^{n} \underbrace{h_{j_m} \dots h_{j_1}}_{O(\|\mathbf{h}\|^k)} \underbrace{\frac{\partial^m f}{\partial x_{j_m} \partial x_{j_{m-1}} \dots \partial x_{j_1}}(\mathbf{x}+\tau\mathbf{h})}_{=\frac{\partial^m f}{\partial x_{j_m} \partial x_{j_{m-1}} \dots \partial x_{j_1}}(\mathbf{x})+o(\mathbf{h})} \\
&= D^k_{\mathbf{h}} f(\mathbf{x}) + o\left(\|\mathbf{h}\|^k\right)
\end{aligned}
$$
□

4.3 Critical points

Definition 4.3.1. $X \subseteq \mathbb{R}^n$ open, $f\colon X \to \mathbb{R}$ has

- *local minimum* at $\mathbf{a} \in X$ if $\exists$ neighbourhood U of $\mathbf{a}$ s.t. $\mathbf{x} \in U \Rightarrow f(\mathbf{x}) \geq f(\mathbf{a})$.
- *local maximum* at $\mathbf{a} \in X$ if $\exists$ neighbourhood U of $\mathbf{a}$ s.t. $\mathbf{x} \in U \Rightarrow f(\mathbf{x}) \leq f(\mathbf{a})$.

Proposition 4.3.2. *If* $\mathbf{a}$ *local min of* $f \in C^1(X, \mathbb{R})$ *then*

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{a}) = 0$$

Proof. IDEA: CONSIDER $\varphi(x_1) := f(x_1, a_2, \dots, a_n)$

φ has local min at $a_1 \Longrightarrow \varphi'(a_1) = 0$. $\varphi'(x_1) = \frac{\partial f}{\partial x_1}(x_1, a_2, \dots, a_n)$ so $\frac{\partial f}{\partial x_1}(\mathbf{a}) = 0$. Do the same $\forall\, i$. □

Corollary 4.3.3. *If* $\mathbf{a}$ *local max of* $f \in C^1(X, \mathbb{R})$ *then*

$$\frac{\partial f}{\partial x_1}(\mathbf{a}) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{a}) = 0$$

Proof. $\mathbf{a}$ is local min of $-f$. □

Corollary 4.3.4. *If* $\mathbf{a}$ *local min/max of* $f \in C^1(X, \mathbb{R})$ *then* $d_{\mathbf{a}} f = 0$.

Definition 4.3.5. $\mathbf{a}$ is a *stationary point* of $f\colon X \to \mathbb{R}$ ($X \subseteq \mathbb{R}^n$ open) if $d_{\mathbf{a}} f = 0$.

Example 4.3.6 (Stationary $\not\Rightarrow$ min/max)**.** $f\colon \mathbb{R}^n \to \mathbb{R}$, $f(\mathbf{x}) = x_1^3$.

$$\frac{\partial f}{\partial x_1}(\mathbf{0}) = 3x_1^2\Big|_{x_1=0} = 0$$

and

$$\frac{\partial f}{\partial x_2}(\mathbf{x}) = \cdots = \frac{\partial f}{\partial x_n}(\mathbf{x}) = 0 \quad \forall\, \mathbf{x} \in X$$

However $\mathbf{0}$ is not a local min or max of f.

Definition 4.3.7. $\mathbf{a}$ is a *critical point* of $f \in C^1(X, \mathbb{R}^p)$ ($X \subseteq \mathbb{R}^n$ open) if

$$\operatorname{rank}\left(\operatorname{Jac}(f)(\mathbf{a})\right) < \min\{n, p\}$$

Remark 4.3.8. $p = 1, \mathbf{a}$ critical $\Longrightarrow \operatorname{Jac}(f)(\mathbf{a}) = 0 \iff \frac{\partial f}{\partial x_k}(\mathbf{a}) = 0 \quad \forall\, k \Longrightarrow$ $\mathbf{a}$ stationary.

4.3.1 Positive definite quadratic forms

$B \in M_n(\mathbb{R})$ symmetric.

