

# Chapter 1: Multilinear maps

## Introduction

A *Banach space* is a normed vector space that is Cauchy-complete under the usual metric induced by its norm.

If  $E$  and  $F$  are Banach spaces over  $\mathbb{R}$ . We will denote the norms on  $E$ , and  $F$  by single lines, so

$$|x| = \|x\|_E \quad \text{and} \quad |y| = \|y\|_F \quad \forall x \in E, y \in F$$

$\mathcal{L}(E, F)$  will denote the space of linear maps between  $E$  and  $F$ . In the category of Banach spaces, the space of morphisms are called *toplinear morphisms* - or *CLMs* (*continuous linear maps*); which we will denote by  $L(E, F)$  for toplinear morphisms between  $E$  and  $F$ .

We use  $\|\cdot\|_{L(E, F)}$  or  $\|\cdot\|$  to denote the operator norm, depending on how much emphasis we wish to place on  $L(E, F)$ . Recall,

$$\begin{aligned} \|\varphi\|_{L(E, F)} &= \inf \left\{ A \geq 0, |\varphi(x)| \leq A|x| \forall x \in E \right\} \\ &= \sup \left\{ |\varphi(x)|, x \in E, |x| = 1 \right\} \end{aligned}$$

By the open mapping theorem: any continuous surjective linear map is an open map. Hence invertible elements in  $L(E, F)$  are naturally called *toplinear isomorphisms*. If  $\varphi \in L(E, F)$  such that  $\varphi$  preserves the norm between the Banach Spaces, that is for every  $x \in E$ ,  $|x| = |\varphi(x)|$  then we call  $\varphi$  an *isometry*, or a *Banach space isomorphism*. If  $E_1$  and  $E_2$  are Banach spaces, we will use the usual *product norm*  $(x_1, x_2) \mapsto \max(|x_1|, |x_2|)$ .

## Bilinear maps

### Definition 2.1: Bilinear map

A map  $\varphi : E_1 \times E_2 \rightarrow F$ , where  $F$  is also a Banach space, is said to be *bilinear* if

$$\varphi(x, \cdot) : E_2 \rightarrow F \quad \text{and} \quad \varphi(\cdot, y) : E_1 \rightarrow F$$

are linear for every  $x \in E_1$  and  $y \in E_2$ .

### Proposition 2.1: Continuity of a bilinear map

Let  $E_1, E_2, F$  be Banach spaces, a bilinear map  $m : E_1 \times E_2 \rightarrow F$  is continuous if and only if there exists a  $C \geq 0$ , where

$$|m(x, y)| \leq C|x||y| \tag{1}$$

*Proof.* Suppose such a  $C$  exists, fix a convergent sequence  $(x_n, y_n) \rightarrow (x, y)$  in  $E_1 \times E_2 = E$ . Because the projection maps are continuous, this means  $x_n \rightarrow x$  and  $y_n \rightarrow y$ . Using inspiration from the proof where  $x_n y_n \rightarrow xy$ , where

$$x_n(y_n - y) + (x_n - x)y = x_n y_n - xy \quad x, y, x_n, y_n \in \mathbb{R}$$

Using the inspiration, and replacing multiplication in  $\mathbb{R}$  with the bilinear map  $m$ , we have:

$$\begin{aligned} m(x_n, y_n - y) + m(x_n - x, y) &= m(x_n, y_n) - m(x, y) \\ |m(x_n, y_n) - m(x, y)| &\leq C[|x_n| \cdot |y_n - y| + |x_n - x| \cdot |y|] \rightarrow 0 \end{aligned}$$

Conversely, if  $m$  is continuous, then it is continuous at the origin  $(0,0) = 0$ . There exists a  $\delta$  where  $|(x,y)| \leq \delta$  implies  $|m(x,y)| \leq 1$ . Now, if  $x, y \neq 0$  are elements in  $E$ , we normalize so that  $(x,y)$  has length  $\delta$

$$|(x|x|^{-1}\delta, y|y|^{-1}\delta)| = \delta|(x|x|^{-1}, y|y|^{-1})| = \delta$$

So that  $|m(x|x|^{-1}\delta, y|y|^{-1}\delta)| \leq 1$ , using bilinearity of  $m$ :

$$|m(x, y)| \leq \delta^{-2}|x| \cdot |y|$$

Setting  $\delta^{-2} = C$  finishes the proof (notice if either  $x$  or  $y$  is 0, then  $m$  is trivially 0 and the inequality holds). ■

**Proposition 2.2:**  $L(E_1, E_2; F)$  is isomorphic to  $L(E_1, L(E_2, F))$

For each bilinear map  $\omega \in L(E_1, E_2; F)$ , there exists a unique map  $\varphi_\omega \in L(E_1, L(E_2, F))$  such that  $|\omega| = |\varphi_\omega|$ ; such that for every  $(x, y) \in E_1 \times E_2$ ,  $\omega(x, y) = \varphi(x)(y)$ .

*Proof.* Let  $\varphi_\omega : E_1 \rightarrow L(E_2, F)$  be the unique map such that  $\varphi_\omega(x)(y) = \omega(x, y)$ . Proposition 2.1 shows that  $\varphi_\omega(x)$  is a continuous linear map into  $F$  at each  $x$ , and  $|\varphi_\omega(x)| \leq |\omega||x|$ . This holds for an arbitrary  $x$ , and  $\varphi_\omega(\cdot)$  is clearly linear, hence  $|\varphi_\omega| \leq |\omega|$ . Reversing the roles of  $\omega$  and  $\varphi$  shows proves the other estimate.

