TAYLOR'S THEOREM IN NORMED VECTOR SPACES

KB-MATH

Contents

1. Basic setup

Let U, V be normed vector spaces and suppose that $X \subset U$ is open. Let $\mathcal{L}(U, V)$ denote the normed vector space of bounded linear maps.

Definition 1.1. We say that $f : X \to V$ is differentiably at $x_0 \in X$ if there exists a bounded linear map $D : U \to V$ and $\epsilon : [0, \infty) \to [0, \infty)$ with $\lim_{r \to 0} \epsilon(r) = 0$ such that

$$\|f(x_0 + u) - f(x_0) - Du\| \leq \epsilon(\|u\|)\|u\|$$

for all u in some open neighbhourhood of 0.

Lemma 1.2. Such a D must satisfy

$$Du = \lim_{t \to 0} \frac{f(x_0 + tu) - f(x_0)}{t\|u\|}$$

for each fixed non-zero $u \in U$ and is hence unique.

Proof. For each fixed $u \in U \setminus \{0\}$ and sufficiently small t we have that

$$\|f(x_0 + tu) - f(x_0) - D(tu)\| = \epsilon(t\|u\|)\|tu\|.$$

Now divide both sides by $|t|$ and use the linearity of D to get the estimate

$$\|\frac{f(x_0 + tu) - f(x_0)}{t} - Du\| \le \epsilon(t\|u\|)\|u\|.$$

The result now follows by letting $t \to 0$. $\square$

Given the uniqueness, we can now define $D = Df(x_0) \in \mathcal{L}(U, V)$ to be the derivative of f at x_0.

Lemma 1.3. If $f : X \to V$ is differentiable at x_0 then it is continuous at x_0.

Proof. By triangle inequality we have that

$$\|f(x_0 + u) - f(x_0)\| \le \|f(x_0 + u) - f(x_0) - Df(x_0)u\| + \|Df(x_0)u\|$$

but both terms on the right hand side converge to zero as $u \to 0$ by the differentiability of f and the boundedness of $Df(x_0)$, respectively. $\square$

2. Higher order derivatives

How do we define higher order derivatives in this setting? If $f : X \to V$ is differentiable on X then $Df : X \to \mathcal{L}(U, V)$. Now $\mathcal{L}(U, V)$ is itself a normed vector space (equipped with operator norm) hence the derivative of Df (assuming it exists) is a map

$$D^2 f : X \to \mathcal{L}(U, \mathcal{L}(U, V)).$$

This means that given $u_1, u_2 \in U$ we have that

$$D(Df)(x_0) \in \mathcal{L}(U, \mathcal{L}(U, V))$$

and hence

$$(D(Df)(x_0))u_1 \in \mathcal{L}(U, V)$$

and hence

$$((D(Df)(x_0))u_1)u_2 \in V.$$

We abbreviate this element in V by

$$(D^2 f)(x_0)(u_1, u_2)$$

and hence we get a **multilinear map**

$$(D^2 f)(x_0) : U \times U \to V$$

called the *second derivative* of f.

Let us be more formal for higher derivatives. Let $\mathcal{L}_U(V) = \mathcal{L}(U, V)$ and define recurisively, $\mathcal{L}_U^1(V) = \mathcal{L}(U, V)$ and $\mathcal{L}_U^n(V) = \mathcal{L}(U, \mathcal{L}_U^{n-1}(V))$.

Thus if $\phi \in \mathcal{L}_U^2(V)$ then $\phi(u_1) \in \mathcal{L}(U, V)$ and thus $(\phi(u_1))(u_2) \in V$. We rewrite this as $\phi(u_1)(u_2)$ to avoid lots of parentheses. So for example, if $\phi \in \mathcal{L}_U^n(V)$ then $\phi(u_1)(u_2) \ldots (u_n) \in V$, where this expression is processed from left to right (i.e., $\phi(u_1) \cdots (u_k)$ is an element of $\mathcal{L}_U^{n-k}(V)$ and it takes u_{k+1} as an argument).

Observe that $\mathcal{L}_U^n(V)$ is naturally a normed-vector space with norm being an operator norm.

Definition 2.1. We define $f : X \to V$ *to be n-differentiable at* $x_0 \in X$ and we define the n-th derivative (if it exists) $(D^n f)(x_0) \in \mathcal{L}^n_U(V)$ by the following recursive definition:

- If $n = 1$, then $D^1 f(x_0) = Df(x_0)$ and 1-differentiable at x_0 means differentiable at x_0.
- If $n > 1$, then we say that f is n-differentiable at x_0 if there is an open neighbhourhoud $X' \subset X$ containing x_0 such that f is $n-1$-differentiable for all $x \in X'$ and the map $X' \to \mathcal{L}^{n-1}_U(V)$ given by $x \mapsto D^{n-1}(f)(x)$ is differentiable at x_0. In this case, we define the n-th derivative

$$(D^n f)(x_0) = D(D^{n-1} f)(x_0) \in \mathcal{L}^n_U(V).$$

We now show that $D^n f(x_0)$ can be naturally realized as a multilinear map $U \times \cdots \times U \to V$, if it exists, and it can be evaluated through iterated directional derivatives. To make this precise it is worth formalizing some notions regarding the continuity of multilinear maps.

Definition 2.2. Given normed vector spaces $U_1, \ldots, U_n$ and V, let $\mathcal{T}(U_1, \ldots, U_n \to V)$ denote the vector space of *bounded* multilinear maps $T : U_1 \times \cdots \times U_n \to V$, where bounded means that

$$\|T\| := \sup\{T(u_1, \ldots, u_m) \mid u_i \in U_i \text{ with } \|u_i\| = 1\} < \infty.$$

Moreover, $\|\cdot\|$ is a norm on the space of bounded multilinear maps.