Definition 4.3.9. B is positive definite if $\forall\, \mathbf{h} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$

$$B\mathbf{h} \cdot \mathbf{h} = \sum_{i,j=1}^{n} b_{ij} h_i h_j > 0$$

Remark 4.3.10. TFAE:

- B positive definite
- All eigenvalues > 0
- $\exists\, c > 0$ s.t. $B\mathbf{h} \cdot \mathbf{h} \geq c\|\mathbf{h}\|^2$
- All principal minors of B positive.

Example 4.3.11.

$$B = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

$$\left.\begin{array}{cl} |2| & = 2 > 0 \\ \begin{vmatrix} 2 & 1 \\ 1 & 2 \end{vmatrix} & = 3 > 0 \\ |B| & = 9 > 0 \end{array}\right\} \implies B \text{ positive definite.}$$

Definition 4.3.12 (Hessian)**.** $f \in C^2(X, \mathbb{R})$ ($X \subseteq \mathbb{R}^n$ open).

$$\text{Hess}\,(f) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{i,j=1}^{n}$$

Theorem 4.3.13 (Sufficient condition for min)**.** *If $f \in C^2(X, \mathbb{R})$ ($X \subseteq \mathbb{R}^n$ open) $\mathbf{a} \in X$ critical point and* Hess (f) *positive definite at* $\mathbf{a}$ *then* $\mathbf{a}$ *local min.*

Proof. X open so $\exists\, \delta_0 > 0$ s.t. $\|\mathbf{h}\| < \delta_0 \Rightarrow \mathbf{a} + \mathbf{h} \in X$. Taylor gives

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \underbrace{\sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(\mathbf{a}) h_j}_{=0 \text{ as } \mathbf{a} \text{ critical}} + \frac{1}{2} \underbrace{\sum_{i,j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) h_i h_j}_{\geq c\|\mathbf{h}\|^2 \text{ as Hess pos def}} + \underbrace{r_2(\mathbf{a}, \mathbf{h})}_{=o(\|\mathbf{h}\|^2)}$$

Then $f(\mathbf{a}+\mathbf{h}) \geq f(\mathbf{a}) + \frac{1}{2}c\|\mathbf{h}\|^2 + r_2(\mathbf{a}, \mathbf{h})$.

$r_2 = o\left(\|\mathbf{h}\|^2\right)$ so $\forall\, \varepsilon > 0\ \exists\, \delta > 0$ s.t.

$$\|\mathbf{h}\| < \delta \Rightarrow \frac{|r_2(\mathbf{a}, \mathbf{h})|}{\|\mathbf{h}\|^2} < \varepsilon$$

Take $\varepsilon = \frac{c}{4}$. Then $\forall\, \mathbf{h}$ with $\|\mathbf{h}\| < \min\{\delta, \delta_0\}$

$$f(\mathbf{a}+\mathbf{h}) \geq f(\mathbf{a}) + \frac{c}{4}\|\mathbf{h}\|^2 \geq f(\mathbf{a})$$

so $\mathbf{a}$ is a local min. ☐

Definition 4.3.14. $f \in C^1(X, \mathbb{R})$ ($X \subseteq \mathbb{R}^n$ open). $\mathbf{a}$ is *saddle point* of f if

- $\mathbf{a}$ a critical point
- $\forall\, \varepsilon > 0\ \exists\, \mathbf{x}, \mathbf{y} \in B(\mathbf{a}, \varepsilon)$ s.t. $f(\mathbf{x}) < f(\mathbf{a}) < f(\mathbf{y})$

Proposition 4.3.15. *$U \subseteq \mathbb{R}^2$ open, $f \in C^2(U, \mathbb{R})$. Let $\mathbf{a} \in U$ be critical point. $D := \det(\text{Hess}\,(f)(\mathbf{a}))$. Then*

1. $D > 0$, $\frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) > 0 \Longrightarrow \mathbf{a}$ *local min.*
2. $D > 0$, $\frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) < 0 \Longrightarrow \mathbf{a}$ *local max.*
3. $D < 0 \Longrightarrow \mathbf{a}$ *saddle point.*

Proof.
1. By theorem 4.3.13 $D > 0$, $\frac{\partial^2 f}{\partial x_1^2}(\mathbf{a}) > 0$

   $\Longrightarrow$ all principal minors > 0

   $\Longrightarrow$ Hess $(f)(\mathbf{a})$ positive definite

   $\Longrightarrow \mathbf{a}$ local min.