The rule as outlined above is linear in  $\omega$ ; and it is not hard to see  $\varphi : L(E_1, E_2; F) \rightarrow L(E_1, L(E_2, F))$  is an injection. By the open mapping theorem, the proposition is proven if  $\varphi$  is a surjection. Fix  $\theta \in L(E_1, L(E_2, F))$ , define a map  $\omega : E_1 \times E_2 \rightarrow F$  such that  $\omega(x, \cdot) = \theta(x)(\cdot)$ . So that  $\omega$  is linear in its second argument. To show  $\omega$  is linear in its first: fix a linear combination  $A = \sum ax$  in  $E_1$ , and  $y \in E_2$ .

$$\omega(A, y) = \theta(\sum ax)(y) = \sum a\theta(x)(y) = \sum a\omega(x, y)$$

Continuity follows from Equation (1), and  $\varphi_\omega = \theta$  as needed. ■

## Notation

We will use the following notation to simplify computations with multilinear maps. Let  $E$  and  $F$  be sets, and  $v_1, \dots, v_k \in E$ .  $f : E \rightarrow F$ .

- Listing individual elements:  $v_{\underline{k}}$  means  $v_1, \dots, v_k$  as separate elements.
- Creating a  $k$ -list:  $(v_{\underline{k}}) = (v_1, \dots, v_k) \in \prod E_{j \leq k}$  if  $v_i \in E_i$  for  $i = \underline{k}$ .
- Double indices:  $(v_{\underline{n_k}}) = (v_{n_{\underline{k}}}) = (v_{n_1}, \dots, v_{n_k})$ , and

$$(v_{\underline{n_k}}) \neq (v_{n_{(1, \dots, k)}})$$

- Closest bracket convention:

$$(v_{(\underline{n_k})}) = (v_{(n_1, \dots, n_k)}) \quad \text{and} \quad (v_{n_{(\underline{k})}}) = (v_{n_{(1, \dots, k)}})$$

- Underlining 0 means it is iterated 0 times:

$$(v_{\underline{0}}, a, b, c) = (a, b, c)$$

- Skipping an index:

$$(v_{\underline{i-1}}, v_{i+\underline{k-i}}) = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_k)$$

for  $i = \underline{k}$ .

- Applying  $f$  to a particular index:

$$(v_{\underline{i-1}}, f(v_i), v_{i+\underline{k-i}}) = (v_1, \dots, v_{i-1}, f(v_i), v_{i+1}, \dots, v_k)$$

Of course, if  $i = 1$ , then the above expression reads  $(f(v_1), v_2, \dots, v_k)$  by the  $\underline{0}$  interpretation.

- In any list using this 'underline' notation, we can find the size of a list by summing over all the underlined terms, and the number of terms with no underline.
- If  $\wedge : E \times E \rightarrow F$  is any associative binary operation,

$$\bigcirc(\underline{v_k}) = v_1 \wedge \dots \wedge v_k$$

### Remark 3.1: Preview of exterior calculus

We can write the formula for the determinant of a  $\mathbb{R}^{k \times k}$  matrix in this notation. Suppose  $a_i \in \mathbb{R}$ , and  $b_i \in \mathbb{R}^{k-1}$  for  $i = \underline{k}$ .

$$M = \begin{bmatrix} a_1 & \cdots & a_k \\ | & & | \\ b_1 & \cdots & b_k \\ | & & | \end{bmatrix}$$

The determinant of  $M$  is a linear combination of determinants of  $k-1$ -sized matrices, given in terms of the columns of  $b$

$$\det(M) = \sum_{i=\underline{k}} (-1)^{i-1} a_i \det(b_{\underline{i-1}}, b_{i+\underline{k-i}})$$

## $k$ -linear maps

### Definition 4.1: $k$ -linear maps

Let  $E_{\underline{k}}, F$  be Banach spaces. A map  $\varphi : \prod E_{\underline{k}}$  is  $k$ -linear if for every  $i = \underline{k}$ ,  $v_i \in E_i$ ,

$$\varphi(\cdot, \cdot, v_i, \cdot, \cdot) : \bigcirc(E_{\underline{i-1}}, E_{i+\underline{k-i}}) \rightarrow F \quad \text{is } (k-1)\text{-linear}$$

The following theorem should give confidence to the notation we have adopted to use.

**Proposition 4.1**

Let  $E_{\underline{k}}$  and  $F$  be Banach spaces, a  $k$ -linear map  $\varphi : \prod E_{\underline{k}} \rightarrow F$  is continuous iff there exists a  $C > 0$ , such that for every  $x_i \in E_i$ ,  $i = \underline{k}$

$$|\varphi(x_{\underline{k}})| \leq C \prod |x_{\underline{k}}|$$

*Proof.* Suppose  $\varphi$  is continuous, then it is continuous at the origin. Picking  $\varepsilon = 1$  induces a  $\delta > 0$  such that for  $|(x_{\underline{k}})| \leq \delta$ ,  $|\varphi(x_{\underline{k}})| \leq 1$ . The usual trick of normalizing an arbitrary vector  $(x_{\underline{k}}) \in \prod E_{\underline{k}}$  does the job:

$$|\varphi(x_{\underline{k}} \cdot |x_{\underline{k}}|^{-1} \cdot \delta)| \leq 1 \implies |\varphi(x_{\underline{k}})| \leq \delta^{-k} \prod |x_{\underline{k}}|$$

Conversely, fix a sequence (indexed by  $n$ , in  $k$  elements in the product space  $\prod E_{\underline{k}}$ ), so

$$(x_{\underline{n}}^k) \rightarrow (x_{\underline{k}}^k) \quad \text{as } n \rightarrow +\infty \quad (2)$$

To proceed any further, we need to prove an important equation that decomposes a difference in  $\varphi$ .

$$\varphi(b^{\underline{k}}) - \varphi(a^{\underline{k}}) = \sum_{i=\underline{k}} \varphi(b^{i-1}, \Delta_i, a^{i+k-i}) \quad (3)$$

where  $(b^{\underline{k}})$  and  $(a^{\underline{k}})$  are elements in  $\prod E_{\underline{k}}$ , and  $\Delta_i = b^i - a^i$  for  $i = \underline{k}$ . The proof is in the following note, which is in more detail than usual - to help the reader ease into the new notation.