Proposition 2.3. There is a canonical isometric isomorphism

$$\mathcal{L}(U_1, \mathcal{T}(U_2, \ldots, U_n \to V)) \cong \mathcal{T}(U_1, \ldots U_n \to V)$$

given by

$$\phi \mapsto ((u_1, \ldots, u_n) \mapsto \phi(u_1)(u_2, \ldots, u_n)).$$

Proof. That is bijective and linear is clear. Let us show that it is an isometry. Let $\phi : U_1 \to T \in \mathcal{T}(U_2, \ldots, U_n \to V)$ be a bounded linear map. Thus for all unit vectors $u_1 \in U_1, \ldots, u_n \in U_n$ we have that

$$\|\phi(u_1)(u_2, \ldots, u_n)\| \leq \|\phi(u_1)\| \leq \|\phi\|,$$

where the first inequality uses the definition of the norm of a multilinear map while the second uses the definition of a the norm of a linear map. This $(u_1, \ldots, u_2) \mapsto \phi(u_1)(u_2, \ldots, u_n)$ is indeed a bounded multilinear map of norm at most $\|\phi\|$. To show the norm of this multilinear map is equal to $\|\phi\|$, take unit vector $u_1 \in U_1$ such that $\|\phi(u_1)\| \geq (1-\epsilon)\|\phi\|$ and now take unit vectors $u_2, \ldots, u_n$ such that $\|\phi(u_1)(u_2, \ldots, u_n)\| \geq (1-\epsilon)\|\phi(u_1)\|$. Thus $\|\phi(u_1)(u_2, \ldots, u_n)\| \geq (1-\epsilon)^2\|\phi\|$.

$\square$

In other words, the norm $\|T\|$ is equal to the operator norm on $\mathcal{L}(U_1, \mathcal{T}(U_2, \ldots, U_n \to V))$. Letting

$$\mathcal{T}_n(U, V) = \mathcal{T}(U_1 \times \cdots \times U_n \to V),$$

we can apply this recursively to get.

Corollary 2.4. We have an isometric isomorphism

$$\mathcal{T}_n(U, V) \cong \mathcal{L}^n_U(V).$$

In this isomorphism, the element $\phi \in \mathcal{L}^n_U(V)$ corresponds to $T \in \mathcal{T}_n(U, V)$ given by

$$T(u_1, \ldots, u_n) = \phi(u_1)(u_2) \cdots (u_n)$$

and the operator norm $\|\phi\|$ is equal to the norm $\|T\|$ on the space of bounded multilinear maps.

Proof. We prove this by induction on n. For $n = 1$ this map is the identity map and $\mathcal{T}_1(U, V) = \mathcal{L}(U, V) = \mathcal{L}_U^1(V)$ and so this mapping $T \mapsto \phi$ is indeed an isomorphism as it is the identity map. Now we let $\iota_k : \mathcal{L}_U^k(V) \to \mathcal{T}_k(U, V)$ by given by

$$\iota_k(\phi)(u_1, \dots u_n) = \phi(u_1)(u_2) \cdots (u_k).$$

By induction hypothesis, ι_{n-1} is an isometric isomorphism, where $n > 1$. Thus we have an isometric isomoprhism $\mathcal{L}(U, \mathcal{L}_U^{n-1}(V)) \to \mathcal{L}(U, \mathcal{T}_{n-1}(U, V))$ which maps ϕ to $\iota_{n-1} \circ \phi$. But by the proposition above, we have an isometric isomorphism $\mathcal{L}(U, \mathcal{T}_{n-1}(U, V)) \to \mathcal{T}_n(U, V)$ which maps φ to the multilinear map

$$(u_1, \dots, u_n) \mapsto \varphi(u_1)(u_1, \dots, u_n).$$

Composing these two isometric isomorphisms, we have an isometric isomorphism mapping $\phi \in \mathcal{L}(U, \mathcal{L}_U^{n-1}(V))$ to the element $T \in \mathcal{T}_n(U, V)$ given by

$$\begin{aligned} T(u_1, \dots, u_n) &= ((\iota_{n-1} \circ \phi)(u_1))(u_2, \dots, u_n) \\ &= \iota_{n-1}(\phi(u_1))(u_2, \dots, u_n) \\ &= \phi(u_1)(u_2) \dots (u_n) \\ &= (\iota_n(\phi))(u_1, \dots, u_n). \end{aligned}$$

Thus we have shown that the mapping $\phi \mapsto \iota_n(\phi)$ is an isomorphism

$$\mathcal{L}(U, \mathcal{L}_U^{n-1}(V)) \to \mathcal{L}(U, \mathcal{T}_{n-1}(U, V)).$$

But by definition, $\mathcal{L}(U, \mathcal{L}_U^{n-1}(V)) = \mathcal{L}_U^n(V)$. $\square$

From now on, we identify $\mathcal{L}_U^n(V)$ with the space of bounded multilinear maps as in the Corollary above. Hence the nth derivative $D^n f(x_0)$ may be recursively defined as bounded multilinear map as follows:

The first derivative is $D^1 f = Df$ (which makes sense since $\mathcal{T}(U \to V) = \mathcal{L}(U, V)$) and

$$D^{(n+1)} f(x_0)(u_1, \dots, u_{n+1}) = [(D(D^n f))(x_0) u_1](u_2, \dots, u_{n+1})$$

if it exists (and if, of course, $D^n f$ exists on X).

The next proposition shows that the n-the derivative is a multilinear form that gives iterated direction derivatives. We let

$$(\partial_u f)(x_0) = \lim_{t \to 0} \frac{f(x_0 + tu) - f(x_0)}{t\|u\|}$$

assuming that this limit exists.

Proposition 2.5. Suppose that $f : X \to V$ is n-times differentiable to x_0. Then

$$D_n f(x_0)(u_1, \dots, u_n) = (\partial_{u_1} \cdots \partial_{u_n} f)(x_0).$$

More precisely, for $k < n$ the iterated limits $(\partial_{u_1} \cdots \partial_{u_k} f)(x)$ exists for all x in some open subset of X, while for $k = n$ it exists for $x = x_0$ and equals the expression given.