2. Let $\tilde{f} = -f$. True by 1.

3. $H :=$ Hess $(f)(\mathbf{a})$ symmetric by thm 4.1.1 since $f \in C^2$. Therefore eigenvalues λ_1, λ_2 of H are real. $\det H = D < 0 \Longrightarrow \lambda_1 > 0, \lambda_2 < 0$ say. Let $\mathbf{v}_1, \mathbf{v}_2$ be eigenvectors. Then by Taylor

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= f(\mathbf{a}) + \frac{1}{2}\sum_{i,j=1}^{2} \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{a}) h_i h_j + r_2(\mathbf{a}, \mathbf{h}) \\ &= f(\mathbf{a}) + \frac{1}{2} H\mathbf{h}\cdot\mathbf{h} + r_2(\mathbf{a}, \mathbf{h}) \quad \text{let } \mathbf{h} = t\mathbf{v}_k \\ &= f(\mathbf{a}) + \frac{t^2}{2} H\mathbf{v}_k \cdot \mathbf{v}_k + r_2(\mathbf{a}, t\mathbf{v}_k) \\ &= f(\mathbf{a}) + \frac{t^2}{2} \lambda_k \mathbf{v}_k \cdot \mathbf{v}_k + r_2(\mathbf{a}, t\mathbf{v}_k) \end{aligned}$$

   Let $\mathbf{x} = \mathbf{a} + t\mathbf{v}_2$, $\mathbf{y} = \mathbf{a} + t\mathbf{v}_1$. □

4.4 Conditional maxima and minima

Example 4.4.1. Find minimum of function $f(x, y, z) = \exp(x - y + z)$ on $\mathbb{S}^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$.

Solution. Define $F(x, y, z) = x^2 + y^2 + z^2 - 1$. $\mathbb{S}^2 = \{\mathbf{x} : F(\mathbf{X}) = 0\}$. $\nabla f(\mathbf{x}) = e^{x-y+z}\begin{pmatrix}1\\-1\\1\end{pmatrix}$, $\nabla F(\mathbf{x}) = \begin{pmatrix}2x\\2y\\2z\end{pmatrix}$.

We need ∇f parallel to ∇F.

$$e^{x-y+z}\begin{pmatrix}1\\-1\\1\end{pmatrix} = 2\lambda\begin{pmatrix}x\\y\\z\end{pmatrix} \implies x = -y = z$$

Also $x^2+y^2+z^2=1$ so we have two solutions $\begin{pmatrix}\frac{1}{\sqrt{3}}\\\frac{-1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\end{pmatrix}$ and $\begin{pmatrix}\frac{-1}{\sqrt{3}}\\\frac{1}{\sqrt{3}}\\\frac{-1}{\sqrt{3}}\end{pmatrix}$. $\mathbb{S}^2$ closed and bdd so sequentially compact, so f has max and min. Evaluate f at each solution to find which max and which min.

Definition 4.4.2. $X \subseteq \mathbb{R}^n$ open. $F\colon X \to \mathbb{R}^p$ $(p \le n)$. $\mathbf{c} \in \mathbb{R}^p$ is a *regular value* of F if $\operatorname{rank} d_{\mathbf{x}}F = p\ \forall\, \mathbf{x} \in X$ s.t. $F(\mathbf{x}) = \mathbf{c}$.

Remark 4.4.3. $\operatorname{rank} d_{\mathbf{x}}F = p \iff \nabla F_1(\mathbf{x}), \dots, \nabla F_p(\mathbf{x})$ linearly independent at every $\mathbf{x}$ s.t. $F_i(\mathbf{x}) = c_i \quad \forall\, i$.

Definition 4.4.4. $M \subset \mathbb{R}^n$. $\gamma \in C^1([-1,1], M)$ called *curve* in M.