**Note 4.1**

We proceed by induction, and eq. (3) follows by setting  $m = k$  in

$$\varphi(a^{\underline{k}}) = \varphi(b^{\underline{m}}, a^{m+k-m}) - \sum_{i=\underline{m}} \varphi(b^{i-1}, \Delta_i, a^{i+k-i}) \quad (4)$$

Base case: set  $m = 1$ , by definition of  $k$ -linearity (def. 4.1) of  $\varphi$ . Since  $a^1 = b^1 - \Delta_1$ ,

$$\varphi(a^{\underline{k}}) = \varphi(b^1 - \Delta_1, a^{1+k-1}) = \varphi(b^1, a^{1+k-1}) - \varphi(\Delta_1, a^{1+k-1})$$

Induction hypothesis: suppose eq. (4) holds for a fixed  $m$ . Since  $a^{m+1} = b^{m+1} - \Delta_{m+1}$ ,

$$\begin{aligned} \varphi(a^{\underline{k}}) &= \varphi(b^{\underline{m}}, a^{m+k-m}) - \sum_{i=\underline{m}} \varphi(b^{i-1}, \Delta_i, a^{i+k-i}) \\ &= \varphi(b^{\underline{m}}, a^{m+1}, a^{(m+1)+k-(m+1)}) - \sum_{i=\underline{m}} \varphi(b^{i-1}, \Delta_i, a^{i+k-i}) \\ &= \varphi(b^{m+1}, a^{(m+1)+k-(m+1)}) - \varphi(b^{m+1}, \Delta_{m+1}, a^{(m+1)+k-(m+1)}) - \sum_{i=\underline{m}} \varphi(b^{i-1}, \Delta_i, a^{i+k-i}) \end{aligned}$$

and this proves eq. (3)

We substitute  $a^i = x^i$ , and  $b^i = x_n^i$  for  $i = \underline{k}$ , and eq. (3) becomes eq. (5)

$$\varphi(x_n^{\underline{k}}) - \varphi(x^{\underline{k}}) = \sum_{i=\underline{k}} \varphi(x_n^{i-1}, x_n^i - x^i, x^{i+k-i}) \quad (5)$$

Then the triangle inequality reads

$$\begin{aligned} \left| \varphi(x_n^{\underline{k}}) - \varphi(x^{\underline{k}}) \right| &\leq \sum_{i=\underline{k}} \left| \varphi(x_n^{i-1}, x_n^i - x^i, x^{i+k-i}) \right| \\ &\leq \sum_{i=\underline{k}} |\varphi| \cdot \bigoplus \left( x_n^{i-1}, \Delta_i, x^{i+k-i} \right) \\ &\leq \sum_{i=\underline{k}} |\varphi| \cdot \left| x_n^i - x^i \right| \bigoplus \left( x_n^{i-1}, x^{i+k-i} \right) \\ &\lesssim_n |\varphi| \sup_{i=\underline{k}} |x_n^i - x^i| \rightarrow 0 \end{aligned}$$

where we identify the product  $\bigoplus(v^{\underline{k}})$  with the product of their norms  $\bigoplus(|v^{\underline{k}}|)$ . ■

**Remark 4.1**

The  $k$ -linear variant of prop. 2.2 holds. We will use but not prove this fact.

**Remark 4.2**

We denote the space of  $k$ -linear maps from  $E$  into  $F$  by  $L(E_{\underline{k}}; F) = L(E^{\underline{k}}, F) = L^k(E, F)$ . *Tensors* on  $E$  are  $k$ -linear maps from the product space of  $E$  into  $\mathbb{R}$ , by replacing  $F$  with  $\mathbb{R}$ .

## Chapter 2: Differentiation

## The derivative

### Definition 1.1: Open sets and neighbourhoods

If  $U$  is an open subset of a topological space  $X$ , we denote this by  $U \subseteq X$ . If  $U$  is a *neighbourhood* of a point  $p \in X$ , we write  $p \in U$ .

We do not require neighbourhoods to be open sets; rather, we say  $U$  is a neighbourhood of  $p$  when the interior of  $U$  contains  $p$ .

### Definition 1.2: Little $o$

A real-valued function in a real variable defined for all  $t$  sufficiently small is said to be  $o(t)$  if  $\lim_{t \rightarrow 0} o(t)/t = 0$ . A map  $\psi : U \rightarrow F$  where  $U \subseteq E$  contains 0 in  $E$ , is said to be  $o(h)$  if  $|\psi(h)|/|h| \rightarrow 0$  as  $h \rightarrow 0$  in  $E$ .

### Definition 1.3: Differentiability

Let  $f : E \rightarrow F$  be a map, replacing  $E$  and  $F$  by their open subsets if necessary. We say  $f$  is *differentiable* at  $x \in E$  when there exists a **continuous linear map on  $E$** :  $\lambda \in L(E, F)$  such that

$$f(x + h) = f(x) + \lambda h + o(h) \quad \text{for sufficiently small } h \quad (6)$$

The role  $o(h)$  plays here is a map from  $U \rightarrow F$ , where  $U$  is some neighbourhood of 0.

### Proposition 1.1: Basic properties of the derivative

If  $f$  is differentiable at  $x$ , then the  $\lambda$  in eq. (6) is unique. We write  $f'(x) = Df(x) = \lambda$  as in ?? . Furthermore, if  $f'(x)$  and  $g'(x)$  exist, then  $(f + g)'(x) = f'(x) + g'(x)$  as linear maps, similar for scalar multiplication.

