SOME REMARKS ON TEACHING MATHS: DIFFERENTIAL CALCULUS

WOLFGANG BERTRAM

This note is the translation of a short text I had first written in French, and which later became the germ of the book [B3] (see list of papers). The general context of *topological differential calculus* is discussed here (Chapter 1). I concentrate in the following on the most basic items: in finite dimension, "topological" and "metric" calculus are equivalent, and the topological approach permits to give very simple, "algebraic" proofs of some basic facts: the first differential $Df(x)$ is a *linear* map; the *chain rule* and *Schwarz' lemma* hold, and there is a *Taylor expansion*.

1. Equivalence of calculi: first order

Let V and W be finite-dimensional real vector spaces, U a non-empty open subset of V and $f : U \to W$ a map. The *(first) prolongation of U* is the set defined by

$$U^{[1]} := \{(x, v, t) | \, x \in U, v \in V, x + tv \in U\}, \tag{1.1}$$

whichis open in $V \times V \times \mathbb{R}$. Recall that, in usual calculus, f is said to be *continously differentiable*, or: *of class C^1*, if all partial derivatives of f exist and are continuous. Equivalently, all *directional derivatives*

$$\partial_v f : U \to W, \quad x \mapsto \partial_v f(x) = \lim_{t \to 0} \frac{f(x + tv) - f(x)}{t} \tag{1.2}$$

exist and are continuous.

Theorem 1.1. *The following statements are equivalent:*

1. *the map f is of class C^1,*
2. *the map f admits a* continuous difference factorizer*, that is, there is a continuous map $f^{[1]} : U^{[1]} \to W$ such that, for all $(x, v, t) \in U^{[1]}$,*

$$f(x + tv) - f(x) = t \cdot f^{[1]}(x, v, t).$$

If this holds, the total differential of f is given by $Df(x)v = f^{[1]}(x, v, 0)$.

Proof. Assume (2) holds. Then $\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(x+tv)-f(x)}{t} = \lim_{\substack{t \to 0 \\ t \neq 0}} f^{[1]}(x, v, t) = f^{[1]}(x, v, 0)$, by continuity of $f^{[1]}$. Thus all directional derivatives $(\partial_v f)(x)$ exist on U, they are given by $f^{[1]}(x, v, 0)$, hence are continuous. In particular, all partial derivatives exist and are continuous, hence f is C^1. Now assume that (1) holds and define a map

$$f^{[1]} : U^{[1]} \to F, \quad (x, v, t) \mapsto \begin{cases} \frac{f(x+tv)-f(x)}{t} & \text{if} \quad t \in \mathbb{R}^\times \\ Df(x)v & \text{if} \quad t = 0. \end{cases}$$

We must show that $f^{[1]}$ is continuous. Since f is continuous, it follows that $f^{[1]}$ is continuous on the set

$$U^{]1[} := \{(x, v, t) \in U^{[1]} | \, t \neq 0\}. \tag{1.3}$$

Thus it remains to prove that $f^{[1]}$ is continous at all points of the form $(x,v,0)$. To show this, note first that, whenever the line segment $[x,x+tv]$ belongs to U, then[1]

$$f^{[1]}(x,v,t) = \int_0^1 Df(x+stv)v\,ds. \tag{1.4}$$

Indeed, for $t=0$, this is obviously true, and for $t \neq 0$, it follows from the fundamental relation between differential and integral calculus, along with a change of variable $r=st$,

$$f(x+tv) = f(x) + \int_0^t Df(x+rv)v\,dr = f(x) + t\int_0^1 Df(x+stv)v\,ds$$

which implies (1.4). Now, the right hand side is the integral of a continuous function, depending continuously on the parameter (x,v,t). Since integration carries over a compact interval, the continuity of the right hand side follows by elementary analysis (no need to invoke more powerful machinery like Lebesgue's theorem).[2] □

2. Linearity of the differential

Theorem 2.1. *Assume V,W are finite-dimensional vector spaces and $f : U \to W$ a map of class C^1 defined on an open subset of V. Then its differential $Df(x) : V \to W$ at each point $x \in U$ is a linear map.*

Proof. Let us show that $Df(x)(v+w) = Df(x)v + Df(x)w$ for $v,w \in V$. For $t \neq 0$, we have

$$\begin{aligned} f^{[1]}(x,v+w,t) &= \frac{f(x+t(v+w)) - f(x)}{t} \\ &= \frac{f(x+tv+tw) - f(xx+tv)}{t} + \frac{f(x+tv)-f(x)}{t} \\ &= f^{[1]}(x+tv,w,t) + f^{[1]}(x,v,t). \end{aligned}$$