Proof. We prove this by induction on n, the $n=1$ case was proven in Lemma 1.2. Now using the induction hypothesis we get

$$\begin{aligned} D^n f(x_0)(u_1,\dots,u_n) &= \left(D(D^{n-1}f)\right)(x_0)(u_1)(u_2,\dots u_n) \\ &= \left(\lim_{t\to 0}\frac{1}{t}\left(D^{n-1}f(x_0+tu) - D^{n-1}f(x_0)\right)\right)(u_2,\dots,u_n) \\ &= \lim_{t\to 0}\frac{1}{t}\left(D^{n-1}f(x_0+tu_1)(u_2,\dots,u_n) - D^{n-1}f(x_0)(u_2,\dots u_n)\right) \\ &= \lim_{t\to 0}\frac{1}{t}\left((\partial_{u_2}\cdots\partial_{u_n}f)(x_0+tu_1) - (\partial_{u_2}\cdots\partial_{u_n}f)(x_0)\right) \\ &= (\partial_1(\partial_{u_2}\cdots\partial_{u_n}f))(x_0) \end{aligned}$$

where we were allowed to put $(u_2,\dots,u_n)$ inside the limit since the convergence takes place with respect to the operator norm (on the space of bounded $n-1$-multilinear forms), hence pointwise. $\square$

Corollary 2.6. If $f:\mathbb{R}^d\to V$ is twice differentiable then

$$D^2f(x_0)(e_i,e_j) = \partial_i\partial_j f(x_0)$$

where ∂_i is partial differentiation with respect to i-th co-ordinate.

3. Basic properties: Chain rule and produt rule

Lemma 3.1. (Chain Rule) Suppose that U,V,W are normed vector spaces with $X\subset U$ and $Y\subset V$ open sets. Suppose $f:X\to Y$ is differentiable at $x_0\in X$ and $g:Y\to W$ is differentiable at $y_0=f(x_0)\in Y$. Then $g\circ f:X\to W$ is differentiable at x_0 and

$$D(g\circ f)(x_0) = D(g)(y_0)\circ D(f)(x_0).$$

Proof. We have

$$(g\circ f)(x_0+u) = g(f(x_0+u)) = g(f(x_0)+Df(x_0)u+R_1(u))$$

for small enough u, where $\|R_1(u)\|\le \epsilon_1(\|u\|)\|u\|$ for some $\epsilon_1:[0,\infty)\to[0,\infty)$ such that $\lim_{r\to\infty}\epsilon(r)=0$.

Observe that for sufficiently small u we have that $Df(x_0)u$ is sufficiently small (as $Df(x_0)$ is bounded and so $\|Df(x_0)u\|\le\|Df(x_0)\|\|u\|$). Thus using the differentiability of g at $y_0=f(x_0)$, we have

$$\begin{aligned} (g\circ f)(x_0+u) &= g(f(x_0)) + Dg(y_0)(Df(x_0)u+R_1(u)) + R_2(Df(x_0)u+R_1(u)). \\ &= g(f(x_0)) + Dg(y_0)Df(x_0)u + Dg(y_0)R_1(u) + R_2(Df(x_0)u+R_1(u)). \end{aligned}$$

It thus remains to bound the remainder terms divided by $\|u\|$ converge to 0. By definition $\|R_2(v)\|\le \epsilon_2(\|v\|)\|v\|$ for ϵ_2 satisfying the same condition as ϵ_1 above. So

$$\begin{aligned} \|R_2(Df(x_0)u+R_1(u))\| &\le \epsilon_2(\|Df(x_0)u+R_1(u)\|)\|Df(x_0)u+R_1(u)\| \\ &\le \epsilon_2(\|Df(x_0)u+R_1(u)\|)\left(\|Df(x_0)\|\|u\|+\|R_1(u)\|\right) \\ &\le \epsilon_2(\|Df(x_0)u+R_1(u)\|)\left(\|Df(x_0)\|\|u\|+\epsilon_1(\|u\|)\|u\|\right) \\ &= \epsilon_2(\|Df(x_0)u+R_1(u)\|)\left(\|Df(x_0)\|+\epsilon_1(\|u\|)\right)\|u\|. \end{aligned}$$

The term in front of $\|u\|$ converges to 0 as $u\to 0$, as desired. Finally, the other remainder term is

$$\|Df(y_0)R_1(u)\|\le\|Df(y_0)\|\|R_1(u)\|\le\|Df(y_0)\|\epsilon_1(\|u\|)\|u\|,$$

as desired. ☐

Remark 3.2. Observe how this proof required the derivative to be a bounded linear map.

Given normed vector spaces U_1, U_2 we will always equip the direct sum with the max norm $\|(u_1, u_2)\| = \max\{\|u_1\|, \|u_2\|\}$, which induces the product topology. Note that if $T : U_1 \times \cdots \times U_n \to V$ is a bounded multilinear map, then it is continuous with respect to this topology as can be seen form the inequality

$$\|T(u_1, \ldots, u_n)\| \leq \|T\|\|u_1\| \cdots \|u_n\|.$$

Lemma 3.3. (Product Rule for multilinear maps) Suppose that $U_1, \ldots, U_n$ are normed vector spaces and $T : U_1 \times \cdots \times U_n \to V$ is a bounded multilinear map. Then T is differentiable at each $\vec{x} = (x_1, \ldots, x_n) \in U_1 \times \cdots \times U_n$ with

$$D(T)(\vec{x})(u_1, \ldots, u_n) = T(u_1, x_2, \ldots, x_n) + T(x_1, u_2, x_3, \ldots, x_n) + \cdots + T(x_1, \ldots x_{n-1}, u_n).$$

Moreover, T is in fact smooth (infinitely differentiable). Moreover, $D^{n+1}(T) = 0$.

Proof. Let $\vec{u} = (u_1, \ldots, u_n)$. Note that

$$T(\vec{x} + \vec{u}) = T(\vec{x}) + \sum_{i=1}^{n} T(x_1, \ldots, x_{i-1}, u_i, x_{i+1}, \ldots, x_n) + R$$

where

$$R = \sum_{(y_1, \ldots, y_n)} T(y_1, \ldots, y_n)$$

where the sum is taken over all $(y_1, \ldots, y_n)$ where at least two of the y_k are equal to u_k and the others are equal to x_j. Thus each a term is at most

$$\|T\|\|y_1\| \cdots \|y_n\| \leq K(\vec{x})\|u\|^2,$$

where $K(\vec{x})$ is some constant depending only on T and $\vec{x}$. This proves that T is differentiable with the derivative as stated.