*Proof.* Suppose  $\lambda_i \in L(E, F)$  are both derivatives of  $f$  at  $x$ . Then,

$$\begin{cases} f(x + h) = f(x) + \lambda_1(h) + o(h) \\ f(x + h) = f(x) + \lambda_2(h) + o(h) \end{cases}$$

And  $(\lambda_1 - \lambda_2)(h) = o(h) = \varphi(h) \cdot |h|$ , where  $\varphi(h) \rightarrow 0$  as  $h \rightarrow 0$ . Using the operator norm, we see that

$$\|\lambda_1 - \lambda_2\|_{L(E, F)} \leq |\varphi(h)| \rightarrow 0$$

This proves uniqueness. Suppose  $f$  and  $g$  are differentiable at  $x$ , denote  $\lambda_f = f'(x)$  (resp.  $g'(x)$ ). The definition of def. 1.3 reads

$$\begin{aligned} f(x + h) + g(x + h) &= (f(x) + g(x)) + (\lambda_f(h) + \lambda_g(h)) + o(h) + o(h) \\ (f + g)(x + h) &= (f + g)(x) + (\lambda_f + \lambda_g)(h) + o(h) \end{aligned} \quad (7)$$

since eq. (7) satisfies eq. (6), the proof is complete. ■



**Proposition 1.2: Chain rule**

Let  $E, F, G$  be Banach spaces. If  $f \in C^1(E, F)$ ,  $g \in C^1(F, G)$ , for every  $x \in E$ ,

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x) \quad (8)$$

*Proof.* Since  $f$  is differentiable at  $x$ ,  $f(x+h) = f(x) + f'(x)(h) + o_1(h)$ , (resp. for  $g$ ,  $o_2(h)$ ). Set  $k(h) = f(x+h) - f(x)$ , and

$$\begin{aligned} g(f(x+h)) &= g(f(x)) + g'(f(x))(k(h)) + o_2(k(h)) \\ &= g(f(x)) + g'(f(x))(f'(x)(h) + o_1(h)) + o_2(k(h)) \\ (g \circ f)(x+h) &= (g \circ f)(x) + g'(f(x)) \circ f'(x)(h) + g'(f(x))(o_1(h)) + o_2(k(h)) \\ (g \circ f)(x+h) &= (g \circ f)(x) + g'(f(x)) \circ f'(x)(h) + o(h) \end{aligned}$$

because  $|A(o_1(h))| \leq |A||o_1(h)|$  for all  $A \in L(E, F)$ ; and  $o(k(h)) = o(h)$  for every continuous  $k : E \rightarrow F$  such that  $k(h) \rightarrow 0$  as  $h \rightarrow 0$ . ■

**Proposition 1.3: Derivatives of CLMs****Proposition 1.4: Product rule in  $k$  variables**

Let  $m : \prod F_{\underline{k}} \rightarrow G$  be a  $k$ -linear map between Banach spaces  $F_{\underline{k}}$  and  $G$ . Suppose  $f_i \in C^1(E, F_i)$  with  $i = \underline{k}$ , writing

$$m(\underline{f}_{\underline{k}})(x) = m(\underline{f}_{\underline{k}}(x)) \quad (9)$$

then  $m(\underline{f}_{\underline{k}})$  is in  $C^1(E, G)$  and for every  $y \in E$ ,

$$Dm(\underline{f}_{\underline{k}})(x)(y) = \sum_{i=\underline{k}} m(\underline{f}_{\underline{i}-1}(x), Df_i(x)(y), \underline{f}_{i+\underline{k}-i}(x)) \quad (10)$$

*Proof.* Let  $x$  be fixed. Equation (10) is proven if we show eq. (11)

$$m(\underline{f}_{\underline{k}})(x+h) = m(\underline{f}_{\underline{k}})(x) + \left( \sum_{i=\underline{k}} m(\underline{f}_{\underline{i}-1}(x), Df_i(x)(h), \underline{f}_{i+\underline{k}-i}(x)) \right) + o(h) \quad (11)$$

and for sufficiently small  $h$  we have

$$f_i(x+h) - f_i(x) = Df_i(x)(h) + o(h^i) \quad (12)$$

We will use the difference formula in eq. (4), with the following substitutions

$$f_i(x+h) = b^i \quad f_i(x) = a^i \quad (13)$$

$$Df_i(x)(h) = c^i \quad o(h^i) = \varepsilon^i \quad (14)$$

$$f_i(x+h) - f_i(x) = c^i + \varepsilon^i \quad \Delta^i = o(h^i) + c^i \quad (15)$$

With these substitutions, the equation we want to prove (eq. (10)) becomes eq. (16)

$$m(b^{\underline{k}}) - m(a^{\underline{k}}) = \left( \sum_{i=\underline{k}} m(a^{i-1}, c^i, a^{i+k-i}) \right) + o(h) \quad (16)$$

Starting from eq. (4),

$$m(b^{\underline{k}}) - m(a^{\underline{k}}) = \sum_{i=\underline{k}} m(b^{i-1}, \Delta^i, a^{i+k-i})$$

We can expand each term, if  $i = \underline{k}$ ,

$$m(b^{i-1}, \Delta^i, a^{i+k-i}) = m(b^{i-1}, c^i, a^{i+k-i}) + m(b^{i-1}, o(h^i), a^{i+k-i}) \quad (17)$$

Let us study the first term in eq. (17), and with  $i$  held fixed, define

$$m_i(z^{i-1}) = m(z^{i-1}, c_i, a^{i+k-i}) \quad (18)$$

Expanding the first term within eq. (17), and because  $m_i$  as defined in eq. (18) is  $i-1$ -linear (because it is a  $k$ -linear map with  $k - (i-1)$  variables held constant); we use eq. (4) again.

$$m_i(b^{i-1}) = \left( \sum_{j=\underline{k}} m_i(b^j, \Delta^j, a^{j+(i-1)-j}) \right) + m_i(a^{i-1}) \quad (19)$$

Unboxing the last term in eq. (19) using the definition of  $m_i$  reads

$$m(b^{i-1}, \Delta^i, a^{i+k-i}) = m(a^{i-1}, c^i, a^{i+k-i}) + \sum_{j=i-1} m_i(b^j, \Delta^j, a^{j+(i-1)-j}) \quad (20)$$

