

# TAYLOR'S THEOREM IN NORMED VECTOR SPACES

KB-MATH

## 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  $f : X \rightarrow V$  is differentiable at  $x_0 \in X$  if there exists a bounded linear map  $D : U \rightarrow V$  and  $\epsilon : [0, \infty) \rightarrow [0, \infty)$  with  $\lim_{r \rightarrow 0} \epsilon(r) = 0$  such that

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

for all  $u$  in some open neighbourhood of 0.

**Lemma 1.2.** Such a  $D$  must satisfy

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

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

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

The result now follows by letting  $t \rightarrow \infty$ . □

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  $f : X \rightarrow V$  is differentiable at  $x_0$  then it is continuous at  $x_0$ .

*Proof.* By triangle inequality we have that

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

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

## 2. HIGHER ORDER DERIVATIVES

How do we define higher order derivatives in this setting? If  $f : X \rightarrow V$  is differentiable on  $X$  then  $Df : X \rightarrow \mathcal{L}(U, V)$ . 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

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

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

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

and hence we get a **multilinear map**

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

called the *second derivative* of  $f$ .

**Lemma 2.1.** If  $f : \mathbb{R}^d \rightarrow V$  is twice differentiable then

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

where  $\partial_i$  is partial differentiation with respect to  $i$ -th co-ordinate.

This can be generalized to higher order derivatives, hence  $D^n f(x_0) : U \times \cdots \times U \rightarrow V$  may be realised as a multilinear map, 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  $U_1, \dots, U_n$  and  $V$ , let  $\mathcal{T}(U_1, \dots, U_n \rightarrow V)$  denote the vector space of *bounded* multilinear maps  $T : U_1 \times \cdots \times U_n \rightarrow V$ , where bounded means that

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

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

**Proposition 2.3.** There is a canonical isometric isomorphism

$$\mathcal{L}(U_1, \mathcal{T}(U_2, \dots, U_n \rightarrow V)) \cong \mathcal{T}(U_1, \dots, U_n \rightarrow V)$$

given by

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

In other words, we could have defined  $\|T\|$  as the metric induced by the canonical isomorphism

$$\mathcal{T}(U_1, \dots, U_n \rightarrow V) \cong \mathcal{L}(U_1, \dots, \mathcal{L}(U_{n-1}, \mathcal{L}(U_n, V)) \dots)$$

**Proposition 2.4.** A multilinear map  $T : U_1 \times \cdots \times U_n \rightarrow V$  is bounded if and only if it is continuous with respect to the product topology.

Hence the  $n$ th derivative  $D^n f : X \rightarrow \mathcal{T}(U, \dots, U \rightarrow V)$  may be recursively defined as  $D^1 f = Df$  (which makes sense since  $\mathcal{T}(U \rightarrow V) = \mathcal{L}(U, V)$ ) 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  $X$ ).

### 3. BASIC PROPERTIES

**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  $f : X \rightarrow Y$  is differentiable at  $x_0 \in X$  and  $g : Y \rightarrow W$  is differentiable at  $y_0 = f(x_0) \in Y$ . Then  $g \circ f : X \rightarrow W$  is differentiable at  $x_0$  and

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

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.

**Lemma 3.2.** (Product Rule for multilinear maps) Suppose that  $U_1, \dots, U_n$  are normed vector spaces and  $T : U_1 \times \dots \times U_n \rightarrow V$  is a bounded multilinear map. Then  $T$  is differentiable at each  $\vec{x} = (x_1, \dots, x_n) \in U_1 \times \dots \times U_n$  with

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

**Lemma 3.3.** (General product rule) Suppose that  $U_1, \dots, U_n$  are normed vector spaces and  $T : U_1 \times \dots \times U_n \rightarrow V$  is a bounded multilinear map and suppose that  $f_i : X \rightarrow U_i$  are differentiable at  $x_0 \in X$ . Then the map

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

is differentiable at  $x_0$  and has derivative

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

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

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

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

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

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

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

**Example 3.4.** Consider the composition map

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

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  $g : X \rightarrow \mathcal{L}(W, U)$  and  $h : X \rightarrow \mathcal{L}(U, V)$  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 3.5.** (Mean value inequality) Suppose  $f : X \rightarrow V$  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

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

*Proof.* We first reduce to the case  $V = \mathbb{R}$  as follows. By the Hahn-Banach theorem we have a linear  $\phi : V \rightarrow \mathbb{R}$  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  $g = \phi \circ f : X \rightarrow \mathbb{R}$  we must have