Definition 4.4.5. Let γ be a curve in M, consider $\mathbf{x}_0 = \gamma(0)$. $\mathbf{v} := \dot\gamma(0)$ called *tangent vector* to M at $\mathbf{x}_0$.

$$T_{\mathbf{x}_0}M = \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \text{ tangent vector to } M \text{ at } \mathbf{x}_0\}$$

Proposition 4.4.6. *$X \subseteq \mathbb{R}^n$ open, $F \in C^1(X, \mathbb{R}^p)$ $(1 \le p < n)$. If $\mathbf{c} \in \mathbb{R}^p$ regular value of F then $\forall\, \mathbf{x}_0 \in M_{\mathbf{c}} = \{\mathbf{x} \in X : F(\mathbf{x}) = \mathbf{c}\}$,*

$$\boxed{T_{\mathbf{x}_0}M_{\mathbf{c}} = \ker d_{\mathbf{x}_0}F}$$

Proof that $T_{\mathbf{x}_0}M_{\mathbf{c}} \subseteq \ker d_{\mathbf{x}_0}F$. Pick $\mathbf{v} \in T_{\mathbf{x}_0}M_{\mathbf{c}}$. Then $\exists\, \gamma \in C^1([-1,1], M_{\mathbf{c}})$ s.t. $\gamma(0) = \mathbf{x}_0$, $\dot\gamma(0) = \mathbf{v}$.

Consider $\varphi(t) = F(\gamma(t))$. $\varphi \in C^1([-1,1], \mathbb{R}^p)$.

$$\forall\, t\colon \gamma(t) \in M_{\mathbf{c}} \implies \varphi(t) = \mathbf{c} \implies$$

$$\begin{aligned} 0 &= \dot\varphi(0) \\ &\underset{2.4.4}{=} d_{\gamma(0)}F(\dot\gamma(0)) \\ &= d_{\mathbf{x}_0}F(\mathbf{v}) \\ &\implies \mathbf{v} \in \ker d_{\mathbf{x}_0}F \end{aligned}$$

□

Proposition 4.4.7. *$X \subseteq \mathbb{R}^n$ open. $f \in C^1(X, \mathbb{R})$, $F \in C^1(X, \mathbb{R}^p)$ $1 \le p < n$. If $\mathbf{c}$ is regular value of F and $\mathbf{x}_0$ is local min/max of f on $M_{\mathbf{c}}$ then*

$$\boxed{\ker d_{\mathbf{x}_0}F \subset \ker d_{\mathbf{x}_0}f}$$

Proof. $\mathbf{v} \in \ker d_{\mathbf{x}_0}F$. $\mathbf{c}$ regular value of F so by 4.4.6, $\mathbf{v} \in T_{\mathbf{x}_0}M_{\mathbf{c}}$. Then $\exists\, \gamma \in C^1([-1,1], M_{\mathbf{c}})$ s.t. $\gamma(0) = \mathbf{x}_0$, $\dot\gamma(0) = \mathbf{v}$.

Consider $\varphi(t) = f(\gamma(t))$. $\varphi \in C^1([-1,1], \mathbb{R})$. $\dot\varphi(0) = d_{\gamma(0)}f(\dot\gamma(0)) = d_{\mathbf{x}_0}f(\mathbf{v})$.

$\mathbf{x}_0$ local min/max of $f \implies 0$ local min/max of $\varphi \implies d_{\mathbf{x}_0}f(\mathbf{v}) = 0$ □

Corollary 4.4.8. *$X \subseteq \mathbb{R}^n$ open, $f \in C^1(X, \mathbb{R})$, $F \in C^1(X, \mathbb{R}^p)$. If $\mathbf{c}$ regular value of F, $\mathbf{x}_0$ local min/max of f on $M_\mathbf{c}$ then $\exists\, \lambda_1, \dots, \lambda_p \in \mathbb{R}$ s.t.*

$$\nabla f(\mathbf{x}_0) = \sum_{j=1}^{p} \lambda_j \nabla F_j(\mathbf{x}_0)$$

Proof. $\mathbf{v} \in \ker d_{\mathbf{x}_0} F$, then

$$\begin{pmatrix} \frac{\partial F_1}{\partial x_1}(\mathbf{x}_0) & \dots & \frac{\partial F_1}{\partial x_n}(\mathbf{x}_0) \\ \vdots & & \vdots \\ \frac{\partial F_p}{\partial x_1}(\mathbf{x}_0) & \dots & \frac{\partial F_p}{\partial x_n}(\mathbf{x}_0) \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \mathbf{0}$$

so $\sum_{j=1}^{n} \frac{\partial F_k}{\partial x_j} v_j = 0 \quad 1 \le k \le p$. Hence $\mathbf{v} \perp \nabla F_k(\mathbf{x}_0)$.