We wish to remove all of the  $b^i$ s. Since  $\Delta^i = c^i + \varepsilon^i$  (eq. (15)), we have

$$\begin{aligned} m(b^{\underline{k}}) - m(a^{\underline{k}}) &= \sum_{i=\underline{k}} m(b^{i-1}, c^i, a^{i+k-i}) + m(b^{i-1}, \varepsilon^i, a^{i+k-i}) \\ &= \left( \sum_{i=\underline{k}} m_i(b^{i-1}) \right) + \sum_{i=\underline{k}} m(b^{i-1}, \varepsilon^i, a^{i+k-i}) \\ &= \left( \sum_{i=\underline{k}} m_i(a^{i-1}) + \sum_{j=i-1} m_i(b^j, \Delta^j, a^{j+(i-1)-j}) \right) + \sum_{i=\underline{k}} m(b^{i-1}, \varepsilon^i, a^{i+k-i}) \\ &= \left( \sum_{i=\underline{k}} m_i(a^{i-1}) \right) + \sum_{\substack{i=\underline{k} \\ j=i-1}} m_i(b^{j-1}, \Delta^j, a^{j+(i-1)-j}) + \sum_{i=\underline{k}} m(b^{i-1}, \varepsilon^i, a^{i+k-i}) \end{aligned} \quad (21)$$

The last term within eq. (21) is  $o(h)$ , since it is a linear combination of  $o(h^i)$ s.

$$\left| \sum_{i=\underline{k}} m(b^{i-1}, \varepsilon^i, a^{i+k-i}) \right| \lesssim_{m,a,b} |o(h)| \quad (22)$$

Each summand in the second last term in eq. (21) is  $o(h)$  as well, as

$$\begin{aligned}
 |m_i(b^{j-1}, \Delta^j, a^{j+(i-1)-j})| &\leq |m_i| \left( \prod (b^{j-1}, \Delta^j, a^{j+(i-1)-j}) \right) \\
 &\leq |m| \cdot \left( \prod (c^i, a^{i+k-i}) \right) \left( \prod (b^{j-1}, \Delta^j, a^{j+(i-1)-j}) \right) \\
 &\lesssim_{m,a,b} \sup_{\substack{i=k \\ j=i-1}} |c^i| \cdot |\Delta^j| \\
 &\lesssim_{m,a,b} \sup_{\substack{i=k \\ j=i-1}} |Df_i(x)(h)| \cdot |f_j(x+h) - f_j(x)| \\
 &\lesssim_{m,a,b} |Df_i(x)||h| \sup_{\substack{i=k \\ j=i-1}} |\Delta^j| \\
 &\lesssim_{m,a,b} |o(h)|
 \end{aligned} \tag{23}$$

for the second last estimate we used  $\Delta^j \rightarrow 0$ . Therefore the second term in eq. (21) is  $o(h)$ , and eq. (11) is proven. Therefore  $m(f_k)$  is differentiable at  $x$ . Continuity of  $Dm(f_k)$  follows from the fact that

$$Dm(f_k)(x) = \sum_{i=k} m(f_{i-1}(x), Df_i(x)(\cdot), f_{i+k-i}(x)) \tag{24}$$

and each of the summands eq. (24) can be broken down as the product of the compositions shown in eqs. (25) and (26)

$$x \mapsto (f_{i-1}(x), f_{i+k-i}(x)) \mapsto m(f_{i-1}(x), \cdot, f_{i+k-i}(x)) \tag{25}$$

$$x \mapsto Df_i(x)(\cdot) \tag{26}$$

which are continuous from  $E$  to  $L(E, F)$ . ■

## Chapter 4: Higher order derivatives

## Introduction

We start with the definition of  $C^p(E, F)$ . Let  $E$  and  $F$  be Banach Spaces, if  $p \geq 1$  is an integer, we define the class  $C^p$  to be the set of maps which are  $p$  times differentiable, and  $D^p f \in C(E, X)$ , where

$$X = L(E, L(E, L(E, \dots F))) \text{ } p \text{ times } \xLeftrightarrow{\mathcal{L}} L(E^p, F)$$

Sometimes we replace  $E$  with an open subset  $U \subseteq E$  if necessary, and we write  $f \in C(U, F)$  if  $D^p \in C(U, X)$ . Note, even if  $f \in C^1(U, F)$ ,  $Df$  is still a map from  $U$  into  $L(E, F)$ .

We will prove two major results in this section.

- The structure of the derivative  $D^p f$ , in particular, if  $f \in C^p(E, F)$ , then  $D^p f(x)$  is a *symmetric multilinear map* in  $p$  arguments.
- Taylor's Theorem

## The second derivative

### Proposition 2.1: Product rule in 2 variables

Let  $E_1$ ,  $E_2$  and  $F$  be Banach spaces, if  $\omega : E_1 \times E_2 \rightarrow F$  is bilinear and continuous, then  $\omega$  is differentiable, and for every  $(x_1, x_2) \in E_1 \times E_2$ ,  $(v_1, v_2) \in E_1 \times E_2$ ,

$$D\omega(x_1, x_2)(v_1, v_2) = \omega(x_1, v_2) + \omega(v_1, x_2)$$

Furthermore,  $D^2\omega(x, y) = D\omega \in L(E^2, F)$ , and  $D^3\omega = 0$ .

*Proof.* By the definition of  $\omega$ , using the familiar interpolation method

$$\omega(x_1 + h_2, x_2 + h_2) = \omega(x_1, x_2) + \omega(x_1, h_2) + \omega(h_1, x_2) + \omega(h_1, h_2)$$

by continuity of  $\omega$ , the last term (which we wish to make  $o(h)$ ):

$$|\omega(h_1, h_2)| \leq \|\omega\| \cdot |(h_1, h_2)|^2$$

so that  $\omega(h_1, h_2) = o(h)$ , and  $D\omega(x_1, x_2)$  exists and is continuous, and is given by the *linear map*  $\omega(x_1, \cdot) + \omega(\cdot, x_2)$ . The rest of the proof follows, if it is not immediately obvious then read the following note.