It remains to prove that T is smooth. For this purpose, let us say that $S : U_1 \times \cdot \times U_n \to W$ is *r-partially multilinear* if there exists $i_1 < \ldots < i_r$ such that $S(x_1, \ldots, x_n) = \tilde{S}(x_{i_1}, \ldots, x_{i_r})$. In other words, S is multilinear in r of the variables and constant in the others. Thus we see that $DT : U_1 \times \cdots \times U_n \to L(U_1 \times \cdots \times U_n, V)$ is a sum of $n-1$-partially multlinear maps. By the same argument as above, we get that

$$D(S)(\vec{x})(u_1, \ldots u_n) = \tilde{S}(u_{i_1}, x_{i_2}, \ldots, x_{i_n}) + \cdots + \tilde{S}(x_{i_1}, \ldots, u_{i_n}).$$

Letting,

$$S^1_{\vec{x}}(u_1, \ldots, u_n) = \tilde{S}(u_{i_1}, x_{i_2}, \ldots, x_{i_n})$$

we have that the map

$$\vec{x} \mapsto S^1_{\vec{x}}$$

is an $r-1$-partially multlinear map $X \to \mathcal{L}(U_1 \times \cdots \times U_n, V)$. Thus we have shown that if $G : U_1 \times \cdots \times U_n \to W$ is a sum of r-partially multilinear maps then it is differentiable and its derivative

$$DG : U_1 \times \cdots \times U_n \to W_2$$

is a sum of $r-1$-partially multilinear maps, where $W_2 = \mathcal{L}(U_1 \times \cdots \times U_n, W)$. This completes the proof that T is smooth. It also shows that D^nT is 0-partially multilinear, i.e., constant. Thus $D^{n+1}(T) = 0$ as claimed. ☐

Example 3.4. Let $T : \mathbb{R}^n \to \mathbb{R}$ be given by$T(x_1, \ldots, x_n) = x_1 \cdots x_n$. This is a bounded multilinear map. By the product rule, its derivative is equal to

$$DT(x_1, \ldots, x_n)(u_1, \ldots u_n) = u_1 x_2 \cdots x_n + \cdots + x_1 x_2 \cdots u_n.$$

By representing the derivative as a matrix, we see that DT is the row vector

$$[x_2 \ldots x_n, x_1 x_3 \cdots x_n, \ldots, x_1 \ldots x_{n-1}].$$

Example 3.5. Let $U_i = \mathbb{R}^n$ for $i = 1, \ldots n$ and let $T : U_1 \times \cdots \times U_n \to \mathbb{R}$ be given by

$$T(x_1, \ldots, x_n) = \det(x_1 | \cdots | x_n)$$

where $x_1 | \cdots | x_n$ denotes the matrix where the i-th column is the column vector $x_i \in \mathbb{R}^n$. This is a multilinear map. Thus, by the product rule, its derivative is

$$T(x_1, \ldots, x_n)(u_1, \ldots, u_n) = \det(u_1 | x_2 | \cdots | x_n) + \cdots + \det(x_1 | \cdots | x_{n-1} | u_n).$$

In particular, we represent the identity matrix $I = e_1 | \cdots | e_n$ where $e_1, \ldots, e_n$ is the standard basis for $\mathbb{R}^n$. Thus

$$DT(I)(A) = \sum_{i=1}^{n} A_{i,i} = Tr(A)$$

where A is an $n \times n$ matrix.

Lemma 3.6. (General product rule) Suppose that $U_1, \ldots, U_n$ are normed vector spaces and $T : U_1 \times \cdots \times U_n \to V$ is a bounded multilinear map and suppose that $f_i : X \to U_i$ are differentiable at $x_0 \in X$. Then the map

$$H(x) = T(f_1(x), \ldots f_n(x))$$

is differentiable at x_0 and has derivative

$$DH(x_0)u = T((Df_1)(x_0)u, f_2(x_0), \ldots, f_n(x_0)) + \cdots + T(f_1(x_0), f_2(x_0), \ldots, Df_n(x_0)u)$$

Proof. The map $F(x) = (f_1(x), \ldots, f_n(x))$ has derivative

$$DF(x_0)u = (Df_1(x_0)u, \ldots, Df_n(x_0)u)$$

and $H = T \circ F$ hence the derivative is

$$\begin{aligned} D(H)(x_0)u &= D(T)(F(x_0))(DF)(x_0)u \\ &= DT(F(x_0))(Df_1(x_0)u, \ldots, Df_n(x_0)u) \end{aligned}$$

and the result now follows from the product rule for multilinear maps.

□

Example 3.7. Let $H : \mathbb{R} \to \mathbb{R}$ be given by $H(x) = h_1(x) \cdots h_n(x)$ where $h : \mathbb{R} \to \mathbb{R}$ are differentiable for all $x \in \mathbb{R}$. Thus $H(x) = T(f_1(x), \ldots, f_n(x))$ where $T(x_1, \ldots, x_n) = x_1 \cdots x_n$ is from Example 3.4, where we computed the derivative already. Thus we have that

$$DH(x)(u) = ((Df_1(x))u) f_2(x) \cdots f_n(x) + \cdots + f_1(x) \cdots f_{n-1}(x)(Df_n(x)u).$$

Now, for $F:\mathbb{R}\to\mathbb{R}$ we have that $DF:\mathbb{R}\to\mathcal{L}(\mathbb{R},\mathbb{R})$. However, $\mathcal{L}(\mathbb{R},\mathbb{R})$ is isomorphic to $\mathbb{R}$ where we identity ϕ with $\phi(1)$. Thus for differentiable $F:\mathbb{R}\to\mathbb{R}$ we identify $DF(x)\in\mathcal{L}(\mathbb{R},\mathbb{R})$ with $F'(x):=DF(x)(1)\in\mathbb{R}$, to get the usual derivative from basic calculus. This means that

$$H'(x)=f_1'(x)f_2(x)\cdots f_n(x)+\cdots+f_1(x)\cdots f_{n-1}(x)f_n'(x)$$

and so we recover the usual product rule from basic calculus.

