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 continuous linear maps; which we will denote by $L(E,F)$ for toplinear morphisms bewteen 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,

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

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 φ preserves the norm between the Banach Spaces, that is for every $x \in E$, $|x| = |\varphi(x)|$ then we call φ 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 $\varphi : E_1 \times E_2 \to F$, where F is also a Banach space, is said to be *bilinear* if

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

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 $m : E_1 \times E_2 \to F$ 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 $(x_n, y_n) \to (x,y)$ in $E_1 \times E_2 = E$. Because the projection maps are continuous, this means $x_n \to x$ and $y_n \to y$. Using inspiration from the proof where $x_n y_n \to xy$, 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:

$$m(x_n,\, y_n - y) + m(x_n - x,\, y) = m(x_n,\, y_n) - m(x,\, y)$$
$$|m(x_n,\, y_n) - m(x,\, y)| \leq C[|x_n| \cdot |y_n - y| + |x_n - x| \cdot |y|] \to 0$$

Conversely, if m is continuous, then it is continuous at the origin $(0, 0) = 0$. There exists a δ 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 δ

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

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 eithe 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 $\varphi_\omega : E_1 \to L(E_2, F)$ 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 ω and φ shows proves the other estimate.

The rule as outlined above is linear in ω; and it is not hard to see $\varphi : L(E_1, E_2; F) \to L(E_1, L(E_2, F))$ is an injection. By the open mapping theorem, the proposition is proven if φ is a surjection. Fix $\theta \in L(E_1, L(E_2, F))$, define a map $\omega : E_1 \times E_2 \to F$ such that $\omega(x, \cdot) = \theta(x)(\cdot)$. So that ω is linear in its second argument. To show ω 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 $v_1, \ldots, v_k \in E$. $f : E \to F$.

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

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

- Closest bracket convention:

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

- 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 $\wedge : E \times E \to F$ is any associative binary operation,

$$\bigwedge(v_{\underline{k}}) = v_1 \wedge \cdots \wedge v_k$$

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

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

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 $i = \underline{k}$, $v_i \in E_i$,

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

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 $\varphi : \prod E_{\underline{k}} \to F$ is continuous iff there exists a $C > 0$, such that for every $x_i \in E_i$, $i = \underline{k}$

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

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

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

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

$$(x_n^{\underline{k}}) \to (x^{\underline{k}}) \quad \text{as } n \to +\infty \tag{2}$$

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

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

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

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

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

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

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

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

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)

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

Then the triangle inequality reads

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

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

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 $L(E_{\underline{k}}; F) = L(E^k, F) = L^k(E, F)$. *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 $U \mathring{\subseteq} X$. If U is a *neighbourhood* of a point $p \in X$, we write $p \mathring{\in} U$.

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 $\lim_{t \to 0} o(t)/t = 0$. A map $\psi : U \to F$ where $U \mathring{\subseteq} E$ contains 0 in E, is said to be $o(h)$ if $|\psi(h)|/|h| \to 0$ as $h \to 0$ in E.

Definition 1.3: Differentiability

Let $f : E \to F$ 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

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

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

Proposition 1.1: Basic properties of the derivative

If f is differentiable at x, then the λ 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 $\varphi(h) \to 0$ as $h \to 0$. Using the operator norm, we see that

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

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

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

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 $x \in E$,

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x) \tag{8}$$

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

$$g(f(x+h)) = g(f(x)) + g'(f(x))(k(h)) + o_2(k(h)) \tag{9}$$
$$= g(f(x)) + g'(f(x))(f'(x)(h) + o_1(h)) + o_2(k(h)) \tag{10}$$
$$(g \circ f)(x+h) = (g \circ f)(x) + g'(f(x)) \circ f'(x)(h) + g'(f(x))(o_1(h)) + o_2(k(h)) \tag{11}$$