### Note 2.1

Write  $E = E_1 \times E_2$  for convenience. The linear map  $A = D\omega(x_1, x_2)$  takes arguments  $E$  into  $F$ , consider the projections  $\pi_1$  and  $\pi_2$ , and  $v \in E_1 \times E_2$ , then

$$A(v) = \omega(x_1, \pi_1 v) + \omega(\pi_2 v, x_2)$$

We can view  $A(x) = D\omega(x_1, x_2) \in L(E, F)$ . It is clear that  $A$  is linear in  $x$ , if we fix  $v \in E$ ,

$$A(x + y, v) = \omega(\pi_1(x + y), \pi_2 v) + \omega(\pi_1 v, \pi_2(x + y)) = A(x, v) + A(y, v)$$

and similarly for scalar multiplication. Hence  $DA(x) = A \in L(E, L(E, F))$  and  $D^2 A(x) = D^3 \omega = 0$ .

Our next result is the following, which states that if  $f : U \rightarrow F$  where  $U \overset{\circ}{\subseteq} E$ , and  $Df, DDf = D^2f$  exists and are continuous maps from  $U$  into  $L(E, F)$  and  $L(E, L(E, F))$  respectively, then  $D^2f(x)$  is a *symmetric bilinear map*. The proof is non-trivial, and relies on computing the 'Lie Bracket':

$$D^2f(x)(v, w) - D^2f(x)(w, v)$$

Which we will prove is equal to 0 for every  $x \in U$ , and  $v, w \in E$ .

**Note 2.2: Notation for open subsets**

The symbol ' $\overset{\circ}{\subseteq}$ ' means  $U$  is an open subset of  $E$ .

**Proposition 2.2: Second derivative is symmetric**

Let  $f \in C^2(U, F)$ , where  $U \overset{\circ}{\subseteq} E$  with the possibility that  $U = E$ . For every point  $x \in U$ , the *second derivative*  $D^2f(x)$  is bilinear and symmetric.

*Proof.* Fix  $x \in U \ni B(r) + x \overset{\circ}{\subseteq} U$ . We restrict our attention to vectors  $v, w \in E$  where  $|v|, |w| < r2^{-1}$  for now, so that the

$$\{x, x + w, x + v, x + v + w\} \subseteq U$$

We will denote the following quantity by  $\Delta$

$$\Delta = f(x + w + v) - f(x + w) - f(x + v) + f(x)$$

By rearranging terms, we see that  $\Delta$  can be approximated in two ways:

- Postponing the discussion about the the domain of  $y$ , set  $g(y) = f(y + v) - f(y)$  is  $C^2$ , and

$$\Delta = g(x + w) - g(x) \tag{27}$$

- Again, for  $y$  sufficiently close to  $x$ , define  $h(y) = f(y + w) - f(y)$ , and

$$\Delta = h(x + v) - h(x) \tag{28}$$

- To find the domain for  $y$ , an easy argument using the Triangle inequality gives us  $g, h \in C^2(B(r2^{-1}) + x, F)$ ,
- Leaving the computations of  $h$  as an exercise, we compute  $Dg$ , recall the shift map  $y \mapsto y + v$  commutes with  $D$ , and

$$Dg(y) = D(\tau_{-v}f)(y) - Df(y) = Df(y + v) - Df(y) \tag{29}$$

Using MVT twice, once on Equation (27) (the line segment  $x + tw$ ,  $0 \leq t \leq 1$  is contained in the domain of  $g$ ), and another time on Equation (29) (with  $y = x + tw$  in the integrand). We obtain:

$$\begin{aligned}\Delta &= g(x + w) - g(x) \\ &= \int_0^1 Dg(x + tw) \cdot w dt \\ &= \int_0^1 \int_0^1 D^2 f(x + tw + sv) \cdot v ds dt \cdot w \\ &= \int_0^1 \int_0^1 D^2 f(x + tw + sv) ds dt \cdot v \cdot w\end{aligned}$$

We can rewrite the application of  $v$  then  $w$  by  $\cdot(v, w)$ , and using the approximation  $D^2 f(x + tw + sv) \cdot (v, w) = D^2 f(x) \cdot (v, w) + \delta_1(tw, sv)$ . Integrating over  $s, t$  gives

$$\Delta = D^2 f(x) \cdot (v, w) + \int_0^1 \int_0^1 \delta_1(tw, sv) ds dt$$

### Note 2.3

The error term  $\delta_1$  in the integrand is given by

$$\delta_1(tw, sv) = D^2 f(x + tw + sv)(v, w) - D^2 f(x)(v, w)$$

for  $v, w$  sufficiently small and  $0 \leq s, t \leq 1$ .

A similar argument for  $h$  shows that  $\Delta = D^2 f(x) \cdot (w, v) + \int_0^1 \int_0^1 \delta_2(tw, sv) ds dt$ . Combining the two together, the following holds for all  $v, w$  sufficiently small:

$$D^2 f(x) \cdot (v, w) - D^2 f(x) \cdot (w, v) = \int_0^1 \int_0^1 \delta_1(tw, sv) ds dt - \int_0^1 \int_0^1 \delta_2(tw, sv) ds dt \quad (30)$$

To show the right hand side is 0, we will need the following note.

### Note 2.4

We wish to show the RHS of Equation (30) is 0. We begin by controlling the RHS and show that it is super-bilinear; meaning it shrinks after than the product  $|v||w|$ . Then, we will prove a lemma which will show the only bilinear map that satisfies this property is the 0 map.