Thus $f^{[1]}(x,v+w,t) = f^{[1]}(x+tv,w,t) + f^{[1]}(x,v,t)$ for all non-zero t in a neighborhood of 0. Since, by Theorem 1.1, both sides are continuous functions of t, it follows that equality also holds for $t=0$, which directly implies the claim. Now we prove that $Df(x)(rv)r = r\,Df(x)v$. For $t \neq 0$ and $r \neq 0$,

$$\begin{aligned} f^{[1]}(x,rv,t) &= \frac{f(x+t(rv)) - f(x)}{t} \\ &= r\frac{f(x+tr\,v) - f(x)}{tr} = r\,f^{[1]}(x,v,tr). \end{aligned}$$

Thus $f^{[1]}(x,rv,t) = rf^{[1]}(x,v,tr)$ for all non-zero t,r in a neighborhood of 0. Again by Theorem 1.1, eqality still holds for $t=0$: $Df(x)(rv) = rDf(x)v$. □

In finite dimension, $\mathrm{Hom}(V,W)$ carries a canonical topology, and it is now easily seen that $Df : U \to \mathrm{Hom}(V,W)$, $x \mapsto Df(x)$ is a continuous map.

---

[1] In France, vector valued integrals like the following are hardly used in teaching – by some Bourbaki tradition, calculus is often taught in the Banach-setting, and there vector valued integrals are refused as kind of impure objects. But in $\mathbb{R}^n$, one may just define the integral componentwise, and prove (or not) that this doesn't depend on the choice of basis.

[2] In [B3], this elementary analysis is developed from scratch.

3. Chain rule

Theorem 3.1. *If g and f are composable mappings of class C^1 defined on open subsets of finite-dimensional vector spaces, then $g \circ f$ is again of class C^1, and its differential is given by*

$$D(g \circ f)(x) = Dg(f(x)) \circ Df(x).$$

Proof. We use again Theorem 1.1 and write $f(x + tv) = f(x) + tf^{[1]}(x, v, t)$. Then, for $t \neq 0$ the difference quotient is

$$\begin{aligned}(g \circ f)^{[1]}(x, v, t) &= \frac{g(f(x + tv)) - g(f(x))}{t} \\ &= \frac{g(f(x) + tf^{[1]}(x, v, t)) - g(f(x))}{t} = g^{[1]}(f(x), f^{[1]}(x, v, t), t).\end{aligned}$$

Our assumptions imply that the right hand side is a continuous function of (x, v, t), and hence a continuous difference factorizer for $g \circ f$ exists, given by this formula. Taking $t = 0$ now yields $D(g \circ f)(x)v = Dg(f(x))(Df(x)v)$, as claimed. □

4. Equivalence of calculi: higher order, and Schwarz' lemma

Theorem 4.1. *With notation as above, the following statements are equivalent:*

(1) f is of class C^2.
(2) f is of class C^1 and $f^{[1]}$ is of class C^1.

Proof. Assume (2) holds. Since $Df(x)v = f^{[1]}(x, v, 0)$, and $f^{[1]}$ is C^1, it follows that Df is also of class C^1, and thus f is of class C^2.

Assume (1) holds. According to theorem 1.1, the continuous difference factorizer $f^{[1]}$ exists. We have to prove that it is C^1. Since f is C^1, the difference factorizer is C^1 on the set $U^{]1[}$ defined by (1.3). Let us prove that it is C^1 on a neighborhood of a point of the form $(x, v, 0)$. To this end, we use once more the integral representation (1.4) of $f^{[1]}$. As in the proof of theorem 1.1, elementary analysis shows that the function defined by the integral over the compact interval $[0, 1]$ is continuously differentiable (integration and differentiation commute), by using the fact that the function in the integral is C^1, by our assumptions. □

If f is C^2, we define the *second order difference factorizer* $f^{[2]} := (f^{[1]})^{[1]}$. By induction, we define $f^{[k]} := (f^{[k-1]})^{[1]}$, and arguments similar as above show:

Theorem 4.2. *The following statements are equivalent:*

(1) f is of class C^k.
(2) f is of class C^{k-1} and $f^{[k-1]}$ is of class C^1.
(3) f is C^1 and $f^{[1]}$ is of class C^{k-1}.