■

Proposition 1.3: Product rule in k variables

Let $m : \prod F_{\underline{k}} \to G$ 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

$$m(f_{\underline{k}})(x) = m(f_{\underline{k}}(x)) \tag{12}$$

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

$$Dm(f_{\underline{k}})(x)(y) = \sum_{i=\underline{k}} m(f_{\underline{i-1}}(x), Df_i(x)(y), f_{i+\underline{k-i}}(x)) \tag{13}$$

Proof. Let x be fixed. Equation (13) is proven if we show eq. (14)

$$m(f_{\underline{k}})(x+h) = m(f_{\underline{k}})(x) + \left(\sum_{i=\underline{k}} m(f_{\underline{i-1}}(x), Df_i(x)(h), f_{i+\underline{k-i}}(x))\right) + o(h) \tag{14}$$

and for sufficiently small h we have

$$f_i(x+h) - f_i(x) = Df_i(x)(h) + o(h^i) \tag{15}$$

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

$$f_i(x+h) = b^i \qquad f_i(x) = a^i \tag{16}$$
$$Df_i(x)(h) = c^i \qquad o(h^i) = \varepsilon^i \tag{17}$$
$$f_i(x+h) - f_i(x) = c^i + \varepsilon^i \qquad \Delta^i = o(h^i) + c^i \tag{18}$$

With these substitutions, the equation we want to prove (eq. (13)) becomes eq. (19)

$$m(b^{\underline{k}}) - m(a^{\underline{k}}) = \left(\sum_{i=\underline{k}} m(a^{\underline{i-1}}, c^i, a^{i+\underline{k-i}})\right) + o(h) \tag{19}$$

Starting from eq. (4),

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

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

$$m(b^{\underline{i-1}}, \Delta^i, a^{i+\underline{k-i}}) = m(b^{\underline{i-1}}, c^i, a^{i+\underline{k-i}}) + m(b^{\underline{i-1}}, o(h^i), a^{i+\underline{k-i}}) \tag{20}$$

Let us study the first term in eq. (20), and with i held fixed, define

$$m_i(z^{\underline{i-1}}) = m(z^{\underline{i-1}}, c_i, a^{i+\underline{k-i}}) \tag{21}$$

Expanding the first term within eq. (20), and because m_i as defined in eq. (21) is $i-1$-linear (because it is a k-linear map with $k-(i-1)$ variables held constant); we use eq. (4) again.

$$m_i(b^{\underline{i-1}}) = \left(\sum_{i=\underline{k}} m_i(b^{\underline{j}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \right) + m_i(a^{\underline{i-1}}) \tag{22}$$

Unboxing the last term in eq. (22) using the definition of m_i reads

$$m(b^{\underline{i-1}}, \Delta^i, a^{i+\underline{k-i}}) = m(a^{\underline{i-1}}, c^i, a^{i+\underline{k-i}}) + \sum_{j=\underline{i-1}} m_i(b^{\underline{j}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \tag{23}$$

We wish to remove all of the b^is. Since $\Delta^i = c^i + \varepsilon^i$ (eq. (18)), we have

$$\begin{aligned}
m(b^{\underline{k}}) - m(a^{\underline{k}}) &= \sum_{i=\underline{k}} m(b^{\underline{i-1}}, c^i, a^{i+\underline{k-i}}) + m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \\
&= \left(\sum_{i=\underline{k}} m_i(b^{\underline{i-1}}) \right) + \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \\
&= \left(\sum_{i=\underline{k}} m_i(a^{\underline{i-1}}) + \sum_{j=\underline{i-1}} m_i(b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \right) + \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \\
&= \left(\sum_{i=\underline{k}} m_i(a^{\underline{i-1}}) \right) + \sum_{\substack{i=\underline{k} \\ j=\underline{i-1}}} m_i(b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) + \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}})
\end{aligned} \tag{24}$$

The last term within eq. (24) is $o(h)$, since it is a linear combination of $o(h^i)$s.

$$\left| \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \right| \lesssim_{m,a,b} |o(h)| \tag{25}$$