By 4.4.7 $\mathbf{v} \in \ker d_{\mathbf{x}_0} f$ so $\sum_{j=1}^{n} \frac{\partial f}{\partial x_j} v_j = 0$. Hence $\mathbf{v} \perp \nabla f(\mathbf{x}_0)$.

Then $\nabla f(\mathbf{x}_0)$ in linear span of $\nabla F_1(\mathbf{x}_0), \dots, \nabla F_p(\mathbf{x}_0)$. ☐

4.4.1 Lagrange multipliers

$X \subseteq \mathbb{R}^n$ open, $f \in C^1(X, \mathbb{R})$, $F \in C^1(X, \mathbb{R}^p)$, $1 \le p < n$, $\mathbf{0}$ is regular value of F.

Problem:

Find min of f on set $F(\mathbf{x}) = \mathbf{0}$.

Consider $G(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{j=1}^{p} \lambda_j F_j(\mathbf{x})$ for $\mathbf{x} \in X, \boldsymbol{\lambda} \in \mathbb{R}^p$.

Theorem 4.4.9. *If $\mathbf{x}_0 \in X$ solution of problem above then $\exists\, \boldsymbol{\lambda}_0 \in \mathbb{R}^p$ s.t. $(\mathbf{x}_0, \boldsymbol{\lambda}_0)$ critical point of G.*

Proof. $\nabla G = 0 \iff \nabla f = \sum_{j=1}^{n} \lambda_j \nabla F_j$; $F_j(\mathbf{x}) = 0 \quad j = 1, \dots, p$. ☐

4.5 Manifolds in $\mathbb{R}^n$

Definition 4.5.1. $M \subseteq \mathbb{R}^n$ is *k-dimensional manifold* in $\mathbb{R}^n$ if $\forall\, \mathbf{x}_0 \in M$ $\exists$ open $U \subset \mathbb{R}^n$, $\mathbf{x}_0 \in U$ and diffeomorphism $\varphi\colon U \to I^n = (-1,1)^n$ s.t.

$$\varphi(M \cap U) = \{\mathbf{y} \in I^n : y_{k+1} = \dots = y_n = 0\}$$

Remark 4.5.2 ($k = 0$). M is 0-D manifold in $\mathbb{R}^n$ iff $\forall\, \mathbf{x}_0 \in M \quad \exists\, U, \varphi$ s.t.

$$\varphi(M \cap U) = \{\mathbf{y} \in I^n : y_1 = \dots = y_n = 0\} = \{\mathbf{0}\}$$

φ diffeomorphism so $M \cap U = \{\mathbf{x}_0\} \implies M$ discrete set.

Conversely suppose $M \subseteq \mathbb{R}^n$ discrete. Take $U = B(\mathbf{x}_0, \delta) \subset \mathbb{R}^n$ open, where $\delta = \min\{1, \inf\{\|\mathbf{x}_0 - \mathbf{x}\| : \mathbf{x}_0 \neq \mathbf{x} \in M\}\}$.

Then $M \cap U = \{\mathbf{x}_0\}$ and $\varphi\colon U \to I^n$, $\varphi(\mathbf{x}) = \mathbf{x} - \mathbf{x}_0$ is diffeomorphism. $\varphi(M \cap U) = \{\mathbf{0}\}$ so M is 0-D manifold in $\mathbb{R}^n$.

Remark 4.5.3 ($k=n$). If M is n-D manifold in $\mathbb{R}^n$, $\forall \mathbf{x}_0 \in M \quad \exists$ neighbourhood U diffeomorphic to I^n.

M open $\implies M$ n-D manifold in $\mathbb{R}^n$.