The next example will be important for our proof of Taylor's theorem and can be used to give a higher order product rule.

Example 3.8. Consider the composition map

$$\mathcal{L}(U,V)\times\mathcal{L}(W,U)\to\mathcal{L}(W,U)$$

given by $(L_1,L_2)\mapsto L_1\circ L_2$. Then it is a bounded bilinear map. Now suppose that we have maps $g:X\to\mathcal{L}(W,U)$ and $h:X\to\mathcal{L}(U,V)$ differentiable at $x_0\in X$, where X is an open subset of a normed vector space U_0. Then the map $M(x)=h(x)\circ g(x)$ is differentiable at x_0 with derivative given by

$$DM(x_0)u=D(h(x_0))u\circ g(x_0)+h(x_0)\circ Dg(x_0)u \quad \text{for } u\in U_0.$$

Lemma 3.9. (Mean value inequality) Suppose $f:X\to V$ is differentiable on each point of the segment x_0+tu for $t\in[0,1]$ and that $t\mapsto Df(x_0+tu)$ is continuous on $[0,1]$. Then

$$\|f(x_0+u)-f(x_0)\|\le\sup_{t\in[0,1]}\|Df(x_0+tu)u\|.$$

Proof. We first reduce to the case $V=\mathbb{R}$ as follows. By the Hahn-Banach theorem we have a linear $\phi:V\to\mathbb{R}$ such that $\phi(f(x_0+u)-f(x_0))=\|f(x_0+u)-f(x_0)\|$ and $\|\phi\|=1$. So if the theorem is true for $g=\phi\circ f:X\to V$ we must have

$$\|f(x_0+u)-f(x_0)\|=\|g(x_0+u)-g(x_0)\|\le\sup_{t\in[0,1]}\|Dg(x_0+tu)u\|.$$

But by the chain rule and the fact that $D\phi(x)=\phi$ for all linear ϕ, we have that $Dg=\phi\circ Df$. Hence as $\|\phi\|=1$ we have that

$$\|Dg(x_0+tu)u\|=\|\phi\left(Df(x_0+tu)u\right)\|\le\|Df(x_0+tu)u\|.$$

We now turn to show that the theorem holds when $V=\mathbb{R}$. Let $h(t)=f(x_0+tu)$. Note that the classical derivative $h'(t)=Dh(t)(1)$ is continuous, hence we may apply the fundamental theorem of calculus to get

$$\|f(x_0+u)-f(x_0)\|=|h(1)-h(0)|\le|\int_0^1 h'(t)dt|\le\sup_{t\in[0,1]}|h'(t)|.$$

But from the chain rule we get

$$\begin{aligned}
h'(t_0)&=Dh(t_0)(1)\\
&=\left[D(f(x_0+ut_0))\circ D(t\mapsto(x_0+tu))(t_0)\right](1)\\
&=\left[D(f(x_0+t_0u))\circ(t\mapsto tu)\right](1)\\
&=D(f(x_0+t_0u))u
\end{aligned}$$

□

4. Smoothness of higher derivatives

5. The commutativity of higher order derivatives

Theorem 5.1. Suppose that $f : U \to V$ is twice differentiable at x_0, then $D^2 f(x_0)$ is a symmetric bilinear form, that is

$$D^2 f(x_0)(u_1, u_2) = D^2 f(x_0)(u_2, u_1) \quad \text{for all } u_1, u_2 \in U.$$

Proof. It is sufficient to show this for all small enough $u_1, u_2 \in U$. Let $g(x) = f(x+u_1) - f(x)$. If $L : U \to V$ is the bounded linear map $L = D^2 f(x_0)(u_1)$ then by the mean value inequality applied to $g - L$ we have that

$$\|g(x_0 + u_2) - g(x_0) - L(u_2)\| \leq \sup_{t\in[0,1]} \|Dg(x_0 + tu_2)u_2 - L(u_2)\| \tag{1}$$

for all small enough $u_1, u_2 \in U$.

By the definition of differentiability we have that

$$Df(x_0 + u) = Df(x_0) + D^2 f(x_0)(u) + R(u)$$

where $\|R(u)\| \leq \epsilon(u)\|u\|$ where $\lim_{u\to 0} \epsilon(u) = 0$. We use this inequality to estimate the upper bound in (1) as follows

$$\begin{aligned} Dg(x_0 + tu_2)u_2 &= Df(x_0 + tu_2 + u_1)u_2 - Df(x_0 + tu_2)u_2 \\ &= D^2 f(x_0)(tu_2 + u_1)u_2 - D^2 f(x_0)(tu_2)u_2 + R(tu_2 + u_1)u_2 - R(tu_2)u_2 \\ &= D^2 f(x_0)(u_1)(u_2) + R(tu_2 + u_1)u_2 - R(tu_2)u_2 \end{aligned}$$

Hence (1) gives the estimate

$$\begin{aligned} \|g(x_0 + u_2) - g(x_0) - D^2 f(x_0)(u_1, u_2)\| &\leq \sup_{t\in[0,1]} \|R(tu_2 + u_1)u_2 - R(tu_2)u_2\| \\ &\leq \epsilon_0(\|u_1\| + \|u_2\|)(\|u_1\| + \|u_2\|)^2 \end{aligned}$$

for some $\lim_{u\to 0} \epsilon_0(\|u\|) = 0$. In particular, by replacing u_i with su_i for $0 < s < 1$ and dividing by s^2 we get the limit

$$\begin{aligned} D^2 f(x_0)(u_1)(u_2) &= \lim_{s\to 0} \frac{g(x_0 + su_2) - g(x_0)}{s^2} \\ &= \lim_{s\to 0} \frac{f(x_0 + su_1 + su_2) - f(x_0 + su_2) - f(x_0 + su_1) + f(x_0)}{s^2} \end{aligned}$$

but this term inside the limit is symmetric in u_1 and u_2. $\square$

We can apply this inductively to get (notice that we don't require $D^n f(x)$ to be defined or continuous on any open set):

Theorem 5.2. Suppose $X \subset U$ is open and that $f : X \to V$ is such that $D^n f(x_0)$ exists at $x_0 \in X$. Then $D^n f(x_0)$ is a symmetric multilinear form.