- For  $j = 1, 2$ , relabel  $\delta = \delta_j$  for convenience. We can use the  $L^1$  inequality, to obtain the estimate

$$\left| \int_0^1 \int_0^1 \delta(tw, sv) ds dt \right| \leq \int_0^1 \int_0^1 |\delta(tw, sv)| ds dt \quad (31)$$

- $\delta(tw, sv)$  is controlled by  $|D^2 f(x + tw + sv) - D^2 f(x)| |v| |w|$ . Take  $y = tw + sv$ , then  $|y| \leq |tw| + |sv|$ . Hence,

$$|\delta_j| \leq |D^2 f(x + tw + sv) - D^2 f(x)| |v| |w| \quad (32)$$

- Let  $A$  denote the span of  $w, v$  for scalars  $s, t \in [0, 1]$ . In symbols,

$$A = \left\{ tw + sv, s, t \in [0, 1] \right\}$$

$A$  is clearly compact, and the continuity of  $D^2f$  means

$$R(v, w, \delta) = \sup_{y \in A} \left| D^2f(x + y) - D^2f(x) \right| \text{ is finite, and } \lim_{(v, w) \rightarrow 0} R(v, w, \delta) = 0 \quad (33)$$

See the remark after this proof for a generalized version of this 'compact linear combination' argument.

- Relabel  $R(v, w)$  to be the maximum across  $R(v, w, \delta_1)$  and  $R(v, w, \delta_2)$ .
- Combining Equations (31) to (33), we obtain the following bound on Equation (30)

$$\begin{aligned} \left| D^2f(x) \cdot (v, w) - D^2f(x) \cdot (w, v) \right| &\leq \left| \iint \delta_1(tw, sv) ds dt - \iint \delta_2(tw, sv) ds dt \right| \\ &\leq \iint |\delta_1| ds dt + \iint |\delta_2| ds dt \\ &\leq |v||w|R(v, w) \end{aligned} \quad (34)$$

The following Lemma gives a useful criterion to check when a multilinear map is identically 0.

### Lemma 2.1

Let  $E$  be a Banach space, and  $k \geq 1$  be an integer. If  $\lambda \in L(E^k, F)$  and there exists another map  $\theta : E^k \rightarrow F$  (defined perhaps on an open neighbourhood of the origin), such that

$$|\lambda(\underline{u}_k)| \leq |\theta(\underline{u}_k)| \cdot \prod |\underline{u}_k|$$

for all  $(\underline{u}_k)$  sufficiently small. And  $\lim_{(\underline{u}_k) \rightarrow 0} \theta(\underline{u}_k) = 0$ , then,  $\lambda = 0$ .

*Proof.* Fix arbitrary  $(\underline{u}_k) \in E^k$ , for  $s > 0$  sufficiently small, the left hand side of the equation reads

$$|s|^k |\lambda(\underline{u}_k)| \leq |\theta(s\underline{u}_k)| \cdot |s|^k \prod |\underline{u}_k|$$

The rest of the argument is Archimedean: divide by  $|s|^k$  and send  $s \rightarrow 0$  (while paying attention to the term with  $\theta$ ): perhaps after relabelling  $v_s = s\underline{u}_k$  for sufficiently small  $s$ , then  $|\theta(v_s)| \rightarrow 0$  as  $s \rightarrow 0$ . ■

■

### Remark 2.1

Generalization of the "compact linear combination" argument used above. Let  $(t_k) \subseteq \mathbb{C}^k$  or  $\mathbb{R}^k$ , and vectors  $v_k \in E$ . Suppose further  $(t_k) \subseteq A$  is compact in  $\mathbb{C}^k$  or  $\mathbb{R}^k$ . It is clear that if  $y = t_i v^i \in E$ ,



where the summation convention is in effect. Then,

$$|y| \lesssim_A |(v^k)|_{E^k}$$

Now, fix a continuous function  $f \in C(E, F)$ , we can approximate the maximum error over all such  $y$

$$\sup_{y \in B} |f(x + y) - f(x)| < \varepsilon \quad \forall |y| \lesssim_A |(v^k)| < \delta$$

where

$$B = \left\{ \sum t_i v^i, (t_k) \subseteq A, (v^k) \in E^k \right\}$$

### The $p$ -th derivatives

If  $f$  is  $p$  times differentiable, and  $f, Df, D^2f, \dots, D^p f$  are all continuous, then we say  $f \in C^p(E, F)$  (replacing  $E$  with an open subset of  $E$  if necessary). A *symmetric,  $k$ -linear* map between vector spaces  $V, W$  is a map  $A \in \mathcal{L}(V^k, W)$  such that for every  $k$ -permutation  $\theta \in S_k$ ,

$$A(v_k) = A(v_{\theta(k)})$$

#### Note 3.1

- We say a map  $F$  is *between* the spaces  $X$  and  $Y$  if  $F : X \rightarrow Y$ .
- $\mathcal{L}(V^K, W)$  denotes the space of  $k$ -linear maps from  $V$  to  $W$  that are not necessarily continuous.

#### Proposition 3.1

If  $f \in C^p(E, F)$ , then  $D^p f(x)$  is symmetric for every  $x \in E$ . (Replace  $E$  with an open set if necessary).