Each summand in the second last term in eq. (24) is $o(h)$ as well, as

$$\begin{aligned}
\left| m_i(b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \right| &\leq |m_i| \; \Pi \; (b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \\
&\leq |m| \cdot \left(\Pi (c^i, a^{i+\underline{k-i}}) \right) \left(\Pi (b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \right) \\
&\lesssim_{m,a,b} \sup_{\substack{i=\underline{k} \\ j=\underline{i-1}}} \left| c^i \right| \cdot \left| \Delta^j \right| \\
&\lesssim_{m,a,b} \sup_{\substack{i=\underline{k} \\ j=\underline{i-1}}} |Df_i(x)(h)| \cdot |f_j(x+h) - f_j(x)| \\
&\lesssim_{m,a,b} |Df_i(x)||h| \sup_{\substack{i=\underline{k} \\ j=\underline{i-1}}} \left| \Delta^j \right| \\
&\lesssim_{m,a,b} |o(h)|
\end{aligned} \tag{26}$$

for the second last estimate we used $\Delta^j \to 0$. Therefore the second term in eq. (24) is $o(h)$, and eq. (14) is proven. Therefore $m(f_{\underline{k}})$ is differentiable at x. Continuity of $Dm(f_{\underline{k}})$ follows from the fact that

$$Dm(f_{\underline{k}})(x) = \sum_{i=\underline{k}} m(f_{\underline{i-1}}(x), Df_i(x)(\cdot), f_{i+\underline{k-i}}(x)) \tag{27}$$

and each of the summands eq. (27) can be broken down as the product of the compositions shown in eqs. (28) and (29)

$$x \mapsto (f_{\underline{i-1}}(x), f_{i+\underline{k-i}}(x)) \mapsto m(f_{\underline{i-1}}(x), \cdot, f_{i+\underline{k-i}}(x)) \tag{28}$$

$$x \mapsto Df_i(x)(\cdot) \tag{29}$$

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

$$X = L(E, L(E, L(E, \cdots F))p \text{ times} \overset{\mathcal{L}}{\rightleftharpoons} L(E^p, F)$$

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 $\omega : E_1 \times E_2 \to F$ is bilinear and continuous, then ω 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 ω, 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 ω, 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 π_1 and π_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 $D^2A(x) = D^3\omega = 0$.

■

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

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

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 ' $\mathring{\subseteq}$ ' means U is an open subset of E.

Proposition 2.2: Second derivative is symmetric

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

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

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

We will denote the following quantity by Δ

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

By rearranging terms, we see that Δ 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

$$\Delta = g(x + w) - g(x) \tag{30}$$

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

$$\Delta = h(x + v) - h(x) \tag{31}$$

- 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

$$Dg(y) = D(\tau_{-w} f)(y) - Df(y) = Df(y + w) - Df(y) \tag{32}$$

Using MVT twice, once on Equation (30) (the line segment $x+tw$, $0 \le t \le 1$ is contained in the domain of g), and another time on Equation (32) (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 δ_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 $0 \le s,t \le 1$.

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:

$$D^2 f(x)\cdot(v,w) - D^2 f(x)\cdot(w,v) = \int_0^1\int_0^1 \delta_1(tw,sv) ds\, dt - \int_0^1\int_0^1 \delta_2(tw,sv) ds\, dt \tag{33}$$

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 (33) 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

$$\left|\int_0^1\int_0^1 \delta(tw,sv) ds\, dt\right| \le \int_0^1\int_0^1 |\delta(tw,sv)| ds\, dt \tag{34}$$

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

$$|\delta_j| \le \left|D^2 f(x+tw+sv) - D^2 f(x)\right||v||w| \tag{35}$$

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

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

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

$$R(v, w, \delta) = \sup_{y \in A} \Big| D^2 f(x+y) - D^2 f(x) \Big| \quad \text{is finite,} \quad \text{and} \quad \lim_{(v,w) \to 0} R(v, w, \delta) = 0 \tag{36}$$