Proof. We prove by induction and assume that $D^{n-1}f(x)$ exists for all $x \in X$ (by shrinking X if necessary). The base case $n = 2$ is shown above.

We let $\mathcal{S}_k(U,V)$ denote those bounded multilinear maps $T : U^k \to V$ such that $T(u_1, \dots, u_k)$ is invariant under permutations of the u_i. Note that $\mathcal{S}_k(U,V)$ is closed in $\mathcal{T}_k(U,V)$.

Suppose that $n \geq 3$ and the result is true for $n-1$. Now observe that for $u_1 \in U$ we have that

$$D(D^{n-1}f)(x_0)(u_1) = \lim_{t \to 0} \frac{1}{t} \left((D^{n-1}f)(x_0 + u_1 t) - (D^{n-1})f(x_0) \right)$$

and thus the left hand side is a limit of elements in $\mathcal{S}_{n-1}(U,V)$ and thus $D(D^{n-1}f)(x_0)(u_1) \in \mathcal{S}_{n-1}(U,V)$. In other words $D^n f(x_0)(u_1, u_2, \dots, u_n)$ is invariant under permutation of the $u_2, \dots, u_n$. It thus remains to show that it is invariant under permuting u_1 and u_2.

By the case applied to $D^{n-2}f$ we have

$$D^2(D^{n-2}f)(x_0)(u_1, u_2) = D^2(D^{n-2}f)(x_0)(u_2, u_1).$$

Thus

$$D^2(D^{n-2}f)(x_0)(u_1, u_2)(u_3, \dots, u_n) = D^2(D^{n-2}f)(x_0)(u_2, u_1)(u_3, \dots, u_n).$$

Finally, the proof is complete as $D^2(D^{n-2}f)(x_0)(u_1, u_2)(u_3, \dots, u_n) = (D^n f)(x_0)(u_1, \dots, u_n)$. ☐

6. Taylor's Theorem

If $h : \mathbb{R} \to \mathbb{R}$ is N times continuously differentiable at $x_0 \in \mathbb{R}$ then the classical Taylor expansion is

$$h(x_0 + u) = h(x_0) + h'(x_0)u + \frac{1}{2!}h^{(2)}(x_0)u^2 + \cdots + \frac{1}{N!}h^{(N)}(x_0)u^N + R(u)$$

with error term

$$R(u) = \frac{1}{(N+1)!} \int_0^u h^{(N+1)}(x_0 + t)t^n dt.$$

We now ask how can this be formulated for $f : U \to \mathbb{R}$ where U is an arbitrary vector space? If we define u^n to be the n-tuple $(u, \dots, u)$ then the expression $Df^n(x_0)u^n$ makes sense, as $Df^n(x_0)$ is a multilinear map. Hence a strategy of deriving a Taylor theorem involves first parametrizing the line segment from x_0 to $x_0 + u$ by defining $L(t) = tu$ and applying the classical Taylor theorem to $h = f \circ L$. To do this we need to compute $D^n(f \circ L)(t_0)$.

Lemma 6.1. Let $L : W \to U$ be a linear map and let $f : U \to V$ be n times differentiable at $L(a)$, where $a \in W$. Then

$$D^n(f \circ L)(a) = D^n f(L(a)) \circ L^n$$

where $L^n(w_1, \dots, w_n) := (Lw_1, \dots, Lw_n)$.

Remark 6.2. The is a generalization of the standard identity $\frac{d}{dt}(h(bt)) = b^n h^{(n)}(bt)$ from classical differential calculus.

Proof. For $n = 1$ this is the chain rule and the fact that $D(L)(a) = L$ for linear maps L. Now we proceed by induction. Assuming the result is true for n, we now suppose that f is $n+1$ differentiable at $L(a)$. Then we get

$$\begin{aligned} D^{n+1}(f \circ L)(a) &= D(D^n(f \circ L))(a) \\ &= D(w \mapsto D^n f(Lw) \circ L^n)(a). \end{aligned}$$

To compute this term we use the product rule, specifically Example 3.8 to the map $w \mapsto D^n f(Lw)$ and the constant map $w \mapsto L^n$. The latter has zero derivative so we only get the first term from Example ((3.8)). This means that

$$\begin{aligned} D^{n+1}(f \circ L)(a)(w_0, w_1, \dots, w_n) &= D^{n+1}(f \circ L)(a)(w_0)(w_1, \dots, w_n) \\ &= (D(w \mapsto D^n f(Lw))(a)(w_0) \circ L^n)(w_1, \dots, w_n) \end{aligned}$$

where we firstly used our canonical isomorphism

$$\mathcal{L}(W_0, \mathcal{T}(W_1, \dots, W_n \to V)) \cong \mathcal{T}(W_0, \dots, W_n \to V)$$

with $W_i = W$ followed by Example 3.8 as described.

Now to compute this final derivative we use the chain rule as follows. Let $F(u) = D^n f(u)$. Then this final derivative is $D(F \circ L)(a)$ hence equal to $D(F)(La) \circ L = D^{n+1} f(La) \circ L$. So we have shown that

$$\begin{aligned} D^{n+1}(f \circ L)(a)(w_0, \dots, w_n) &= \left((D^{n+1} f(La) \circ L)w_0 \circ L^n)\right)(w_1, \dots, w_n) \\ &= D^{n+1} f(La)(Lw_0)(Lw_1, \dots, Lw_n) \end{aligned}$$

and the proof is complete once one again applies the canonical isomorphism. □

Theorem 6.3. (Co-ordinate free general Taylor theorem) Suppose $f : X \to \mathbb{R}$ is $(n+1)$ times continuously differentiable and let $x_0 \in X$ and $u \in U$ be such that the line segment $x_0 + tu$ is in X. Then we have that

$$f(x_0 + u) = f(x_0) + \sum_{n=1}^{N} \frac{1}{n!} D^n f(x_0) u^n + R(u)$$

where $u^n = (u, \dots, u) \in U^n$ and

$$R(u) = \frac{1}{(N+1)!} \int_0^1 D^{N+1} f(x_0 + tu) u^{n+1} t^n dt.$$

Proof. By translation invariance of derivatives, it is enough to prove this for $x_0 = 0$. Now write $f(x_0 + tu) - f(x_0) = h(1) - h(0)$ and use Taylor's theorem, where $h = f \circ L$ where $L : \mathbb{R} \to U$ is given by $L(t) = tu$. The result now follows from the Lemma 6.1 □