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

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

$$\|Dg(x_0 + tu)u\| = \|\phi Df(x_0 + tu)u\| \leq \|Df(x_0 + tu)u\|.$$

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

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

But from the chain rule we get

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

□

#### 4. THE COMMUTATIVITY OF HIGHER ORDER DERIVATIVES

**Theorem 4.1.** Suppose that  $f : U \rightarrow V$  is twice differentiable at  $x_0$ , then  $D^2f(x_0)$  is a symmetric bilinear form, that is

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

*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  $L : U \rightarrow V$  is the bounded linear map  $L = D^2f(x_0)(u_1)$  then by the mean value inequality applied to  $g - L$  we have that

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

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

By the definition of differentiability we have that

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

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

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

Hence (1) gives the estimate

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

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

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

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

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 4.2.** Suppose that  $f : U \rightarrow V$  is such that  $D^n f(x_0)$  exists at  $x_0 \in U$ , then  $D^n f(x_0)$  is a symmetric multilinear form.

## 5. TAYLOR'S THEOREM

If  $h : \mathbb{R} \rightarrow \mathbb{R}$  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  $f : U \rightarrow \mathbb{R}$  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  $L(t) = x_0 + tu$  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 5.1.** Let  $L : W \rightarrow U$  be a linear map and let  $f : U \rightarrow V$  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 5.2.** This is a generalization of the standard identity  $h^{(n)}(at) = a^n h^{(n)}(t)$  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 3.4 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 ((3.4)). 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

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

with  $W_i = W$  followed by Example 3.4 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

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

and the proof is complete once one again applies the canonical isomorphism.  $\square$

**Theorem 5.3.** (Co-ordinate free general Taylor theorem) Suppose  $f : X \rightarrow \mathbb{R}$  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

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

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  $L : \mathbb{R} \rightarrow U$  is given by  $L(t) = tu$ . The result now follows from the Lemma 5.1  $\square$

## 6. 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

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

by the discussion surrounding Lemma 2.1 on iterated derivatives, where  $\partial_i$  is partial differentiation with respect to the  $e_i$  direction. Hence by multilinearity we get that

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

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

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

where  $\partial_i^k$  means  $\partial_i$  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  $\{1, \dots, d\}^n$  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

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

## 7. 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 7.1.** Suppose that  $X \subset \mathbb{R}^d$  is open and  $f : X \rightarrow V$  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

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

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  $\|u_i e_i\| < \frac{1}{d}\delta$  and let  $w_i = x_0 + u_i e_i$  and  $w_0 = x_0$  (notice that this ensures each  $\|w_i - w_0\| < \delta$ ). We have

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

where the remainder term  $R_i$  satisfies

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

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

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

which verifies differentiability.  $\square$

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

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

More generally, the space of  $k$ -multilinear map  $T : U \times \dots \times U \rightarrow V$  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  $\{e_1, \dots, e_d\}^k$  map to 0.

**Theorem 7.2.** Suppose that  $X \subset U = \mathbb{R}^d$  is open and  $f : X \rightarrow V = \mathbb{R}^m$  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  $f : X \rightarrow \mathbb{R}^m$  has the property that all its  $n + 1$ -order partial derivatives exist and are continuous. The induction hypothesis says that  $D^n f : X \rightarrow \mathcal{T}(U, \dots, U \rightarrow V)$  exists and is continuous. But with respect to the standard basis for  $\mathcal{T}(U, \dots, U \rightarrow V)$  the coefficients of  $D^n f(x_0)$  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.  $\square$

## 8. APPLICATION: LAPLACIAN AND WAVE OPERATOR

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

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

In elementary courses one shows that

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

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

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

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

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

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

**8.2. 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  $V \times V \rightarrow \mathbb{R}$ .

**Lemma 8.1.** If  $B : V \times V \rightarrow \mathbb{R}$  is a bilinear form, then there exists a unique linear map  $L = L_B : V \rightarrow V$  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  $\mathcal{T}(V, V \rightarrow \mathbb{R}) \rightarrow \mathcal{L}(V \rightarrow V)$ . It is injective since  $\langle -, - \rangle$  is non-degenerate. It is therefore surjective since its domain and codomain have equal dimensions  $(\dim V)^2$ .  $\square$

If  $Q : V \rightarrow V$  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 8.2.** 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.  $\square$

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

**Corollary 8.3.** 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, notice that the Laplacian and D'Alembertian of a function  $f$  may be written as the trace of  $L_{D^2 f(x)}$  with respect to the Euclidean norm and Minkowski pseudo norm, respectively. On the other hand, by Lemma REF we have that

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

i.e.,  $D^2(f \circ Q)(x)(v_1, v_2) = D^2(f(Qx))(Qv_1, Qv_2)$ , so the corollary applies.