Assume f is of class C^2. Let us give an explicit formula for the second order difference factorizer, assuming that the scalars by which we devide are non-zero:

$$\begin{aligned}f^{[2]}\big((x, v, t), (x', v', t'), t''\big) &= \tfrac{1}{t''}\Big(f^{[1]}\big((x, v, t) + t''(x', v', t')\big) - f^{[1]}(x, v, t)\Big) \\ &= \frac{f(x + t''x' + (t + t''t')(v + t''v')) - f(x + t''x')}{t''(t + t''t')} - \frac{f(x + tv) - f(x)}{t''t}\end{aligned}$$

Theorem 4.3. *Assume f is of class C^2. Then we have, for all $x \in U$ and $v, w \in V$,*

$$\partial_v(\partial_w f)(x) = \partial_w(\partial_v f)(x),$$

and the second order differential *at x is a symmetric bilinear map*

$$D^2 f(x) : V \times V \to W, \quad (v, w) \mapsto \partial_v(\partial_w f)(x).$$

Proof. The formula given above yields, for $t' = 0$, $v' = 0$ and $t = t'' \neq 0$,

$$f^{[2]}((x, v, t), (x', 0, 0), t) = \frac{f(x + tx' + tv) - f(x + tx') - f(x + tv) + f(x)}{t^2}.$$

Obviously, this is symmetric in v and x'. Thus, for all $t \neq 0$,

$$f^{[2]}((x, v, t), (x', 0, 0), t) = f^{[2]}((x, x', t), (v, 0, 0), t).$$

Since both sides are continuous functions of t, we have equality also for $t = 0$. But

$$\partial_v(\partial_w f)(x) = \partial_{(v,0,0)} f^{[1]}(x, w, 0) = f^{[2]}((x, w, 0), (v, 0, 0), 0), \tag{4.1}$$

whence the first claim, i.e., symmetry of $D^2 f(x)$. Moreover, by Theorem 2.1, first differentials are linear maps, so this expression is linear with respect to v. By symmetry, it is then also linear with respect to w. □

5. Second order Taylor expansion

Let $f : U \to W$ be of class C^2. Fix $x \in U$. If $(x, v, t) \in U^{[1]}$, we can write

$$\begin{aligned} f(x + tv) &= f(x) + t f^{[1]}(x, v, t) \\ &= f(x) + t\big(f^{[1]}(x, v, 0) + t f^{[2]}((x, v, 0), (0, 0, 1), t)\big) \\ &= f(x) + t Df(x)v + t^2 f^{[2]}((x, v, 0)(0, 0, 1), 0) + t^2 R_2(x, v, t), \end{aligned}$$

where the remainder $R_2(x, v, t) = f^{[2]}((x, v, 0), (0, 0, 1), t)) - f^{[2]}((x, v, 0), (0, 0, 1), 0))$ depends continuously on t and satisfies $R_2(x, v, 0) = 0$. The following lemma shows that this expansion coincides indeed with the second order Taylor expansion:

Lemma 5.1. *If f is C^2, then $f^{[2]}((x, v, 0), (0, 0, 1), 0) = \frac{1}{2} D^2 f(x)(v, v)$.*

Proof. A conceptual proof is given in [17], [B2] or [B3, XIV.3.3]. Another, more "elementary" proof is proposed in Exercice B.7.7 of [B3]: first, observe that

$$\frac{\partial}{\partial t} f^{[1]}(x, v, t) = f^{[2]}((x, v, t), (0, 0, 1), 0). \tag{5.1}$$

Deriving both sides of $f(x + tv) - f(x) = t \cdot f^{[1]}(x, v, t)$ partially with respect to t,

$$f^{[1]}(x + tv, v, 0) = f^{[1]}(x, v, t) + t \cdot \frac{\partial}{\partial t} f^{[1]}(x, v, t),$$

we see that $\frac{\partial}{\partial t} f^{[1]}(x, v, t)$ is a continuous difference factorizer, as follows:

$$\frac{\partial}{\partial t} f^{[1]}(x, v, t) = f^{[2]}((x, v, t), (v, 0, -1), t). \tag{5.2}$$

Comparing (5.1) and (5.2) for $t = 0$, and by (4.1) and linearity, the claim follows. □

E-mail address: wolfgang.bertram(Klammeraffe)univ-lorraine.fr