7. Recovering the co-ordinate based Taylor theorem

To recover the Taylor theorem most often presented in undergraduate vector calculus, one needs to expand $D^n f(x_0) u^n$ as follows. Writing

$$u = \sum_i u_i e_i$$

where $e_1, \dots, e_d$ is a basis for U, we have that

$$D^n f(x_0)(e_{i_1}, \dots, e_{i_n}) = \partial_{i_1} \cdots \partial i_n f(x_0),$$

where ∂_i is partial differentiation with respect to the e_i direction. Hence by multilinearity we get that

$$D^n f(x_0) u = \sum_{i_1, \dots, i_n} \partial_{i_1} \cdots \partial_{i_n} f(x_0) u_{i_1} \dots u_{i_n}.$$

Now by commutativity of derivatives the order of $i_1, \ldots, i_n$ does not matter, hence it pays to express this in terms of partitions as follows. Given $\alpha = (a_1, \ldots, a_d) \in \mathbb{Z}^d_{\geq 0}$ such that $\sum_i a_i = n$ we define

$$D^\alpha f(x_0) = \partial_1^{a_1} \cdots \partial_n^{a_d} f(x_0)$$

where ∂_i^k means ∂_i applied k times. Define also

$$u^\alpha = \prod_i u_i^{a_i}.$$

Then by grouping the terms together we get the expression

$$D^n f(x_0)u = \sum_{|\alpha|=n} \binom{n}{\alpha} D^\alpha f(x_0) u^\alpha$$

where $|\alpha| = a_1 + \cdots + a_d$ and

$$\binom{n}{\alpha} = \frac{n!}{a_1! \cdots a_d!}$$

is the multinomial coefficient which counts the number of sequences $\{1, \ldots, d\}^n$ where i appears a_i times. Finally, this means that the Taylor theorem may be written as

$$f(x_0 + u) = f(x_0) + \sum_{n=1}^N \sum_{|\alpha|=n} \frac{1}{\alpha!} D^\alpha f(x_0) u^\alpha + R(x)$$

where

$$\alpha! = a_1! \cdots a_d!$$

8. Sufficient conditions for differentiability

It is all well and good proving abstract theorems about derivatives, but so far we don't have a mechanism for quickly telling whether a certain function is differentiable. Fortunately we have the following criteria.

Proposition 8.1. Suppose that $X \subset \mathbb{R}^d$ is open and $f : X \to V$ has the property that all its partial derivatives exist and are continuous. Then f is continuously differentiable on X.

Proof. The hypothesis says that each limit

$$\partial_i f(x) = \lim_{t \to 0} \frac{f(x + te_i)}{t}$$

exists and is continuous as a function of x. To show differentiability at $x_0 \in X$, we first construct the derivative $D = Df(x_0)$ and then show it satisfies the definition. Without loss of generality, we will use the 1-norm $\| \cdot \| = \| \cdot \|_1$ for convenience. We define D to be the linear map

$$Du = \sum_{i=1}^n u_i \partial_i f(x_0)$$

for $u = \sum_i u_i e_i$. To show that D satisfies the definition of the derivative fix $\epsilon > 0$ and find a corresponding $\delta > 0$ such that $\|\partial_i f(x) - \partial_i f(x_0)\| < \epsilon$ for all x such that $\|x - x_0\| < \delta$. Choose u such that each $\|u_i e\| < \frac{1}{d}\delta$ and let $w_i = u_1 + \ldots + u_i$ and $w_0 = 0$ (notice that this ensures each $\|w_i\| < \delta$). We have

$$
\begin{aligned}
f(x_0+u)-f(x_0) &= \sum_{i=1}^{d} f(w_i)-f(w_{i-1}) \\
&= \sum_{i=1}^{d} f(w_{i-1}+u_ie_i)-f(w_i) \\
&= \sum_{i=1}^{d} (u_i\partial_i f(w_{i-1})+R_i) \\
&= Du+\sum_{i=1}^{d} u_i\left(\partial_i f(w_{i-1})-\partial_i f(x_0)\right)+\sum_{i=1}^{d} R_i
\end{aligned}
$$

where the remainder term R_i satisfies

$$\|R_i\| \leq \sup_{t\in[0,1]} \|u_i\partial_i f_i(w_{i-1}+tu_ie_i)-u_i\partial_i f_i(w_0)\| \leq 2\epsilon\|u_i\|$$

by the mean value inequality applied to the function $u \mapsto f(w_{i-1}+ue_i)$. So altogether we have shown that if $\|u_ie_i\| < \frac{1}{d}\delta$ then we have

$$\|f(x_0+u)-f(x_0)-Du\| \leq 3\epsilon\|u\|$$

which verifies differentiability. □

This generalizes to higher order differentiability. We will focus on finite dimensional spaces. If U has basis $e_1,\dots,e_d$ and V has basis $b_1,\dots,b_m$ then the corresponding standard basis for $\mathcal{L}(U,V)$ is given by $\delta_{i,j}$ for $1\leq i\leq m$, $1\leq j\leq d$ where

$$\delta_{i,j}\left(\sum_{\ell} u_\ell e_\ell\right) = u_jb_i.$$

More generally, the space of k-multilinear map $T: U\times\cdots\times U\to V$ has a standard basis where the elements are indexed by $i,j_1,\dots,j_k$ where such an element maps $(e_{j_1},\dots,e_{j_k})$ to b_i but all other tuples in $\{e_1,\dots,e_k\}^k$ map to 0.

Theorem 8.2. Suppose that $X\subset U=\mathbb{R}^d$ is open and $f: X\to V=\mathbb{R}^m$ has the property that all its n-order partial derivatives exist and are continuous. Then f is continuously n times continuously differentiable on X.