4.5.1 1-D manifolds in $\mathbb{R}^n$

If M 1-D manifold then $\varphi(M \cap U) = \{\mathbf{y} \in I^n : y_2 = \cdots = y_n = 0\}$ a line on 1st co-ordinate axis of length 2.

$M \cap U$ is preimage of this line, so 1-D manifold is locally a curve in $\mathbb{R}^n$.

Example 4.5.4 (Manifolds in $\mathbb{R}^2$)**.**

ellipse	✓
parabola	✓
figure 8	✗

Figure 8 not a manifold as $\nexists$ neighbourhood of intersect point which can map cross into straight line.

4.5.2 k-D manifolds in $\mathbb{R}^n$

Example 4.5.5 (Co-ordinate plane in $\mathbb{R}^n$)**.**

$$P := \{\mathbf{x} \in \mathbb{R}^n : x_{k+1} = \cdots = x_n = 0\}$$

is k-D manifold in $\mathbb{R}^n$.

$\forall \mathbf{x}_0 \in P$ define $U = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{x}_0\|_\infty < 1\}$, $\varphi(\mathbf{x}) = \mathbf{x} - \mathbf{x}_0$ diffeomorphism.

Example 4.5.6 (Graph of a function)**.** $f\colon \mathbb{R}^k \to \mathbb{R}^q$. Let $n = k+q$.

$$\operatorname{graph}(f) = \{(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^k \times \mathbb{R}^q : \mathbf{y} = f(\mathbf{x})\}$$

If $f \in C^1$ then graph (f) is k-D manifold in $\mathbb{R}^n$.

Proof. Consider $g\colon \mathbb{R}^k \times \mathbb{R}^q \to \mathbb{R}^k \times \mathbb{R}^q$, $g(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y} - f(\mathbf{x}))$ diffeomorphism. $g(\operatorname{graph}(f)) = P$ co-ordinate plane. Construct $\tilde{U}, \tilde{\varphi}$ as for plane above $\forall \mathbf{x}_0 \in P$.

Then $U = g^{-1}(\tilde{U})$, $\varphi = \tilde{\varphi} \circ g$. ☐

Theorem 4.5.7. *If $F \in C^r(\mathbb{R}^n, \mathbb{R}^p)$, $r \geq 1$, $1 \leq p < n$, $\mathbf{c} \in \mathbb{R}^p$ regular value of F then $M_{\mathbf{c}}$ is $(n-p)$-D manifold in $\mathbb{R}^n$.*

Example 4.5.8. $\mathbb{S}^{n-1} \subset \mathbb{R}^n$ is $(n-1)$-D manifold in $\mathbb{R}^n$.

Consider $f(\mathbf{x}) = x_1^2 + \cdots + x_n^2$. $f \in C^1(\mathbb{R}^n, \mathbb{R})$. $\mathbb{S}^{n-1} = \{\mathbf{x} : f(\mathbf{x}) = 1\}$.

Check 1 is regular value:

$$\mathbf{x} \in \mathbb{S}^{n-1} \Rightarrow \exists\, k\colon \left(x_k \neq 0 \Rightarrow \nabla f\Big|_{\mathbb{S}^{n-1}} \neq 0\right) \Rightarrow 1 \text{ regular value of } f$$

Therefore $\mathbb{S}^{n-1} \subset \mathbb{R}^n$ is $(n-1)$-D manifold in $\mathbb{R}^n$ by thm 4.5.7.

Definition 4.5.9 (Tangent space)**.** M k-D manifold in $\mathbb{R}^n$, $\mathbf{x}_0 \in M$.

$$T_{\mathbf{x}_0}M = \{\mathbf{v} \in \mathbb{R}^n : \exists \gamma \colon [-1,1] \to M,\ \gamma(0) = \mathbf{x}_0\ \dot{\gamma}(0) = \mathbf{v}\}$$

Theorem 4.5.10. *If M k-D manifold in $\mathbb{R}^n$ then $\forall\, \mathbf{x}_0 \in M$, $T_{\mathbf{x}_0}M$ is a vector space with* $\dim(T_{\mathbf{x}_0}M) = k$*.*