*Proof.* The main proof proceeds as follows. We will use induction on  $p$ , with  $p = 2$  serving as the base case. Our induction hypothesis is that for every  $f \in C^{p-1}(E, F)$ , for every permutation  $\beta \in S_{p-1}$ , at every point  $x \in E$ , for every possible choice of  $p - 1$  vectors  $(v_2, \dots, v_p) = (v_{1+\underline{p-1}})$ ,

$$D^{p-1} f(x)(v_{1+\underline{p-1}}) = D^{p-1} f(x)(v_{1+\beta(\underline{p-1})})$$

To prove the assertion for  $p$ , it suffices to show  $D^p f(x)(v_p)$  is invariant under transpositions of indices; since the transpositions generate  $S_p$ . Furthermore, the transpositions in  $S_p$  are generated by

- the transposition  $(1, 2, \dots) \mapsto (2, 1, \dots)$  where the omitted indices are held fixed, and
- the transpositions which leave the first index fixed:

$$(1, 1 + \underline{p-1}) \mapsto (1, 1 + \beta(\underline{p-1}))$$

where  $\beta \in S_{p-1}$

so it suffices to prove invariance under those two types of transpositions. Let  $g = D^{p-2}f$ , so  $g \in C^2(E, L(E^{p-2}, F))$ . Because the application of vectors (currying) on a multilinear map  $A \in L(E^p, F)$  is associative, illustrated as follows:

$$(A \cdot v_1) \cdot v_2 = A \cdot (v_1, v_2) = A(v_1, v_2, \cdot) \in L(E^{p-2}, F)$$

Then, let  $\lambda : L(E^{p-2}, F) \rightarrow F$  be the evaluation map at  $(v_3, \dots, v_p) = (v_{2+\underline{p-2}})$ . Using the base case on  $D^{p-2}f = g \in C^2(E, L(E^{p-2}, F))$ ,

$$(D^2g)(x)(v_1, v_2) = (D^2g)(x)(v_2, v_1) \implies \lambda((D^2g)(x)(v_1, v_2)) = \lambda((D^2g)(x)(v_2, v_1))$$

But  $\lambda$  is the map that *applies* the rest of the vectors, and

$$(D^2g)(x)(v_1, v_2) \cdot (v_{2+\underline{p-2}}) = (D^2g)(x)(v_2, v_1) \cdot (v_{2+\underline{p-2}}) \quad (35)$$

Since  $D$  commutes with continuous linear maps (and  $\lambda$  is continuous because  $(v_{2+\underline{p-2}})$  is fixed),

$$\lambda(D^2(D^{p-2}f)) = D(\lambda(D(D^{p-2}f))) = D(D\lambda \circ D^{p-2}f) = D^2(\lambda \circ D^{p-2}f) \quad (36)$$

Substituting Equation (35) for the rightmost hand side of Equation (36) gives the result.

### Note 3.2

There are no magic 'identifications' being made here. To be perfectly clear, for each  $x \in E$ ,  $g(x)$  is an element in  $L(E^{p-2}, F)$ , and  $(D^2g)(x) \in L(E^2, L(E^{p-2}, F))$ . Evaluating  $g$  at a point  $x$  gives a bilinear map that takes values in the Banach space  $L(E^{p-2}, F)$ .

For the second case, beginning from the induction hypothesis. If  $\theta$  is a  $p$ -permutation that leaves the first coordinate unchanged, then there exists a unique  $p-1$ -permutation  $\beta \in S_{p-1}$  such that

$$\begin{aligned} (\theta(\underline{p})) &= (1, \theta(1 + \underline{p-1})) \\ &= (1, 1 + \beta(\underline{p-1})) \end{aligned} \quad (37)$$

Using a similar argument as the first case, set  $g = D^{p-1}f$  and  $\lambda, \lambda' \in L(E^{p-1}, F)$  to be the evaluation maps of  $(v_1, v_{1+\underline{p-1}}) = (v_{\underline{p}})$  and  $(v_1, v_{1+\beta(\underline{p-1})})$  respectively. Rehearsing the same proof as before:

$$\begin{aligned} (D^p f)(x)(v_{\underline{p}}) &= D(\lambda D^{p-1} f)(x)(v_1) && \text{Equation (36)} \\ &= D(\lambda' D^{p-1} f)(x)(v_1) && \text{ind. hyp.} \\ &= (D^p f)(x)(v_{\theta(\underline{p})}) && \text{Equation (36)} \end{aligned}$$

This proves the induction step, and the proof is complete. ■

Before stating and proving Taylor's Theorem, an important remark on the 'postcomposition' of linear maps. Summarized in the following note.

### Note 3.3

Let  $f \in C^p(E, F)$ , and  $\lambda \in L^p(F, G)$ .  $\lambda$  induces a map between  $L(E^p, F)$  and  $L(E^p, G)$  by post-composing any multi-linear map  $A \in L(E^p, F)$  by  $\lambda$ . Denoting this map by  $\lambda_*$ ,

$$\lambda_* : L(E^p, F) \rightarrow L(E^p, G)$$

It is clear  $\lambda_*$  is linear and continuous. And its action on  $A$ , evaluated at  $(v_p) \in E^p$  is given by

$$\lambda_*(A) \in L(E^p, G) \quad (\lambda_*(A))(v_p) = \lambda(A(v_p)) = (\lambda \circ A)(v_p)$$

Now, recall that for  $p = 1$

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

To simplify the notation, we want to 'move' the evaluation  $x$  outside of the brackets, and somehow write  $x \mapsto \lambda[(Df)(x)]$  as one map between  $E$  and  $L(E, G)$ . We further *identify*  $\lambda$  as this map, so that

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

Dropping the  $x$  from the expression, for  $p \geq 2$  *assuming a similar formula holds*, then we write  $[D^p(\lambda \circ f)] = \lambda_* \circ D^p f$ . We make a final identification, of  $\lambda = \lambda_*$  (thereby conflating the two different maps, the first is a map from  $E$  to  $F$ , the second is a map from  $L(E^p, F)$  into  $L(E^p, G)$ ).

### Proposition 3.2

If  $p \geq 2$ ,  $f \in C^p(E, F)$ ,  $\lambda \in L(F, G)$ , then

$$D^p(\lambda \circ f) = \lambda \circ D^p f$$

Where we have identified  $\lambda$  as the same map that acts on  $L(E^p, F)$  to produce another map in  $L(E^p, G)$ , and suppressed the point  $x$ .

*Proof.* Use induction on  $p$ . ■

$$\ln(e^a) = e^{\ln(a)} = a$$