Proof. We have shown the $n=1$ case. Proceeding by induction, suppose that it holds for n. Suppose that $f: X\to\mathbb{R}^m$ has the property that all its $n+1$-order partial derivatives exist and are continuous. The induction hypothesis says that $D^nf: X\to\mathcal{T}(U,\dots,U\to V)$ exists and is continuous. But with respect to the standard basis for $\mathcal{T}(U,\dots,U\to V)$ the coefficients of $D^nf(x_0)$ are n-order partial derivatives of f at x_0 and their partial derivatives are the order $(n+1)$ partial derivatives, which exist and are continuous by hypothesis. Applying the already shown $n=1$ case establishes the inductive step. □

9. Application: Laplacian and Wave Operator

9.1. **Motivation.** The Laplacian of a twice differentiable $f:\mathbb{R}^3\to\mathbb{R}$ is defined as

$$\triangle f = \partial_1^2 f+\partial_2^2 f+\partial_3^2 f.$$

In elementary courses one shows that

$$(\triangle f)\circ Q = \triangle(f\circ Q)$$

for $Q \in O(3)$. In mathematical physics and PDEs, one is often interested in the wave operator (D'Alembertian) of $f : \mathbb{R}^4 \to \mathbb{R}$ as

$$\Box f = \partial_1^2 f + \partial_2^2 f + \partial_3^2 f - \partial_4^2 f$$

and one shows that it is invariant under Lorentz transformations, i.e., if $L \in O(3,1)$ then

$$\Box(f \circ L) = (\Box f) \circ L.$$

Our aim is to prove these in an elegant and co-ordinate free way in the framework developed here.

9.2. **Trace of a bilinear form.** Let V be a finite dimensional vector space and suppose that

$$(v_1, v_2) \mapsto \langle v_1, v_2 \rangle$$

is a symmetric, non-degenerate bilinear form $V \times V \to \mathbb{R}$.

Lemma 9.1. If $B : V \times V \to \mathbb{R}$ is a bilinear form, then there exists a unique linear map $L = L_B : V \to V$ such that

$$B(v_1, v_2) = \langle v_1, L v_2 \rangle.$$

Proof. The map $L_B \mapsto B$ is a well defined linear map $\mathcal{T}(V, V \to \mathbb{R}) \to \mathcal{L}(V \to V)$. It is injective since $\langle -, - \rangle$ is non-degenerate. It is therefore surjective since its domain and codomain have equal dimensions $(\dim V)^2$. ◻

If $Q : V \to V$ is a linear automorphism then we say it is *orthogonal* (with respect to $\langle -, - \rangle$) if

$$\langle Q v_1, Q v_2 \rangle = \langle v_1, v_2 \rangle.$$

Lemma 9.2. If Q is orthogonal then

$$L_{B \circ Q^2} = Q^{-1} L_B Q$$

where $Q^2(v_1, v_2) = (Q v_1, Q v_2)$.

Proof. For each fixed $v_0 \in V$ we have that

$$\langle v_0, L_{B \circ Q^2} v \rangle = (B \circ Q^2)(v_0, v) = B(Q v_0, Q v) = \langle Q v_0, L_B Q v \rangle = \langle v_0, Q^{-1} L_B Q v \rangle.$$

Since this holds for all $v_0 \in V$ and $\langle -, - \rangle$ is non-degenerate, we have the desired equality. ◻

If $Tr(L)$ denote the trace of a linear operator $L : V \to V$ then we obtain the following corollary.

Corollary 9.3. If Q is orthogonal then

$$Tr(L_B) = Tr(L_{B \circ Q^2})$$

To see how this demonstrates the symmetries of the Laplacian and D'Alembertian, we observe the following. First, let $c_1, \ldots, c_n \in \mathbb{R} \setminus \{0\}$ be non-zero real numbers. We now consider on $\mathbb{R}^n$ the symmetric, non-degenerate bilinear form given by

$$\langle v, w \rangle = \sum_{i=1}^{n} c_i v_i w_i.$$

Now fix a twice differentiatble function $f : \mathbb{R}^n \to \mathbb{R}$. For each $x \in \mathbb{R}^n$, we have a unique linear operator $L_{D^2 f(x)}$ such that

$$D^2 f(x)(v, w) = \langle L_{D^2 f(x)} v, w \rangle.$$

Let $e_1, \ldots, e_n$ be the standard basis for $\mathbb{R}^n$. Since

$$\frac{1}{c_i}\langle e_i, e_j\rangle = \delta_{i,j}$$

we compute the trace to be

$$\begin{aligned} Tr(L_{D^2 f(x)}) &= \sum_{i=1}^{n} \frac{1}{c_i}\langle L_{D^2 f(x)} e_i, e_i\rangle \\ &= \sum_{i=1}^{n} \frac{1}{c_i} D^2 f(x)(e_i, e_i) \\ &= \sum_{i=1}^{n} \frac{1}{c_i}\partial_i^2 f. \end{aligned}$$

Now observe that for the case $c_1 = \cdots = c_n = 1$ this trace is the Laplacian. While for $c_1 = \cdots = c_{n-1} = 1$, $c_n = -c^2$ this trace is the D'Alembertian for the wave equation with speed $c > 0$. We know by Lemma 6.1 that if $Q : \mathbb{R}^n \to \mathbb{R}^n$ is al linear transformation then

$$D^2(f \circ Q)(x) = D^2 f(Qx) \circ Q^2.$$

Now if Q is orthogonal with respect to our quadratic form, then

$$Tr(L_{D^2(f\circ Q)(x)}) = Tr(L_{D^2 f(Qx)})$$

by the corollary above. We have thus shown

Theorem 9.4. If Q is orthogonal with respect to $\langle -, - \rangle$ then

$$Tr(L_{D^2(f\circ Q)(x)}) = Tr(L_{D^2 f(Qx)}).$$

In particular, this shows that

$$\triangle(f \circ Q) = (\triangle f) \circ Q$$

for orthogonal linear transformation Q on Euclidean space $\mathbb{R}^n$. The same holds if for $\square$ in place of $\triangle$ if Q is a Lorentz transformation.