  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 (34) to (36), we obtain the following bound on Equation (33)

$$\begin{aligned} \Big| D^2 f(x) \cdot (v, w) - D^2 f(x) \cdot (w, v) \Big| &\le \Big| \iint \delta_1(tw, sv) ds\, dt - \iint \delta_2(tw, sv) ds\, dt \Big| \\ &\le \iint |\delta_1| ds\, dt + \iint |\delta_2| ds\, dt \\ &\le |v||w| R(v, w) \end{aligned} \tag{37}$$

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

$$|\lambda(u_{\underline{k}})| \le |\theta(u_{\underline{k}})| \cdot \prod |u_{\underline{k}}|$$

for all $(u_{\underline{k}})$ sufficiently small. And $\lim_{(u_{\underline{k}}) \to 0} \theta(u_{\underline{k}}) = 0$, then, $\lambda = 0$.

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

$$|s|^k |\lambda(u_{\underline{k}})| \le |\theta(s u_{\underline{k}})| \cdot |s|^k \prod |u_{\underline{k}}|$$

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

■

Remark 2.1

Generalization of the "compact linear combination" argument used above. Let $(t_{\underline{k}}) \subseteq \mathbb{C}^k$ or $\mathbb{R}^k$, and vectors $v_{\underline{k}} \in E$. Suppose further $(t_{\underline{k}}) \subseteq A$ 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,

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

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

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

where

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

The p-th derivatives

If f is p times differentiable, and $f, Df, D^2 f, \ldots, D^p f$ 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 $\theta \in S_{\underline{k}}$,

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

Note 3.1

- We say a map F is *between* the spaces X and Y if $F : X \to Y$.
- $\mathcal{L}(V^K, W)$ 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 $(v_2, \ldots, v_p) = (v_{1+\underline{p-1}})$,

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

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

- the transposition $(1, 2, \ldots) \mapsto (2, 1, \ldots)$ 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 $\lambda : L(E^{p-2}, F) \to F$ be the evaluation map at $(v_3, \dots, v_p) = (v_{2+\underline{p-2}})$. Using the base case on $D^{p-2}f = g \in C^2(E, L(E^{p-2}, F))$,

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

But λ is the map that *applies* the rest of the vectors, and

$$(D^2 g)(x)(v_1, v_2) \cdot (v_{2+\underline{p-2}}) = (D^2 g)(x)(v_2, v_1) \cdot (v_{2+\underline{p-2}}) \tag{38}$$

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

$$\lambda(D^2(D^{p-2}f)) = D(\lambda(D(D^{p-2}f)) = D(D\lambda \circ D^{p-2}f) = D^2(\lambda \circ D^{p-2}f) \tag{39}$$

Substituting Equation (38) for the rightmost hand side of Equation (39) 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 $(D^2 g)(x) \in L(E^2, L(E^{p-2}, F))$. 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 θ 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

$$\begin{aligned}(\ \theta(\underline{p})\ ) &= \big(1, \theta(1 + \underline{p-1})\big) \\ &= \big(1, 1 + \beta(\underline{p-1})\big)\end{aligned} \tag{40}$$

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:

$$\begin{aligned}(D^p f)(x)(v_{\underline{p}}) &= D\big(\lambda D^{p-1}f\big)(x)(v_1) && \text{Equation (39)} \\ &= D\big(\lambda' D^{p-1}f\big)(x)(v_1) && \text{ind. hyp.} \\ &= (D^p f)(x)(v_{\theta(\underline{p})}) && \text{Equation (39)}\end{aligned}$$

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 $\lambda \in L^p(F, G)$. λ 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 λ. Denoting this map by λ_*,

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

It is clear λ_* is linear and continuous. And its action on A, evaluated at $(v_{\underline{p}}) \in E^p$ is given by

$$\lambda_*(A) \in L(E^p, G) \quad \big(\lambda_*(A)\big)(v_{\underline{p}}) = \lambda\big(A(v_{\underline{p}})\big) = (\lambda \circ A)(v_{\underline{p}})$$

Now, recall that for $p = 1$

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

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

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

Dropping the x from the expression, for $p \geq 2$ *assuming a similar formula holds*, then we write $\big[D^p(\lambda \circ f)\big] = \lambda_* \circ D^p f$. 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 λ 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. ■

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